OF INDIVIDUAL AND COLLABORATIVE EXPERIENCES: TRAINING AND LEARNING IN IMMERSIVE ENVIRONMENTS FOR MEDICAL EDUCATION

DISSERTATION

zur Erlangung des akademischen Grades

Doktoringenieur (Dr.-Ing.)

angenommen durch die Fakultät für Informatik
der Otto-von-Guericke-Universität Magdeburg

von **M.A. Danny Schott**

geb. am 22.02.1990 in Osterburg

Gutachter

Prof. Dr. Bernhard Preim
Prof. Dr. Daniel Roth
Prof. Dr. Dieter Schmalstieg

Magdeburg, den 03.04.2025

Dedicated to
— E. & T. —

ACKNOWLEDGMENTS

I did it for me. I liked it. I was good at it.
And I was really... I was alive.

— Walter White

The work for this thesis was partially funded by the Federal Ministry of Education and Research within the Forschungscampus STIMULATE under Grant 13GW0473A and by the German Federal Ministry for Economic Affairs and Climate Action under Grant 16KN093942.

ABSTRACT

The ongoing digital transformation is fundamentally reshaping the landscape of medical education. Mixed Reality (MR) technologies contribute significantly to these advancements by introducing immersive simulations that enhance user engagement and learning outcomes. However, designing such systems for educational purposes involves addressing interdisciplinary challenges. A crucial aspect of this process is understanding students' needs and the pedagogical framework while integrating technological considerations. Therefore, the overarching question in this thesis is: How can immersive experiences be designed to enrich medical education?

The first part of this work addresses the design of virtual environments by exploring the impact of visual and interactive fidelity in medical task simulations. This investigation provides a practical framework for balancing realism, development effort, and user needs. The findings indicate that higher fidelity enhances the user experience, while no significant differences in task performance were observed across the selected interaction modalities.

The second part focuses on the field of anatomical education and introduces a fully immersive virtual environment designed to foster individual knowledge construction. Through interdisciplinary collaboration, a system was developed that leverages natural hand interaction to enhance understanding of embryonic heart development. The results indicate that effective visualizations in MR must strike a balance between simplicity and sufficient contextual detail, while interactions should cater to varying levels of user expertise and spatial reasoning. Furthermore, when this application was used as a supplementary learning tool during exam preparation by medical students, it demonstrated a measurable knowledge gain.

Building on these insights, the third part focuses on the development of collaborative approaches across different MR systems and compares them to individual learning applications. The study concluded that there were no significant differences in educational outcomes between individual learning environments and collaborative setups, as both effectively supported knowledge acquisition for embryonic heart development. While individual environments facilitated learning with greater user control, collaborative approaches enhanced social presence and teamwork dynamics.

The final part extends the focus to advanced practical training in the context of liver surgery. A cross-modality MR-based platform was developed to explore different modalities and learning approaches using curated clinical use cases. A study involving teachers and stu-

dents demonstrated the potential of this cross-modality system to effectively support both collaborative and explorative learning in liver surgery education.

This thesis explores how user-centered principles can guide the design of MR systems for medical education, addressing technical and pedagogical challenges through varied technologies, learning approaches, and interaction principles. It further advances understanding by examining human-computer interaction methods and interaction design decisions to improve MR-based educational tools.

ZUSAMMENFASSUNG

Die fortschreitende digitale Transformation verändert die Landschaft der medizinischen Ausbildung grundlegend. Mixed-Reality-Technologien (MR) tragen wesentlich zu diesen Fortschritten bei, indem sie immersive Simulationen ermöglichen, die sowohl die Motivation der Lernenden erhöhen als auch das Lernen verbessern. Die Entwicklung solcher Systeme erfordert jedoch die Bewältigung interdisziplinärer Herausforderungen. Ein zentraler Aspekt dabei ist das Verständnis der Bedürfnisse von Lernenden und der pädagogischen Rahmenbedingungen, während gleichzeitig technologische Überlegungen einbezogen werden müssen. Die zentrale Fragestellung dieser Arbeit lautet daher: Wie können immersive Erlebnisse gestaltet werden, um die medizinische Ausbildung zu bereichern?

Der erste Teil dieser Arbeit beschäftigt sich mit der Gestaltung virtueller Umgebungen, indem der Einfluss visueller und interaktiver Qualitäten in medizinischen Aufgabensimulationen untersucht wird. Diese Untersuchung liefert einen praxisorientierten Rahmen, um Realismus, Entwicklungsaufwand und Benutzerbedürfnisse in Einklang zu bringen. Die Ergebnisse zeigen, dass eine höhere visuelle Qualität die Nutzererfahrung verbessert, während bei den gewählten Interaktionsmodalitäten keine signifikanten Unterschiede in der Aufgabenleistung festgestellt wurden.

Der zweite Teil konzentriert sich auf den Bereich der anatomischen Ausbildung und stellt eine vollständig immersive virtuelle Umgebung vor, die darauf ausgelegt ist, individuelle Wissensbildung zu fördern. In einer interdisziplinären Zusammenarbeit wurde ein System entwickelt, das natürliche Interaktionen nutzt, um das Verständnis der embryonalen Herzentwicklung zu verbessern. Die Ergebnisse zeigen, dass effektive Visualisierungen in MR ein Gleichgewicht zwischen Einfachheit und ausreichendem kontextuellem Detail finden müssen, während die Interaktionen auf unterschiedliche Kompetenzstufen und räumliches Vorstellungsvermögen der Nutzer abgestimmt sein sollten. Darüber hinaus wurde bei der Verwendung dieses Systems als unterstützendes Lernwerkzeug während der Prüfungsphase von Medizinstudierenden ein messbarer Wissenszuwachs festgestellt.

Aufbauend auf diesen Erkenntnissen befasst sich der dritte Teil mit der Entwicklung kollaborativer Ansätze auf verschiedenen MR-Systemen und vergleicht diese mit individuellen Lernanwendungen. Die Studie ergab keine signifikanten Unterschiede in den Lernergebnissen zwischen individuellen Lernumgebungen und kollaborativen Ansätzen, jedoch unterstützten beide effektiv die Wissensver-

mittlung zur embryonalen Herzentwicklung. Während individuelle Umgebungen das Lernen durch mehr Benutzerkontrolle erleichterten, verbesserten kollaborative Ansätze die soziale Interaktion und die Teamdynamik.

Der abschließende Teil erweitert den Fokus auf einen fortgeschrittenen Ausbildungszweig im Kontext der Leberchirurgie. Eine plattformübergreifende MR-basierte Anwendung wurde entwickelt, um verschiedene Modalitäten und Lernansätze mithilfe kuratierter klinischer Anwendungsfälle zu erforschen. Eine Studie mit Lehrenden und Studierenden zeigte das Potenzial dieses plattformübergreifenden Systems, um sowohl kollaboratives als auch exploratives Lernen in der Leberchirurgie-Ausbildung effektiv zu unterstützen.

Diese Arbeit untersucht, wie nutzerzentrierte Prinzipien die Gestaltung von MR-Systemen für die medizinische Ausbildung leiten können, indem technische und pädagogische Herausforderungen durch unterschiedliche Technologien, Lernansätze und Interaktionsprinzipien adressiert werden. Sie erweitert das Verständnis, indem Methoden und Interaktionsdesign-Entscheidungen analysiert werden, um MR-basierte Systeme für die medizinische Ausbildung zu verbessern.

CORE PUBLICATIONS

The *Core* contributions I made to this dissertation are based on the publications listed below. Some ideas, illustrations, and wordings in this thesis have previously appeared in these publications. Unless explicitly stated otherwise, in cases where I am listed as the first author, I was responsible for designing, planning, and implementing the research methodology, as well as writing the paper.

[Core1] **Danny Schott**, Matthias Kunz, Florian Heinrich, Jonas Mandel, Anne Albrecht, Rüdiger Braun-Dullaeus, and Christian Hansen. “Stand Alone or Stay Together: An In-situ Experiment of Mixed-Reality Applications in Embryonic Anatomy Education.” In: *Proceedings of the 2024 Symposium on Virtual Reality Software and Technology (VRST)*. Trier, Germany: ACM, 2024, pp. 1–10. ISBN: 979-8-4007-0535-9. DOI: 10.1145/3641825.3687706.

[Core2] **Danny Schott**, Florian Heinrich, Matthias Kunz, Jonas Mandel, Anne Albrecht, Rüdiger Braun-Dullaeus, and Christian Hansen. “CardioCoLab: Collaborative Learning of Embryonic Heart Anatomy in Mixed Reality.” In: *Eurographics Workshop on Visual Computing for Biology and Medicine*. Ed. by Laura Garrison and Daniel Jönsson. The Eurographics Association, 2024. ISBN: 978-3-03868-244-8. DOI: 10.2312/vcbm.20241191.

[Core3] Matthias Kunz, **Danny Schott**, Tom Wunderling, Martin Halloul, Christian Hansen, Anne Albrecht, and Rüdiger Braun-Dullaeus. “Embryonic heart development as an immersive experience: Unveiling learning effects and influential factors in virtual learning environments.” In: *Computers in Biology and Medicine* 187 (2025), p. 109638. ISSN: 0010-4825. DOI: 10.1016/j.compbiomed.2024.109638.

[Core4] **Danny Schott**, Matthias Kunz, Jonas Mandel, Lovis Schwenderling, Rüdiger Braun-Dullaeus, and Christian Hansen. “An AR-Based Multi-User Learning Environment for Anatomy Seminars.” In: *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE. 2024, pp. 949–950. DOI: 10.1109/VRW62533.2024.00271.

[Core5] **Danny Schott**, Florian Heinrich, Lara Stallmeister, Julia Moritz, Bennet Hensen, and Christian Hansen. "Is this the vReal Life? Manipulating Visual Fidelity of Immersive Environments for Medical Task Simulation." In: *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE. 2023, pp. 1171–1180. DOI: 10.1109/ISMAR59233.2023.00134.

[Core6] **Danny Schott**, Matthias Kunz, Tom Wunderling, Florian Heinrich, Rüdiger Braun-Dullaeus, and Christian Hansen. "Cardiogenesis4d: Interactive morphological transitions of embryonic heart development in a virtual learning environment." In: *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 29.5 (2023), pp. 2615–2625. DOI: 10.1109/TVCG.2023.3247110.

[Core7] **Danny Schott**, Patrick Saalfeld, Gerd Schmidt, Fabian Joeres, Christian Boedecker, Florentine Huettl, Hauke Lang, Tobias Huber, Bernhard Preim, and Christian Hansen. "A VR/AR Environment for Multi-User Liver Anatomy Education." In: *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. 2021, pp. 296–305. DOI: 10.1109/VR50410.2021.00052.

FURTHER PUBLICATIONS

During my doctoral studies, I have also contributed to *Further* publications that have influenced my research. These publications are listed in the following section.

[Further1] Dorothea Wagnerberger, **Danny Schott**, Lovis Schwenderling, Christian Hansen, and Dominik Schumacher. "Empowering Patients: Improve Gender-Sensitive Medical Knowledge Through Interactive Edutainment." In: *Proceedings of the Nordic Conference on Human-Computer Interaction (NordiCHI)*. Uppsala, Sweden: ACM, 2024, pp. 1–12. ISBN: 979-8-4007-0966-1. DOI: 10.1145/3679318.3685500.

[Further2] Vuthea Chheang, **Danny Schott**, Patrick Saalfeld, Lukas Vradelis, Tobias Huber, Florentine Huettl, Hauke Lang, Bernhard Preim, and Christian Hansen. "Advanced liver surgery training in collaborative VR environments." In: *Computers & Graphics* 119 (2024), p. 103879. DOI: 10.1016/j.cag.2024.01.006.

[Further3] Josefine Schreiter, Florian Heinrich, Benjamin Hatscher, **Danny Schott**, and Christian Hansen. "Multimodal Human-Computer Interaction in Interventional Radiology and Surgery: A Systematic Literature Review." In: *Int J CARS* (2024). DOI: 10.1007/s11548-024-03263-3.

[Further4] Tonia Mielke, Fabian Joeres, **Danny Schott**, and Christian Hansen. "Interactive Registration Methods for Augmented Reality in Robotics: A Comparative Evaluation." In: *2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. 2023, pp. 501–506. DOI: 10.1109/ISMAR-Adjunct60411.2023.00109.

[Further5] **Danny Schott**, Julia Moritz, Christian Hansen, and Fabian Joeres. "The UUXR-Framework: A Draft Classification for Using Extended Reality in Usability and User Experience Research." In: *2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. 2023, pp. 460–465. DOI: 10.1109/ISMAR-Adjunct60411.2023.00100.

[Further6] Josefine Schreiter, Tonia Mielke, **Danny Schott**, Maximilian Thormann, Jazan Omari, Maciej Pech, and Christian Hansen. "A multimodal user interface for touchless control of robotic ultrasound." In: *International journal of computer assisted radiology and surgery* 18.8 (2023), pp. 1429–1436. DOI: 10.1007/s11548-022-02810-0.

[Further7] Vuthea Chheang, **Danny Schott**, Patrick Saalfeld, Lukas Vradelis, Tobias Huber, Florentine Huettl, Hauke Lang, Bernhard Preim, and Christian Hansen. "Towards Virtual Teaching Hospitals for Advanced Surgical Training." In: *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 2022, pp. 410–414. DOI: 10.1109/VRW55335.2022.00089.

[Further8] **Danny Schott**, Florian Heinrich, Dominic Labsch, Bennet Hensen, and Christian Hansen. "Towards multimodal interaction for needle-based procedures in a virtual radiology suite." In: *Current Directions in Biomedical Engineering* 8.1 (July 2022), pp. 70–73. DOI: 10.1515/cdbme-2022-0018.

[Further9] **Danny Schott**, Florian Heinrich, Lara Stallmeister, and Christian Hansen. "Exploring object and multi-target instrument tracking for AR-guided interventions." In: *Current Directions in Biomedical Engineering* 8.1 (2022), pp. 74–77. DOI: 10.1515/cdbme-2022-0019.

[Further10] Josefine Schreiter, **Danny Schott**, Lovis Schwenderling, Christian Hansen, Florian Heinrich, and Fabian Joeres. "AR-Supported Supervision of Conditional Autonomous Robots: Considerations for Pedicle Screw Placement in the Future." In: *Journal of Imaging* 8.10 (2022). ISSN: 2313-433X. DOI: 10.3390/jimaging8100255.

[Further11] Christian Boedecker, Tilman Borchardt, Florentine Huettl, Natascha Müller, **Danny Schott**, Patrick Saalfeld, Christian Hansen, Hauke Lang, and Tobias Huber. "In virtual reality durch die pandemie: Digitale lehre mit modernster technologie in der leberchirurgie." In: *Zeitschrift für Gastroenterologie* 59.08 (2021), e218. DOI: 10.1055/s-0041-1733640.

[Further12] Fabian Joeres, Florian Heinrich, **Danny Schott**, and Christian Hansen. "Towards natural 3D interaction for laparoscopic augmented reality registration." In: *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization* 9.4 (2021), pp. 384–391. DOI: 10.1080/21681163.2020.1834877.

[Further13] Remigiusz Kwapik, Julia Moritz, Bennet Hensen, Benedikt Janny, Enrico Pannicke, **Danny Schott**, Georg Rose, Oliver Speck, and Frank Wacker. "Virtual reality-based usability laboratory for interventional mr applications." In: *Proc 5th Conf on Image-Guided Interventions*. 2021, pp. 51–52. URL: https://www.igic.de/deutsch/igic-2021/proceedings/index.html.

[Further14] **Danny Schott**, Benjamin Hatscher, Fabian Joeres, Mareike Gabele, Steffi Hußlein, and Christian Hansen. "Lean-Interaction: passive image manipulation in concurrent multitasking." In: *Proceedings of Graphics Interface 2020*. GI 2020. University of Toronto: Canadian Human-Computer Communications Society / Société canadienne du dialogue humain-machine, 2020, pp. 404–412. ISBN: 978-0-9947868-5-2. DOI: 10.20380/GI2020.40.

CONTENTS

1 INTRODUCTION 1
- 1.1 Motivation 2
- 1.2 Research Questions 3
- 1.3 Research Contribution 4
- 1.4 Thesis Structure 5

2 BACKGROUND 7
- 2.1 Introduction to Mixed Reality 8
  - 2.1.1 A kind of taxonomy 8
  - 2.1.2 Perceptual Constructs in MR Systems 11
  - 2.1.3 Terminology Clarifications for this thesis 13
  - 2.1.4 Technical Overview 14
- 2.2 Educational Background 18
  - 2.2.1 Learning Theories & Concepts 19
  - 2.2.2 From Theory to Practice 20
- 2.3 Medical Education 21
  - 2.3.1 General Advantages 21
  - 2.3.2 Anatomy Education 22
  - 2.3.3 Application-Specific Use Cases in this Thesis 23
  - 2.3.4 Requirements for implementation 23
- 2.4 Methodology Background 24
  - 2.4.1 Human-Centered Design 24
  - 2.4.2 Measurement Tools 26
- 2.5 Chapter Summary 31

3 RELATED WORK 33
- 3.1 Effects of Manipulating Realism 34
  - 3.1.1 The Role of Realism in Virtual Environments 34
  - 3.1.2 Advanced Input and Feedback 35
- 3.2 Immersive Learning Experiences in General 37
- 3.3 Advantages of MR in Anatomy Education 37
- 3.4 Collaborative Learning 39
- 3.5 Learning Embryology 40
- 3.6 Learning Liver Anatomy 43
- 3.7 Chapter Summary 45

4 VISUAL AND INTERACTION FIDELITY 47
- 4.1 Introduction 48
- 4.2 Experiment 1 50
  - 4.2.1 Materials 50
  - 4.2.2 Evaluation 57
  - 4.2.3 Results 65
  - 4.2.4 General discussion 70
  - 4.2.5 Conclusion 74
- 4.3 Ongoing Work 75

4.3.1 Interaction and Visualization Rationale 75
4.3.2 Apparatus Design 76
4.3.3 Planned Variables 78
4.4 Chapter Summary 79
5 INDIVIDUAL LEARNING ENVIRONMENTS 81
5.1 Introduction 82
5.2 Experiment 2 84
5.2.1 Material & Methods 84
5.2.2 Evaluation 91
5.2.3 Results 96
5.2.4 Discussion 101
5.2.5 Limitations 102
5.2.6 Conclusion 103
5.3 Experiment 3 104
5.3.1 Material 104
5.3.2 Evaluation 106
5.3.3 Results 112
5.3.4 Discussion 116
5.3.5 Conclusion 119
5.4 Chapter Summary 119
6 COLLABORATIVE LEARNING ENVIRONMENTS 121
6.1 Introduction 122
6.2 Experiment 4 125
6.2.1 Material 125
6.2.2 Evaluation 128
6.2.3 Results and Discussion 130
6.2.4 Conclusion and Future Research 131
6.3 Experiment 5 131
6.3.1 Material 132
6.3.2 Evaluation 134
6.3.3 Results 135
6.3.4 Discussion & Conclusion 138
6.4 Experiment 6 139
6.4.1 Motivation 139
6.4.2 Material and Methods 140
6.4.3 Evaluation 144
6.4.4 Results 150
6.4.5 Discussion & Conclusion 155
6.5 Chapter Summary & Ongoing Work 158
7 CROSS-MODALITY LEARNING 161
7.1 Introduction 162
7.2 Experiment 7 164
7.2.1 Material 164
7.2.2 Development of the CMVLE 164
7.2.3 Evaluation 172
7.2.4 Results 176

7.2.5 Discussion 178
7.2.6 Conclusion 181
7.3 Chapter Summary 182
8 CLOSING 183
8.1 Thesis Summary 184
8.1.1 Effects of Visual and Interaction Fidelity on Medical Task Simulations 184
8.1.2 Individual Learning Environments 185
8.1.3 Collaborative Learning Environments 186
8.1.4 Cross-Modality Learning Environments 188
8.2 Limitations & Future Directions 190
8.2.1 Experimental Design and Fidelity 190
8.2.2 Representation of Users and Interaction Modalities 190
8.2.3 Evaluation Metrics and Cognitive Load 191
8.2.4 Development Limitations and Hardware Variability 191
8.2.5 Generalizability 191
8.3 General Contribution 192
9 APPENDIX 193
9.1 Videos 194
9.2 Demographic Data Sheet 195
9.3 Anatomy Knowledge Test 196
9.3.1 Category 1: Shape 196
9.3.2 Category 2: Time 197
9.3.3 Category 3: Location 199
9.4 Mental Rotation Test 200
9.5 Advanced Embryology Knowlege Test 202
9.6 Moderators Training Questions 205
9.7 Expert Interview Questions 207

BIBLIOGRAPHY 209

LIST OF FIGURES

Figure 1 Reality-virtuality continuum 9
Figure 2 Fidelity Spectrum of MR Systems 18
Figure 3 VLE Model 21
Figure 4 Training System Requirements 25
Figure 5 Workflow of MRI-guided interventions 52
Figure 6 Visual Fidelity in a Virtual Environment 55
Figure 7 Input modalities and realized Visual Fidelity . 57
Figure 8 Realization of subtasks in reality and MR . . . 59
Figure 9 Significant main effects of the visual fidelity . 66
Figure 10 Significant interaction effects on realism and UX 67
Figure 11 Significant main and interaction effects on PD 70
Figure 12 Executed Needle Insertion Task 77
Figure 13 Overview of the new apparatus. 78
Figure 14 Physical and virtual representations 87
Figure 15 Collage of multiple stages of the heart tube . . 88
Figure 16 Modeling mode and Navigation Cube 88
Figure 17 Learning Mode with *passive* features 90
Figure 18 User interacting in lab 93
Figure 19 Descriptive results of active and passive features 95
Figure 20 Correlation between MRT and AKT results . . 98
Figure 21 Descriptive results of the individual first and second training 99
Figure 22 Overview of all 3D model deformations 105
Figure 23 Enhanced application interface 107
Figure 24 Study design flowchart of MR application . . . 108
Figure 25 Comparison of knowledge test points scored . 114
Figure 26 Results of the questionnaires measuring *Sense of Presence*, immersive tendencies, and UX . . . 115
Figure 27 Results of the N-TLX 116
Figure 28 User wearing HMD and see-through 127
Figure 29 Evaluated configurations 129
Figure 30 Photograph through MR glasses 133
Figure 31 Photomontage of the setup 134
Figure 32 ARI, SUS and TAM ratings 137
Figure 33 GEQ ratings 138
Figure 34 Screenshots of evaluated MR systems 141
Figure 35 Overview of student's UI 142
Figure 36 Overview of instructor's UI 144
Figure 37 Study design flowchart of the MR applications 146
Figure 38 Differences of means of investigated measures 151

Figure 39	Results of knowledge tests in relation to the learning groups	152
Figure 40	Results of knowledge tests in relation to learning environments	152
Figure 41	Interactive stations	166
Figure 42	User interacting with virtual organ	167
Figure 43	Virtual DICOM Data	168
Figure 44	Multi-user scene view	169
Figure 45	System architecture	172
Figure 46	VST MR and Desktop-Mode	173
Figure 47	IPQ and SUS scores	176

LIST OF TABLES

Table 1	Breakdown of the different visual fidelities	56
Table 2	ANOVA and robust ANOVA results summary	63
Table 3	Summary of descriptive results for the dependent variables related to presence	64
Table 4	Summary of descriptive results for the dependent variables related to UX and performance	65
Table 5	Summary of descriptive results with respect to demographics	94
Table 6	Summary of MRT Pearson correlation test results	98
Table 7	Summary and quantity of statements	99
Table 8	Demographic Data Summary	111
Table 9	Correlation overview	113
Table 10	Summary of qualitative feedback	118
Table 11	Demographic participant data.	148
Table 12	Summary of descriptive results for ILE	153
Table 13	Summary of descriptive results for CLE	153
Table 14	Summary of statistical analyses on dependent variables	154
Table 15	Sorting types and corresponding parameters	165
Table 16	Collected statements of respective stations	177

ACRONYMS & ABBREVIATIONS

2D	Two-dimensional
3D	Three-dimensional
4D	Four-dimensional
ADL	Advanced Detail Level
AI	Artificial Intelligence
AKT	Anatomy Knowledge Test
ANOVA	Analysis of Variance
AR	Augmented Reality
ARIQ	Augmented Reality Immersion Questionnaire
AV	Augmented Virtuality
Avg	Average
CT	Computer Tomography
CLE	Collaborative Learning Environment
CMVLE	Cross-Modality Virtual Learning Environment
CAVE	Cave Automatic Virtual Environment
CP	Co-presence
DICOM	Digital Imaging and Communications in Medicine
DOF	Degree of Freedom
EDL	Essential Detail Level
EKT	Embryology Knowledge Test
GEK	General Embryology Knowledge
GEQ	Game Experience Questionnaire
HCI	Human-Computer Interaction
HCD	Human-Centered Design
HEK	Heart Embryology Knowledge
HMD	Head-Mounted Display
ILE	Individual Learning Environment
IM	Interaction Modality
IPQ	Igroup Presence Questionnaire
ITQ	Immersive Tendencies Questionnaire
M	Mean
Mdn	Median

MRT	Mental Rotation Test
MRI	Magnetic Resonance Imaging
MR	Mixed Reality
MRTK	Mixed Reality Toolkit
MPR	Multiplanar Reconstruction
N-TLX	NASA Task Load Index
PD	Placement Deviation
PU	Perceived Usefulness
PEU	Perceived Ease of Use
RVC	Reality-Virtuality Continuum
SDK	Software Development Kit
SDL	Standard Detail Level
SI	Social Interaction
SD	Standard Deviation
SE	Standard Error
SUS	System Usability Scale
TCT	Task Completion Time
TAP	Think Aloud protocol
TAM	Technology Acceptance Model
UEQ	User Experience Questionnaire
UI	User Interface
UX	User Experience
VE	Virtual Environment
VLE	Virtual Learning Environment
VBO	Virtual Body Ownership
VF	Visual Fidelity
VR	Virtual Reality
VST	Video-See-Through
XR	eXtended Reality

INTRODUCTION

ch. 1

SYNOPSIS *This introduction outlines the motivation behind this thesis, formulates the guiding research questions, and provides a comprehensive overview of its structure.*

1.1 MOTIVATION

How can immersive experiences be designed to enrich medical education? This is the guiding question of this thesis. To begin, I would like to share a personal motivation that has driven me throughout this work. From the start of my doctoral studies, I have been inspired by the prospect of creating systems that support those who dedicate their lives to helping others: healthcare professionals.

To truly support them, improving their education might be a vital step. This thesis focuses on addressing the challenges involved in integrating Mixed Reality (MR) into medical education. While MR cannot replace traditional methods, this work seeks to demonstrate how these technologies can be utilized as complementary tools to enhance the learning process. Therefore, equally compelling is the scientific motivation underpinning this work: to explore and advance systems that bridge technology and pedagogy, ultimately enriching learning experiences and improving educational outcomes.

The ongoing digital transformation is already reshaping the landscape of medical education [64]. The growing field of "Digital Health" encompasses a broad spectrum of technologies, including MR, Artificial Intelligence (AI), robotics in surgery, and various digital health applications. These innovations are expanding the possibilities of medical education, from enhanced anatomical understanding to improved medical skill training, while minimizing patient risks [213, 214, 231]. MR technologies add a new dimension to these advancements, offering immersive simulations that enable repeated practice without ethical dilemmas, significantly improving the comprehension of complex spatial relationships and boosting learner engagement and performance outcomes [91, 102, 296].

However, designing MR systems for education also entails interdisciplinary challenges. As Drey [74] highlights, "The obstacles such research has to overcome are further interdisciplinary, as interaction theory driven by Human-Computer Interaction (HCI) strongly influences the MR learning environment's design and, consequently, learning outcomes." Moreover, a crucial aspect of creating impactful systems is a deep understanding of user needs. As an interaction designer, this one aspect has always been paramount to me: engaging directly with people—in this case, the potential users of these systems—to uncover their genuine needs. This commitment to understanding user needs is at the core of the Human-Centered Design (HCD) approach employed in this thesis. It serves as the foundational methodology for addressing the guiding question and for developing meaningful, learner- and teacher-centered solutions that bridge the gap between technology and pedagogy.

Ultimately, this thesis is shaped by the motivation to create MR-based systems that enhance the education of healthcare professionals while addressing the scientific challenge of advancing medical education through user-centered approaches.

1.2 RESEARCH QUESTIONS

To address the challenges of integrating MR systems into medical education, this thesis is guided by a series of research questions, explored through seven distinct experiments. The initial step involved identifying components that play a crucial role in the design of Virtual Environments (VEs). For this purpose, a medical task simulation was developed to investigate the effects of visual and interaction fidelity in MR. This effort led to the following research question:

RQ1 | What are the critical visual and interaction fidelity factors that contribute to creating engaging and effective medical task simulations in MR?

Subsequently, the focus shifted to identifying a specific application for medical education. This thesis emphasizes anatomical education, with a particular problem space defined around embryonic heart development. An Individual Learning Environment (ILE) tailored for this purpose was developed, resulting in the following research question:

RQ2 | How can suitable visualizations and interactions in MR be designed to effectively represent embryonic heart development?

Following this, an investigation was conducted to determine whether the use of this ILE, tailored for understanding embryonic heart development, has a measurable effect on learning outcomes and influential factors. This inquiry led to the next research question:

RQ3 | Are there measurable learning effects when using MR to understand embryonic heart development, and which factors influence these outcomes?

Given the motivation to explore various technical and pedagogical approaches, the development of a Collaborative Learning Environment (CLE) became a priority. This effort required examining how collaborative approaches could be technically implemented and pedagogically integrated from the perspective of educators. This gave rise to the following research question:

RQ4 | What are the technical and pedagogical requirements for a collaborative MR-based system to effectively support the learning of embryonic heart development?

Building on these insights into feasibility, the next step was to determine how the CLE could be practically integrated into medical training. This was explored through a simulated anatomy seminar, resulting in the following research question:

RQ5 | How can a collaborative MR-based learning environment for understanding embryonic heart development be effectively integrated into an anatomy seminar setting?

To address a gap in the literature, which lacks direct comparisons of ILEs and CLEs for anatomy education within MR-based learning environments, two distinct Virtual Learning Environments (VLEs) were developed. This effort aimed to answer the following research question:

RQ6 | How do individual and collaborative MR-based learning environments differ in supporting educational outcomes for embryonic heart development?

Finally, while this thesis predominantly focuses on foundational anatomy education (e.g., embryonic heart development), it extends to advanced practical training in the context of liver surgery. To explore this, a cross-modality MR-based platform was developed to investigate different modalities and address the following research question:

RQ7 | What design principles can enhance collaborative MR environments for advanced medical training, specifically in liver anatomy education, by integrating real clinical cases and accommodating varying levels of immersion?

1.3 RESEARCH CONTRIBUTION

This thesis contributes to the understanding of how HCD principles can be effectively applied in the design of MR systems for medical education. It provides insights into technical challenges by exploring various technologies, visualization techniques, and interaction principles, while also addressing pedagogical challenges through the examination of different learning approaches. Additionally, the thesis delves into the use of HCI research methods, accompanied by comprehensive system descriptions. By discussing interaction design decisions, it aims to advance the understanding and development of MR-based medical educational applications.

This dissertation describes and discusses the findings of seven peer-reviewed publications, referenced as *Core* contributions [Core1] - [Core7], which are incorporated into this thesis through seven distinct experiments. These works have been published in conference proceedings and journals and were presented at international conferences, including ACM VRST, IEEE VR, IEEE ISMAR, and EG VCBM. Among these, one article [Core6] was published in the journal TVCG, and another [Core3] was under review at a journal at the time of submission. Two additional papers [Core2, Core4], while representing smaller contributions as a poster and a short paper respectively, served as foundational research leading to the outcomes presented in

[Core1]. These smaller works are integral links in the iterative overall process of this dissertation.

Research is inherently collaborative, and I would like to emphasize that the *Core* publications included in this thesis are the result of teamwork among all contributing authors. To ensure better readability and seamless integration into the overall flow of this dissertation, the content is primarily presented in a passive voice.

Beyond the contributions central to this thesis, my PhD journey encompassed additional scientific projects that significantly shaped my research direction. While not all of these works align directly with the focus of this dissertation, they have been influential and are documented as *Further* contributions [Further1] - [Further14]. These include both first-authored and co-authored publications in conference proceedings and journals, as well as presentations at international and national conferences, such as ACM NordiCHI, Computers & Graphics, and Current Directions in Biomedical Engineering.

1.4 THESIS STRUCTURE

This dissertation is structured as follows:

- *Chapter 2* discusses the conceptual and technological foundations of MR systems and VE, focusing on their specific requirements and challenges. Building on this, it delves into the pedagogical background, translating relevant learning theories into practical requirements. Additionally, the chapter highlights the core aspect of this thesis: the application of MR in medical education, particularly in anatomical training. Finally, it outlines fundamental approaches and methods to effectively address the identified challenges.

- *Chapter 3* provides an overview of related work in medical education and medical simulations, focusing on the impact of manipulating visual and interaction fidelity in VEs, exploring different didactic approaches such as collaborative settings, and examining the use of MR technologies in anatomy education. It also highlights two distinct focal areas examined in this thesis: the development of the embryonic heart and the training for liver surgery.

- *Chapter 4* investigates the effects of various levels of fidelity in MR, focusing on user performance, User Experience (UX), and *Presence* in medical task simulations. It examines how different degrees of Visual Fidelity (VF) and Interaction Modalitys (IMs) replicate real-world interactions. The presented experiment evaluates these aspects to establish design guidelines for future MR-based medical training tools, emphasizing the nuanced impact of realism on user engagement and performance.

The chapter concludes with an outlook, outlining plans for further development of these methods.

- *Chapter 5* introduces a VLE aimed at supporting early stage medical education, helping students better understand the dynamic morphological changes that occur during embryonic heart development. The VLE seeks to make these complex processes more accessible. The chapter begins with an overview of the medical background, current teaching methods, and the transition to Four-dimensional (4D) heart models, followed by a user study. A subsequent experiment further develops the application and methods, assessing its effectiveness and investigating questions related to its impact on learning outcomes and influential factors in MRs system characteristics.

- *Chapter 6* presents the iterative development of a collaborative MRs-based learning system, shaped by the requirements of both experts and students. Through three interlinked experiments, the potential integration of a VLEs into concrete anatomical training scenarios is explored. This culminates in a comparison of different didactic approaches within a shared VLEs, aimed at examining their effectiveness in enhancing teaching and learning outcomes.

- *Chapter 7* presents the development of a collaborative and cross-modal learning MRs platform designed as a tool for liver anatomy education. With support from surgeons, curated clinical use cases were implemented into diverse MRs approaches, and various teaching and learning scenarios were explored.

- Finally, *Chapter 8* summarizes the findings of this thesis by addressing the research questions posed at the beginning and throughout the individual experiments. Furthermore, it discusses limitations and highlights directions for future research.

A video demonstration of the applications described in this thesis is provided in the Appendix (see Section 9.1).

BACKGROUND

ch. 2

SYNOPSIS *How can MR be integrated within the medical context, and what requirements and challenges does this entail? To shed light on these questions, this chapter will first discuss the conceptual and technological foundations of mixed reality systems and virtual environemnts. Building on this, it will address the educational background and translate relevant learning theories into practical requirements. Furthermore, the focus will shift to the core aspect of this work: the use of MR in medical education and training, particularly with regard to anatomical education. Finally, foundational approaches and methods will be explored to address the aforementioned challenges effectively.*

ABOUT THIS CHAPTER *Selected passages in this section draw on previously published peer-reviewed work by the author included in the Core publications ([Core1] - [Core7]). These passages have been adapted and incorporated to provide a coherent and comprehensive background for the related work discussed in this chapter. Reuse of content complies with the respective publishers' policies: articles published by IEEE are reused with permission in accordance with IEEE's thesis reuse policy; open access publications under Creative Commons licenses are reused in compliance with CC BY 4.0; the Elsevier article is reused in accordance with Elsevier's thesis reuse policy; and the ACM article is reused in accordance with ACM's open access guidelines.*

2.1 INTRODUCTION TO MIXED REALITY

This section aims to clarify definitions, introduce relevant theories, and outline key technologies within the MR landscape.

2.1.1 *A kind of taxonomy*

As philosophers might say, "The beginning of wisdom is the definition of terms." But defining a term or concept that lacks clear consensus within the HCI research community, or is often open to varied interpretations, can be confusing and challenging. Nevertheless, in this thesis, I will primarily use the term MR for the theories and technologies discussed. My rationale for this choice is outlined below.

THE CLASSIC DEFINITION When searching recent literature for a definition of MR, a wide range of interpretations emerges [165, 179, 190, 221, 287]. Looking back several years, one of the most widely cited and recognized taxonomies in the research community is the understanding of MR as part of a continuum—the Reality-Virtuality Continuum (RVC) introduced by Milgram and Kishino in 1994 [177]. In their paper, this continuum is defined by two extremes: at one end, the real environment, which "consists solely of real objects," and at the other, the VE, which "consists solely of virtual objects." Everything in between—blending real and virtual elements, though not including these endpoints—is considered MR [274]. Figure 1 shows a schematic representation of this concept, which contains an extension that will be explained in the course of this section.

Milgram and Kishino elaborate that "within this [reality-virtuality] framework it is straightforward to define a generic MR environment as one in which real world and virtual world objects are presented

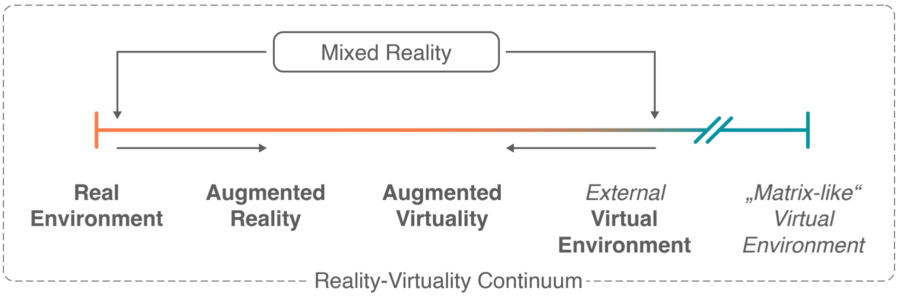


Figure 1: Adapted illustration of Milgram and Kishino's reality-virtuality continuum [177] (■) with an extension based on the revised version by Skarbez et al. [274] (■). Skarbez adds the term "external" to the VE as used by Milgram, and labels the final stage as *Matrix-like VR*, which is also indicated in italics. Licensed under CC BY 4.0.

together within a single display" [177]. Skarbez et al. expand on this, interpreting MR as "any display (interpreted broadly) that presents a combination of real and virtual objects that are perceived at the same time" [274]. They suggest that this combination of real and virtual can be achieved in various ways, such as overlaying virtual objects on the real world with optical- or Video-See-Through (VST) displays, or incorporating real-world content into a virtual world via embedded live video or tracked haptic objects.

Between the extremes of the RVC, there are two main gradations of MR, which vary depending on the amount of virtual content that is added or removed. When a predominantly real environment is supplemented by some virtual elements, such systems are typically referred to as Augmented Reality (AR), as the quantity of real content exceeds that of virtual objects. If the opposite is the case—"either completely immersive, partially immersive, or otherwise, to which some amount of (video or texture mapped) "reality" has been added" [178]—the term Augmented Virtuality (AV) is used. According to this definition, Virtual Reality (VR) is not part of MR but is represented as the "Virtual Environment" anchor at the end of the continuum, describing an interface that contains no real content. AR, on the other hand, is considered a subtype of MR.

THE GRADATIONS Beyond this understanding, researchers have sought to define the individual terms AR and VR more specifically. The most widely recognized definition of AR, also used in textbooks [70, 247], comes from Azuma, who describes AR as "systems that have the following three characteristics: combines real and virtual; interactive in real time; registered in 3D" [16]. The goal of AR, therefore, is not only to merge reality with virtual components but also to create content that can be precisely aligned with the real world through geometric registration. Additionally, the virtual overlay should be capable of adapting to changes in the environment. Definitions of

VR, on the other hand, are often less precise. Example definitions, referenced by Doerner et al. [70], include: "VR refers to the use of three-dimensional displays and interaction devices to explore real-time computer-generated environments" (Steve Bryson, Call for Participation, 1993 IEEE Symposium on Research Frontiers in Virtual Reality) and "Virtual Reality refers to immersive, interactive, multi-sensory, viewer-centered, three-dimensional computer-generated environments and the combination of technologies required to build these environments" (Carolina Cruz-Neira, SIGGRAPH '93 Course Notes, Virtual Reality Overview). In these definitions, the concepts of interaction and real-time are central. The use of Three-dimensional (3D) displays and interaction devices further distinguishes VR from other computer-generated environments, such as simulations and games.

Although these definitions and taxonomies are over two decades old, they remain valid to this day. Skarbez et al. support the "classic" definition of MR and advocate for the continued use of this term, adding, "Instead of requiring that real and virtual objects be combined within a single display, we propose that real and virtual objects and stimuli could be combined within a single percept" [274]. They argue that MR should encompass everything traditionally referred to as VR [274] and that the RVC is, in fact, discontinuous, as the endpoint of "virtual reality" is practically unattainable. In their opinion paper, *It Is Time to Let Go of "Virtual Reality"*, they further support this view by stating, "Every so-called "virtual reality" experience is actually a mixture of virtual (most commonly computer-generated visual and auditory stimuli, but other sensory modalities are also employed) and real (for example, the sensation of the floor under one's feet, the feeling that gravity is down), and hence, mixed reality."[275]. Despite advancing technology and the fact that most prior definitions focused primarily on visual cues and simulations—while modern MR technologies engage more than just visual perception—a truly fully immersive environment remains unattainable with current technology, as our senses would need to be entirely overwritten. Consequently, Skarbez et al. [274] proposed viewing the virtuality anchor of the continuum as a universal "matrix-like" VE—one that cannot yet be achieved with today's technology. They consider traditional VR to be "external virtual environments" and part of MR, as this technology cannot manipulate interoceptive senses, such as proprioception. Figure 1 shows the revised version of the RVC, illustrating that classic VR remains part of MR. Similarly, Wienrich et al. [323] proposed a modification to the RVC, suggesting that the distinction between AR and AV should not rely solely on which content occupies more screen space. Instead, they argue that whether reality or virtuality is augmented depends on the referential influence of the relevant real and virtual elements.

META DEFINITIONS The term eXtended Reality (XR) is often used as an umbrella term for both VR and MR, thus encompassing all previously described concepts as well [70]. Further definitions frame this within the realm of *Mediated Reality*: an overarching concept that influences real-world perception by adding or removing information in real time [164]. This manipulation is typically achieved through electronic devices such as smartphones or headsets, which act as visual filters or overlays on the real world. When information is intentionally removed to manipulate perception of reality, it is referred to as *Diminished Reality* [165].

Another emerging term in the research community is *Cross Reality*, which refers to systems that enable seamless integration and smooth transitions across gradients on the RVC [14, 169, 260, 269, 312]. In this context, *Transitional Interfaces* are also discussed—systems that can be seamlessly integrated as usable, consistent, and coherent user interfaces from various locations within a shared physical space [123].

Additional terms that have gained traction in recent years, particularly in the consumer sector through commercial manufacturers like Apple's Vision Pro[1] and Meta's Quest HMDs[2], include *Metaverse* and *Spatial Computing*. Simon Greenwold first defined *Spatial Computing* in 2003 as "human interaction with a machine in which the machine retains and manipulates referents to real objects and spaces" [97]. Expanding on this, *Spatial Computing* facilitates the digitization of real-world objects, enables interactions between sensors, and enhances the digital representation of the physical world. This process integrates spatial mapping to validate tracking and enables seamless interaction between users and objects across physical and digital realms, ultimately creating an interconnected experience where digital and physical elements interact cohesively [19]. The *Metaverse* is a post-real universe—a persistent, continuous multi-user environment where physical reality and digital virtuality merge. It relies on technologies that facilitate multisensory interactions with VEs, digital objects, and other people, emphasizing connected social experiences within a shared digital space [188].

2.1.2 *Perceptual Constructs in MR Systems*

To evaluate a user's experience in VEs, a variety of subjective constructs have been proposed in the literature [272, 274]. One of the most prevalent—and perhaps most significant is *Presence*, which describes the sense of "being there" within VEs [25]. *Presence* is often defined as the sensation of being in a computer-generated world or a mediated environment [73, 281, 289, 329]. It refers to the subjective experience of being in another space or location, even though one is

1 Apple Inc.: `https://www.apple.com/apple-vision-pro/`
2 Meta Inc.: `https://www.meta.com/quest/`

physically elsewhere [317]. *Presence* can be divided into two main dimensions: *Being There*, which describes the sensation of actually being at a specific location, and *Perceived Realism*, the perception of coherence and reality within the VE. The latter dimension is increasingly seen as crucial for enhancing the overall *Sense of Presence* [316]. *Presence* also includes three primary subcategories, as originally proposed by Lee (2004) and later summarized by Oh, Bailenson, and Welch in their systematic review (2018) [150, 194]:

- *Telepresence* (spatial presence) is defined by Steuer as "the extent to which one feels present in the mediated environment rather than in the immediate physical environment" [289] and refers to the vividness with which a user perceives the spatial qualities of the VE. When telepresence is strongly felt, individuals may lose awareness that their experiences are being mediated by technology [157].

- *Self-presence* is described by Aymerich-Franch et al. as "the extent to which the "virtual self is experienced as the actual self" [15]. Unlike telepresence, self-presence does not focus on the vividness of the environment but rather on the connection the user feels to their virtual body, emotions, or identity [220].

- *Social Presence*, or Co-presence (CP), is described by Biocca et al. as "the sense of being with another" [27] and is influenced by the ease with which a user perceives the intentions and sensory cues of others [26]. Unlike the first two subcategories, *Social Presence* requires a co-present entity that feels sentient, allowing users to experience others as social beings rather than artificial entities [151].

Immersion refers to the objective level of sensory fidelity provided by an MR system [276], while *Presence* is the user's subjective psychological reaction to engaging with that system [24]. Wilkinson summarizes this distinction as: "Presence is an experiential quality within VEs, whereas immersion pertains to the technical attributes of a virtual system that facilitate the user's *Sense of Presence*" [328]. Thus, presence relies on the system's immersion capabilities (such as interaction options, tracking accuracy, graphical resolution, etc.) [316], and the degree to which users "steps inside" a VE can vary based on the application. This experience is also influenced by the extent to which the physical environment remains perceptible [202, 222].

According to Slater, immersive experiences primarily engage the perceptual system rather than the cognitive one. While the senses react directly and automatically to virtual stimuli and environmental dynamics, it is only upon reflection that the cognitive system recognizes the artificial nature of the experience [278]. Slater et al. also distinguish between the "Place Illusion", the user's feeling of being in the location shown, and the “Plausibility Illusion“, the feeling that actions are taking place [277]. When the “Place Illusion“ and the “Plausibility Illusion“ are created in the best possible way, a parallel reality to the real world can be created [279], resulting in high subjectively-perceived “realism“ for the user of a VR application [128].

2.1.3 *Terminology Clarifications for this thesis*

This section does not cover all relevant terms, as the vast array of theories and taxonomies reflects the ongoing scientific debate and underscores the inherent complexity of defining terminology in this field. In summary, establishing a universally accepted definition of all these terms will become increasingly challenging as new technological advancements emerge. Eventually, technology may progress to a point where distinctions are no longer made, and what constitutes MR will largely depend on the context [287]. Since these technologies were not available at the time of this thesis’s submission, I will use the term MR in reference to the approaches developed throughout this work or when discussing VEs enriched with virtual content of any kind. I view this approach as aligned with the “classic definition” mentioned earlier and with Skarbez et al.’s interpretation [177, 274], even though a broader understanding might encompass AR, VR, or something in between or beyond.

In this thesis, the term *Presence* is frequently mentioned in connection with evaluation methods and results. A more detailed explanation of the instruments used to measure *Presence* will be provided later in this chapter. Here, *Presence* refers to the subjective feeling of being in a VE. This concept can be further divided into dimensions as defined by Regenbrecht and Schubert [224]: Spatial *Presence* (the feeling of “being there”), Involvement, and Experienced Realism, which together reflect users’ perceptions of coherence, realism, and their degree of immersion in a VE. This framework aligns well with the two main dimensions of presence highlighted here: Being There and Perceived Realism [316].

Last but not least, I prefer to differentiate between the term VE, which I use to refer to software-based immersive environments where users (single or multiple) interact within a synthetic world [30, 75, 147], and the term MR, which will be used to describe (hardware) systems, technologies or devices that process and render these environments. The term VLE will be primarily used when discussing VEs specifically designed to enhance the educational experience by providing interactive and accessible learning spaces for students and educators [65].

2.1.4 *Technical Overview*

This section provides a technical overview of the fundamental components and tracking methods used in MR systems, specifically focusing on Head-Mounted Displays (HMDs), tracking techniques, input devices, fidelity considerations, and user interaction components.

2.1.4.1 *Display Systems*

HMDs are widely used in MR applications and play a central role in the experiments of this thesis. As summarized by Dörner et al. [69], HMDs are positioned directly in front of the user's eyes and often feature integrated tracking systems that adjust camera angles and positions based on orientation and location, enabling precise spatial positioning. HMDs can be categorized into two types: closed HMDs, which isolate the user from the surrounding environment to create fully immersive experiences—typically using LCD or OLED displays—and see-through HMDs that allow the user to view the real environment with overlaid digital content. See-through HMDs are further divided into Optical See-Through HMDs, which overlay digital information transparently in the field of view, and VST HMDs, which capture the real world via cameras, integrate digital elements, and display the combined view to the user.

The immersive experience in VEs can be categorized into three types of MR systems [163]:

NON-IMMERSIVE SYSTEMS Users experience a VE on a screen, using conventional monitors or displays. These systems do not fully engage the user's senses and do not create the feeling of physically being in the virtual space. Classic examples include video games or MR applications on smartphones and tablets.

SEMI-IMMERSIVE SYSTEMS In semi-immersive systems, users wear headsets or engage with environments augmented by projection-based technologies, creating a more captivating experience. While users remain cognitively connected to the physical world, they interact with virtual objects that coexist alongside real ones. Classic examples include multi-sided displays such as CAVE (Cave Automatic Virtual Environment), which, however, have certain limitations regarding depth perception and full sensory engagement.

FULLY-IMMERSIVE SYSTEMS Users are fully immersed in a VE, experiencing a high level of sensory interaction, often through a combination of various stimuli (visual, auditory, and haptic). Typically, closed HMDs are used to completely block out the real world, combined with interaction capabilities facilitated by tracking systems that monitor user movements and gestures, enabling natural interaction with virtual objects.

A common issue that arises with immersive MR systems is known as cybersickness. This phenomenon, extensively researched, occurs when the sensory signals received by the brain do not align [69]. This creates a conflict between visual input (perceived movement) and vestibular (inner ear) sensations, which can lead to eye strain, nausea, and dizziness. Beyond visual challenges such as cybersickness, MR systems also explore other sensory outputs to enrich the user experience. As indicated in previous sections, MR systems are by no means limited to visual displays targeting the human visual system. Output devices in the MR context can also engage haptic, auditory, and olfactory senses [147]. Multisensory MR systems can positively impact the overall experience [174], though they have less effect on *Presence* [84].

2.1.4.2 *Tracking*

Tracking in an MR system is crucial for detecting a user's movements and position. This process typically involves pinpointing an object's location within a coordinate system and defining how it relates spatially to the user or a camera [69]. Various approaches can be used to achieve tracking, including acoustic tracking, which measures differences in sound waves; magnetic field-based tracking, which generates artificial magnetic fields to determine location; inertial tracking, which uses accelerometers; laser-based tracking, which measures distances using lasers; satellite-based tracking, such as GPS; and camera-based or optical tracking, which relies on visual information captured by cameras [69]. Each method offers specific advantages, depending on the requirements of the application. Especially in HMDs, two primary types of position tracking are distinguished [69]:

INSIDE-OUT TRACKING In this approach, the sensors or cameras are located on the device itself, facing outward to determine the position of the HMD in relation to the physical environment. This can be achieved either by using the natural features of the physical environment or by placing specific markers in the physical environment to track (marker-based tracking). These systems continuously send data to the environment, making their accuracy dependent on visual information quality and the available processing power. Often, these technologies function as standalone devices, such as the Meta Quest[3], with the main advantage being that they require no complex setup, allowing for greater freedom of movement for the user.

3 Meta Platforms, Inc.: https://www.meta.com/quest/

OUTSIDE-IN TRACKING To monitor the position of the HMD and its input devices, external sensors and cameras can also be used. These track markers on the devices (e.g., infrared LEDs on the HMD and controllers) to determine their position in space. This setup requires a more complex installation, as multiple cameras (Lighthouses) need to be distributed throughout the physical space to achieve higher accuracy. Outside-in tracking generally provides greater accuracy and lower latency than inside-out tracking, especially when multiple cameras are used, making it well-suited for applications requiring precise motion capture.

2.1.4.3 *Input Devices*

Commercial MR systems, primarily used in the experiments of this thesis, are typically supplied with controllers, which remain the standard for interaction within VEs [88]. These controllers enable users to interact with virtual interfaces through buttons and triggers while providing haptic feedback (such as vibrations) to enhance tactile experience. A more natural form of interaction can be achieved through optical hand and finger tracking. Using cameras or sensors—often mounted directly on an HMD [313]—the system captures hand positions and gestures in real time, allowing direct interaction with virtual objects. This natural interaction can positively impact UX and *Presence*, though the absence of haptic feedback may negatively affect performance [5]. Hand tracking is increasingly preferred over controllers and has been found to provide a greater sense of value, although the level of *Presence* depends on the task type [308]. For example, grasping tasks elicit more *Presence* and realism than typing tasks. Another option for interaction is wearable data gloves, which can also capture detailed hand and finger movements and provide tactile feedback. However, data gloves require additional setup, can be bulky, and may limit user mobility and natural hand movements, making them less comfortable for extended use. Additionally, physical objects can be equipped with markers or sensors, allowing the system (mostly using the same tracking technology as controllers and HMDs) to track their position and orientation in virtual space. Taylor et al. further suggest using attachable trackers for real objects and introduce approaches for tracking deformable objects, including the use of neural networks to enhance object tracking in MR [295].

2.1.4.4 *Interaction and Interface*

Interaction in MR systems is defined as the extent of a user's ability to modify the form and content of the MR experience [289]. This culminates in the design of a User Interface (UI) that allows users to operate the system effectively. In MR systems, users generally have three main ways to interact with the system [69]:

- *Selection*: The identification of a specific point, area, or volume in the environment (e.g., to place an object) or the selection of a meaningful subset within the environment (e.g., a virtual object or sub-object for manipulation).
- *Manipulation*: The interactive adjustment of parameters that characterize a virtual object, such as its position, orientation, size, shape, weight, velocity, or appearance (including attributes like color, texture, or shading).
- *Navigation*: In the real world, this involves orienting oneself by determining one's position, planning a route to a target location, and performing the necessary actions to reach it. In MR systems, depending on the level of virtuality, navigation can be achieved through locomotion techniques like traveling, walking, or teleportation.

Providing intuitive tools for selection, manipulation, and navigation enables MR interfaces to support natural interaction with virtual elements. UI design in MR systems plays a crucial role in shaping the UX and supporting effective interaction within immersive environments. As traditional Two-dimensional (2D) interface elements like windows and menus may not translate well to 3D environments, MR systems require innovative UI components and interaction methods [147]. Designing effective UIs for VEs thus presents unique challenges, including the need for intuitive interaction techniques, efficient use of spatial 3D input devices, and the integration of tracking systems.

2.1.4.5 *Fidelity*

In the context of this thesis, I would like to emphasize the importance of fidelity in VEs. The concept of fidelity in MR refers to how accurately and effectively a VE replicates the real world [8, 128]. Realism is achieved through the integration of multi-sensory stimuli, encompassing visual, auditory, tactile, and agency cues [128]. Fidelity can encompass various dimensions, including visual fidelity, interaction fidelity, and overall realism. Research has demonstrated that these aspects of the realism of MR can influence UX [95], perception [326], and the *Sense of Presence* [128].

Bonfert et al. recently introduced the IntFi model, which identifies eight distinct aspects of interaction fidelity in MR [33]. This model expands the traditional definition of fidelity beyond simply simulating reality (realism) and thus distinguishes fidelity from realism. It aims to cover the entire user interaction process with the VE—from user input, through system processing, to the user-perceived output. The level of fidelity is described on a spectrum, ranging from low to medium, high, and maximum fidelity. To illustrate this concept, Figure 2 has been created, placing findings from experiments in this thesis within this spectrum. Fidelity has a positive effect on user performance, UX, and *Presence* [36, 234], which is why this effect is specifically examined in Chapter 4 within the context of medical simulations. Gonçalves et al. [95] most recently presented a review article on studies investigating the impact of fidelity in MR. They determined, that in general, the fidelity positively influences UX. For this thesis, I will differentiate between VF and "interaction fidelity"—or "IMs"—while using "fidelity" as a generic term encompassing both input and output media. This definition covers interactions in virtual environments, including the rendering of user interactions and system responses.

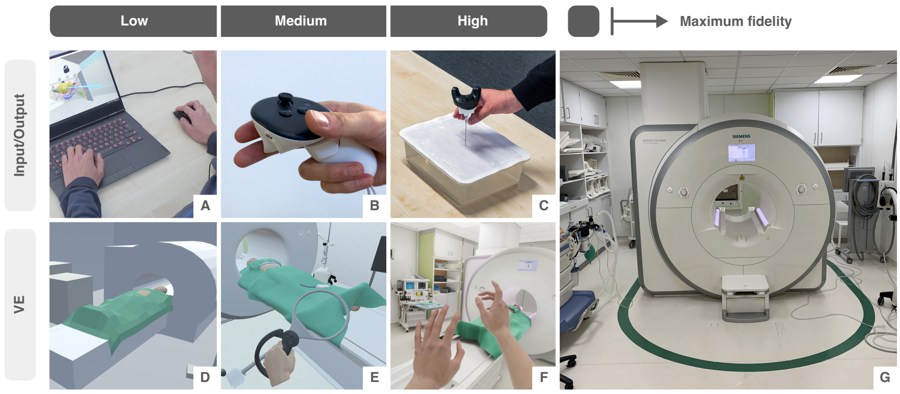


Figure 2: Fidelity spectrum of MR systems as proposed by Bonfert [33], covering low, medium, high, and maximum fidelity levels. Licensed under CC BY 4.0. The structure of this illustration is adapted from Bonfert's model and enhanced with example images from experiments in this thesis. The spectrum is divided into Input/Output and corresponding VE categories. Examples include images from: mouse/keyboard input (A) from Chapter 7, controller input (B) from Chapter 6, and hand tracking combined with tactile tracked objects (C) from Chapter 4. The lower images (D–G) showcase various fidelity levels of an intervention room as presented in Chapter 4, including an image of a real intervention room to illustrate maximum fidelity.

2.2 EDUCATIONAL BACKGROUND

Learning can be broadly defined as the process through which individuals acquire a relatively lasting change in behavior or knowledge through experience or practice [140]. This section provides an overview of foundational learning theories and concepts, emphasizing their relevance for MR systems. By bridging theoretical understanding with practical educational strategies, it also explores how VLEs uniquely contribute to the educational experience.

2.2.1 *Learning Theories & Concepts*

A variety of learning theories address essential aspects of pedagogy, such as instructional methods, knowledge transfer, and motivation. These theories are often combined in practice to create enriched learning environments.

- *Instructional theory* posits that learning is supported not only through individual discovery but also through targeted guidance and structured teaching methods [263]. This approach implies the facilitation of learning by a teacher who structures the learning process, enabling learners to progressively build an understanding of the content. The theory still emphasizes the essential role of active participation and interaction, as these elements foster comprehension. A learner-centered approach is also highlighted, which includes regular feedback and the adaptation of teaching methods to meet learners' needs, as well as individualized pacing to support optimal learning [225].

- *Cognitivism*, on the other hand, is a learning theory that focuses on internal mental processes rather than solely on observable behaviors [77]. Within this framework, the *Cognitive Load Theory* complements *Cognitivism* by combining principles such as activating prior knowledge, embedding learning in a relevant context, and using visualizations to form analogies. In contrast, *Behaviorism* is a learning theory centered on observable behaviors, positing that all behaviors are learned through interaction with the environment [263].

- *Constructivist theory*, which emphasizes that learners build knowledge through hands-on experience and reflection, plays a crucial role in promoting active engagement and individualized exploration [263]. This approach highlights the value of problem-based learning to foster curiosity, customized support during early learning stages, and collaborative learning to enrich the educational journey in MR [63, 109, 114].

As discussed by de Freitas et al., the constructivist approach in VLE illustrates a shift from traditional teaching methods to immersive, exploratory, and learner-centered experiences. This shift reinforces constructivist principles by focusing on the need for tailored guidance and the advantages of collaborative learning [63, 114]. A study by Aiello et al. supports the constructivist approach to learning with MR systems, emphasizing that learning is an active, situational process in which learners build knowledge through personal experiences and interactions within their environment. VEs facilitate immersion, which in turn fosters cognitive development and knowledge construction through direct engagement and sensorimotor interaction [6].

2.2.2 *From Theory to Practice*

As we move from learning theories to practical applications, it is important to recognize the role of pedagogy and didactics. Pedagogy is the broader field that encompasses theories, principles and practices of education and focuses on how best to support learning and development [9, 315]. Didactics is a branch of pedagogy that deals with the methods and principles of teaching. It deals with the planning, implementation and evaluation of educational processes [321]. Didactics involves deciding which teaching methods, activities and resources best support the learning objectives in a particular context [110].

Blended Learning—also known as hybrid learning—combines traditional in-person instruction with online learning components [96]. This approach integrates various modalities of face-to-face and online materials to create a more flexible and personalized learning environment, which can enhance the overall learning experience. Students can participate in both synchronous (real-time) and asynchronous (self-paced) learning activities, allowing for greater adaptability.

According to Kaminska [132], MR in education offers numerous advantages, including enhanced visualization, inclusivity, and virtually unlimited access to information. By fostering engagement, cooperation, and self-directed learning, MR supports an approach that encourages both collaboration and individual knowledge-building. MR systems can further enhance learning by immersing learners in interactive, cognitively engaging VE that support experiential and collaborative learning. They provide immediate feedback to reinforce understanding, facilitating not only a deeper grasp of complex concepts but also enriching the overall learning experience [242].

By reviewing research spanning over two decades, Dalgarno and Lee investigate the learning affordances of VLEs, focusing on unique characteristics such as representational fidelity and learner interactivity (see Figure 3) and how these aspects enhance learning compared to traditional approaches [60]. They identify five key advantages:

- *Enhanced Spatial Knowledge Representation*: 3D VLEs enable learners to develop a better spatial understanding of the domain under study.
- *Experiential Learning*: 3D VLEs facilitate realistic, immersive experiences that are difficult or impossible to replicate in the real world.
- *Increased Motivation and Engagement*: The interactivity and immersive qualities of 3D VLEs lead to higher intrinsic motivation.

- *Improved Knowledge Transfer*: Learning within a realistic 3D context aids in transferring knowledge and skills to real-world applications.
- *Richer Collaborative Learning*: Multi-user 3D VLEs support more effective collaboration, as learners can interact in shared spaces and work together on tasks.

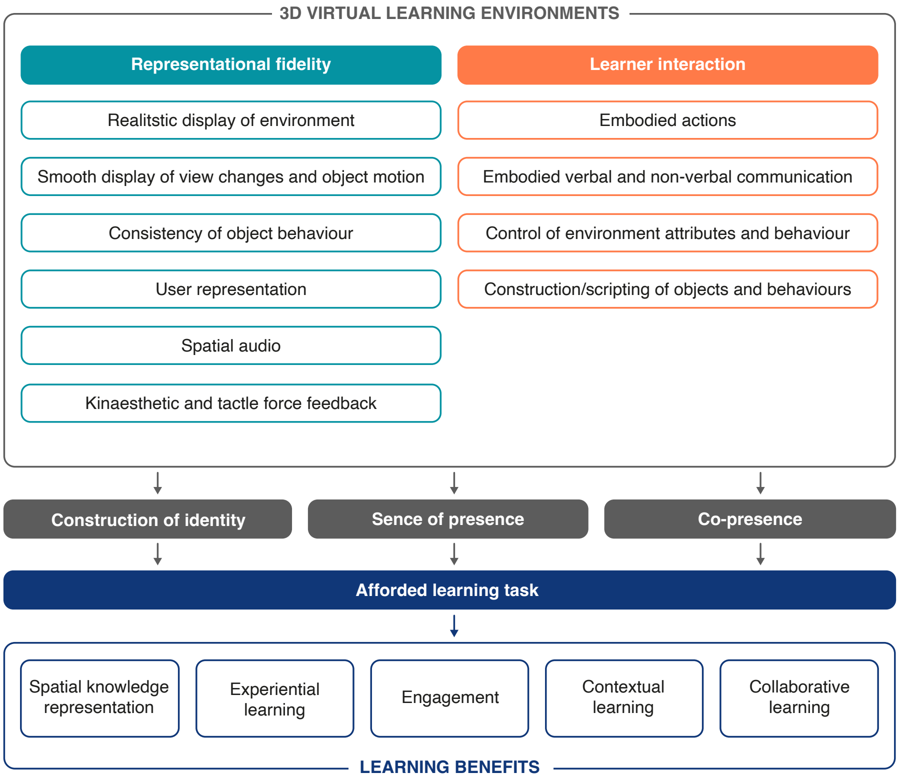


Figure 3: Adapted model based on Dalgarno and Lee [60], illustrating the potential and anticipated learning outcomes achievable in VLEs. Copyright Wiley. Used for non-commercial academic purposes.

2.3 MEDICAL EDUCATION

The previous section has discussed learning theories and the advantages VLEs offer as a supplementary educational medium. The following section will focus on the primary application area of the approaches developed in this thesis: the use of MR medical education and training.

2.3.1 *General Advantages*

Medical education has significantly evolved beyond the traditional "see one, do one, teach one" approach [325]. Today, it incorporates a range of teaching formats, including lectures, seminars, workshops, hands-on exercises, and internships. This blend, enhanced by digital media, promotes solution-oriented learning methods and fosters deeper engagement [135].

With the rise of new educational tools, traditional resources are now being augmented to enrich learning experiences [203]. Since the COVID-19 pandemic, students have gained access to online videos, virtual workshops, and various learning apps, further expanding educational possibilities [197]. The increasing availability of MR technologies has sparked widespread interest in their educational applications, extending well beyond medicine to a variety of fields [236]. This trend has led to greater integration of MR within medical education, highlighting its potential to transform learning [47, 132, 211, 294].

The advantages of MR in medical education are compelling: procedures can be simulated realistically and repeated as often as needed, without risk to actual patients. Compared to traditional resources, MR technologies also improve accessibility and can achieve this with relatively lower financial investment [23].

MR systems provide intuitive interaction, a tangible sense of spatial understanding, and a strong feeling of presence and immersion, all of which contribute to increased motivation and enhanced learning outcomes [109, 213].

2.3.2 *Anatomy Education*

While this work aims to cover a broad range of applications of MR in medical education and training, with emphasis on both individual and collaborative learning approaches, a specific focus is placed on applications for anatomy education. As such, the following section delves into the use of MR in supporting anatomical education, examining both the fidelity of these environments and their unique educational benefits.

MR-based training not only offers the opportunity to train practical procedures but can also be used to teach theoretical principles in the early stages of medical education, such as basic anatomy knowledge. Anatomy education aims to equip medical students with a comprehensive understanding of the morphology, function, location, and spatial relationships of anatomical structures, serving as a critical foundation for understanding diseases and their treatment [214]. In the field of anatomy education, this relates to *regional anatomy* that focuses on parts of the body and *surgical anatomy*, the application and study of anatomy to avoid complications and guide surgeons during interventions [213].

Traditionally, this education involves lectures, textbooks and atlases, and the dissection of human bodies provided by donors [213]. The dissection course, in particular, is invaluable as an active learning approach, reinforcing topographical knowledge and manual skills while helping students comprehend the connections between anatomical structures [38]. To effectively convey this complex information, new teaching methods are required that integrate these various educational aspects cohesively [125].

Due to numerous advantages, such as the capacity for realistic simulation, the ability to repeat procedures, and fewer ethical concerns, MR technologies are especially promising for anatomy education [294]. Where traditional teaching methods may be limited, immersive technologies provide an effective means for grasping complex spatial structures. Interactive 3D visualizations help students develop mental models of intricate anatomical regions, enhancing their spatial understanding. The high levels of immersion and interactivity offered by VE contribute to this deeper comprehension, supporting a new and improved approach to learning in anatomy education [213].

2.3.3 *Application-Specific Use Cases in this Thesis*

This thesis focuses on two key areas of anatomical education at different stages of medical training, with tailored MR applications developed for each. The first area, relevant in the early stages of medical studies, is embryology—particularly the study of human development in the womb, which is fundamental for understanding both normal anatomical structures and congenital diseases. Embryology poses unique challenges due to complex processes like embryogenesis and organogenesis, as well as the rapid morphological changes in structures such as the embryonic heart, involving simultaneous growth and dynamic 3D shape changes over a brief period.

The second area, liver anatomy, becomes crucial in advanced anatomy education for specializations in fields like liver surgery. Liver surgery is a highly specialized and complex discipline typically performed at dedicated clinical centers. Training in this area is demanding, not only because of the intricacies of the surgical procedures but also due to the complexity of the associated disease patterns.

An introduction to specific applications within these distinct medical fields covered in this thesis, as well as an overview of related work and educational methods in anatomy, is provided in Chapter 3.

2.3.4 *Requirements for implementation*

As demonstrated in the previous section, there is a growing trend in the use of MR within medical education, leading to substantial improvements in learner engagement and performance [296]. However, effectively implementing these systems—while addressing the needs of both learners and educators—presents significant challenges [291].

Given the broad range of applications and the diversity of MR technologies, this thesis specifically focuses on HMD-based MR environments.

Recently, Pedram et al. presented requirements for the efficient integration of HMD-based MR systems, aimed at enhancing learning outcomes and supporting skill acquisition among medical students [204, 205]. These requirements are broadly categorized into three main areas:

- *Design Considerations*: The layout and functionality of the MR simulation environment should create an immersive and realistic experience that aligns with specific learning goals, particularly for medical scenarios.
- *Learning Mechanisms*: These include interactive components, feedback loops, and assessment tools embedded as fundamental educational elements within the VE to ensure effective learning outcomes.
- *Implementation Considerations*: This includes the practical aspects of integrating MR, focusing on technical requirements, hardware compatibility, accessibility, and ensuring a seamless fit with existing educational programs.

Figure 4 provides an overview of these categories. These requirements underscore the importance of a user needs-driven validation framework to ensure long-term effectiveness and user acceptance.

2.4 METHODOLOGY BACKGROUND

This section provides an overview of the methodological approach used throughout this thesis, highlighting the foundational principles and processes that guided this thesis.

2.4.1 *Human-Centered Design*

This work is rooted in the field of HCI—the study and exploration of how people interact with computer technology [67]. MR technologies form a part of this field, representing some of the more advanced technologies that are not (yet) widely integrated into everyday life. Familiar technological examples include interactions via mouse and keyboard, smartphones, and household appliances.

A key component of the approach in this thesis is the discipline firmly embedded in HCI: HCD. According to ISO 9241-210:2019 [116], the HCD process is described as an "approach to systems design and development that aims to make interactive systems more usable by focusing on the use of the system and applying human factors/ergonomics and *Usability* knowledge and techniques."

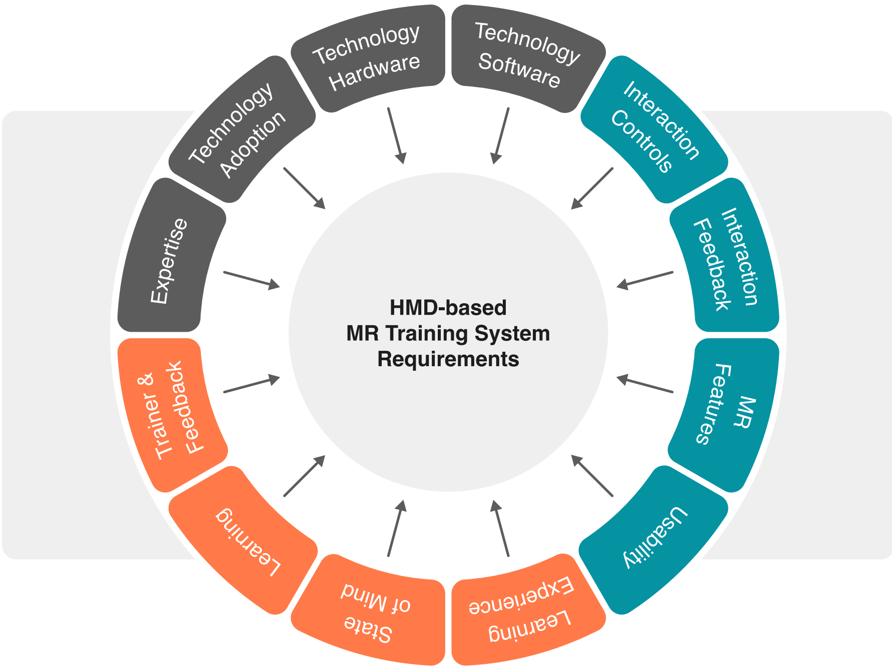


Figure 4: Framework adapted from Pedram et al. [204], illustrating requirement statements grouped into three categories: ■ Design Considerations—focused on the MR simulation and synthetic environment design, ■ Learning Mechanisms—centering on the design of learning experiences within VE, and ■ Implementation Considerations—factors influencing the practical implementation of the design. Licensed under CC BY 4.0.

2.4.1.1 *Usability*

Usability is a core aspect of the HCD process and is standardized under ISO 9241-11:2018 as the degree to which a specific product can be used by designated users in a particular context to achieve defined goals effectively, efficiently, and satisfactorily [115]. The latter components are intended to assess a system's quality during use. *Effectiveness* refers to the extent to which users achieve specific tasks and goals. This is further divided into components: accuracy, which indicates how closely the actual results align with the intended outcomes, and completeness, which reflects the extent to which users attain all intended results. *Efficiency* is defined as the resources utilized relative to the results achieved, depending on the *Usability* objectives. *Satisfaction* is defined as the degree to which the physical, cognitive, and emotional responses of the user, resulting from the use of a system, product, or service, align with the user's needs and expectations.

2.4.1.2 *User Experience*

A fundamental aspect of *Usability* is the UX—the way people perceive the system before, during, and after use. To clarify the distinction between *Usability* and UX: a goal can be achieved effectively, efficiently, and satisfactorily, yet the experience along the way may evoke positive or negative reactions, ultimately impacting overall perception. Thus, a highly usable solution does not automatically ensure a positive UX, as the quality of the journey throughout the process is essential.

Achieving a high level of *Presence* is associated with an overall "better" experience in VEs [244]. Studies suggest that *Usability* and UX correlate with *Presence*, and that UX can be significantly shaped by the degree of *Presence* felt, as the design and immersive qualities of the MR systems contribute substantially to both *Presence* and UX in VE [37, 44]. Strengthening these factors in context of VLEs is essential for enriching the learning experience and supports long-term benefits, including improved learning effectiveness, motivation, and the accessibility of MR applications [35, 60].

2.4.1.3 *Approach in this Thesis*

According to ISO 9241-210:2019 [116], the HCD process is inherently iterative, involving repeated cycles of design and testing for refinement, with user feedback integrated at each stage. By aligning user needs with *Usability* goals, this approach enables continuous adaptation and improvement, ultimately enhancing the UX. This iterative approach is foundational in this thesis and is particularly applied in Chapter 5 and Chapter 6, where design thinking principles [42, 58] guide the development of systems tailored to user needs—primarily, the students engaging with these tools.

Through the integration of learner-centered design principles, this work aims to ensure that learners are not only the primary users but also active participants in shaping their learning environments. This involvement fosters engagement and better aligns the tools with the needs and learning behaviors of students [146, 216].

Furthermore, participatory methods are employed, directly involving users in the development process [31, 117]. This thesis adopts an interdisciplinary approach by collaborating closely with medical experts, who provide valuable insights to ensure the systems created are optimally designed for their intended users. From a teacher-centered perspective, educators are empowered as co-designers, enabling them to shape teaching tools in alignment with their pedagogical goals [90].

2.4.2 *Measurement Tools*

According to Rogers et al., "Data can be numbers, words, measurements, descriptions, comments, photos, sketches, films, videos, or almost anything that is useful for understanding a particular design, stakeholders' goals, and people's behavior. Data can be quantitative or qualitative" [235]. To evaluate users' perceptions, interactions, and responses within MR systems, a variety of measurement tools were employed in this thesis. In this thesis, a mixed-methods approach—i.e., the integration of qualitative and quantitative data collection—is used. As a form of triangulation, this approach investigates a phenomenon from multiple perspectives to enhance the robustness and validity of the research findings [235]. The combination of methods aims to provide a deeper and more precise analysis of the research object by ensuring both statistical generalizability and content depth [288].

2.4.2.1 *Qualitative Measures*

Qualitative methods are employed to capture users' subjective impressions and to gain insights into how they perceive and interact with MR systems, contributing to a more holistic understanding of the effectiveness of the developed MR system and identifying areas for improvement.

INTERVIEWS Interviews are an effective method for gathering feedback on an application. The four primary types of interviews are unstructured (or open-ended), structured, semi-structured, and group interviews [235]. In this thesis, semi-structured interviews and group interviews (focus groups) are primarily used. In semi-structured interviews, the interviewer follows a script that serves as a guide and includes a mix of open- and closed-ended questions. A focus group is a small group of individuals who come together to answer questions and discuss a specific topic in a moderated setting. This method was used in this thesis when experts (instructors) were involved in the evaluation. Focus groups are particularly useful for exploring shared issues, as the mix of different participants encourages individuals to contribute their unique perspectives [235].

SELF-REPORTING A variety of methods can be used to understand users' perceptions of their experience with a technology, where this understanding depends on users' conscious impressions [71]. Self-reporting methods, such as the Think Aloud protocol (TAP) [285], allow users to continuously verbalize their thoughts while interacting with a system. This approach provides insight into users' cognitive processes, helping to understand their experiences, challenges, and thought patterns in real-time. Following the evaluation, open-ended feedback is collected as part of a qualitative self-assessment method, allowing users to reflect on their experiences and offer further insights and suggestions for improvement.

2.4.2.2 *Quantitative Measures*

The following sections provide a detailed description of each measurement instrument, explaining the rationale for its selection, the constructs it measures, and its relevance to the research context in this thesis. Quantitative instruments include standardized questionnaires and assessments that offer objective insights, capturing specific aspects of the overall experience with the proposed MR systems in this thesis. Each test was chosen based on its relevance to the context and objectives of each experiment. Custom rating scales tailored to individual experiments (e.g., individual preferences), custom questionnaires (for knowledge assessment), and demographic data collected are detailed in the respective chapters or included in Chapter 9. Various statistical tests, individually selected for each experiment, were also conducted and are presented in the respective chapters. Where an official (validated) German version of a questionnaire was available, it was used; otherwise, the English version was employed.

SENSE OF PRESENCE In recent years, an increasing number of tools have been developed to measure the *Sense of Presence* [300], encompassing both objective and subjective assessments of its various dimensions [24]. In this thesis, the Igroup Presence Questionnaire (IPQ) was used to estimate presence-related qualia, selected for its high reliability and suitability for user studies within a reasonable time frame [265]. The IPQ, developed by Schubert et al. [262], is a validated tool designed to quantify the *Sense of Presence* in VEs. The questionnaire consists of 14 items rated on a 7-point Likert scale and includes three primary subscales—Spatial Presence (the sense of physically "being there" within the VE), Involvement (the degree of engagement and focus within the VE), and Experienced Realism (the extent to which the environment feels authentic)—as well as an overarching item called General Presence, with each subscale considered an independent factor.

CO-PRESENCE & SOCIAL INTERACTION *Social Presence* supports group learning, promotes group cohesion, and enhances interpersonal communication, making the learning experience more engaging, motivating, and supportive, which are crucial elements for an effective learning environment [136]. Studies also suggest a strong correlation between *Social Presence*, CP, and satisfaction in VLEs, which can positively impact students' satisfaction with their experiences [43].

In this thesis, MR systems are evaluated to compare interactions between multiple users (student-student, lecturer-student) and between users and a virtual guide. To examine the effects on additional factors (e.g. UX, Usabilty, ...), surveys on CP and Social Interaction (SI) within the applications were conducted. CP and SI were assessed through a custom questionnaire inspired by Poeschl et al. [207], who adapted items from Biocca et al. [28]. Responses were recorded on a 7-point Likert scale (1 = Strongly Disagree, 4 = Neutral, 7 = Strongly Agree).

The questionnaire included two main areas: (a) SI, with four items such as "I had the feeling of interacting with other human beings," "I felt connected to the other people," "I felt able to interact with people in the virtual room," and "I had the impression that the audience noticed me in the virtual room"; and (b) (co-)presence of others, with three items: "I was aware that other people were with me in the virtual room," "I felt that I perceived other people in the virtual room," and "I felt alone in the virtual environment."

IMMERSIVE TENDENCIES To account for UX differences attributable to individual tendencies toward immersion, the Immersive Tendencies Questionnaire (ITQ) [329] was utilized in this thesis. This psychometric tool assesses individual differences in both the capacity for and inclination towards immersion—essentially, how readily and deeply a person can engage with a media experience while detaching from the real world. Studies indicate that users with stronger immersive tendencies report a greater sense of co-presence, highlighting the importance of capturing these tendencies to foster immersive and socially interactive environments [43]. In the experiments conducted in this thesis, responses were recorded on a 7-point Likert scale (1 = extremely disagree; 4 = neither; 7 = extremely agree) across the subscales of Focus, Game, and Involvement.

TECHNOLOGY ACCEPTANCE In this thesis, the use of novel technologies in new (medical) contexts is explored, making the Technology Acceptance Model (TAM) a fitting choice for evaluating user acceptance. The TAM, developed in 1989 by Fred D. Davis [61], was originally intended to predict and explain technology acceptance in the workplace. Today, it is regarded as one of the leading models for predicting and explaining user acceptance [62]. To gain insights into user behavior and attitudes toward the systems developed in this thesis, the TAM was assessed using a 7-point Likert scale (1 = extremely disagree; 4 = neither; 7 = extremely agree). A 12-question variant was used, consisting of 6 questions for Perceived Usefulness (PU) and six questions for Perceived Ease of Use (PEU). These two key factors, PU and PEU, are regarded as additional indicators for assessing the effectiveness of MR systems in an educational context.

AR IMMERSION The Augmented Reality Immersion Questionnaire (ARIQ) is based on the gaming immersion model by Brown and Cairns [39], which includes levels of engagement, engrossment, and total immersion. It was specifically developed to measure immersion in location-based AR settings and is used to assess and compare user immersion experiences in VEs. It has also been applied to measure immersion in MR various systems [299, 333], including educational contexts [241]. This instrument evaluates immersion using a 7-point

Likert scale (1 = completely disagree; 4 = neutral; 7 = completely agree) across 21 items, with the overall immersion score calculated by averaging responses. In this thesis, the questionnaire was primarily used to assess the quality of different concepts and user comfort.

USER EXPERIENCE The User Experience Questionnaire (UEQ) is a comprehensive tool designed to assess user satisfaction and acceptance by capturing users' subjective impressions of a product or system. It provides insights into both *Usability* (pragmatic qualities) and enjoyment and aesthetic appeal (hedonic qualities). In this thesis, the classic (long) and short versions of the UEQ were applied as needed, depending on the context, to cover a broad spectrum of UX qualities and allow for straightforward comparisons across different conditions [145, 259]. The classic UEQ includes six scales—Attractiveness, Perspicuity, Efficiency, Reliability, Stimulation, and Originality—and consists of 26 items rated on a 7-point scale from -3 (negative) to +3 (positive). The short version of the UEQ was also used in contexts requiring a more condensed assessment to form an overall UX score. This version uses a 7-point Likert scale to rate 8 contrasting adjective pairs across dimensions like attractiveness, comprehensibility, efficiency, stimulation, and originality.

GAMING EXPERIENCE To further evaluate user experience and social components, elements of the Game Experience Questionnaire (GEQ) were used in this thesis. The GEQ is widely utilized by game researchers across a broad range of game genres to capture various facets of the gaming experience. Due to its modular structure and coverage of different aspects of player experience, the GEQ is adaptable to various game types and research scenarios and contributes to the validation of real-time indicators for player experiences. This adaptability has also led to its application in the study of MR systems [122, 286, 324]. In this thesis, the *Core*, *Social Presence*, and *Post-game* components of the GEQ were used to evaluate the overall experience of the application [113].

USABILITY SCORE The System Usability Scale (SUS) is a standardized tool introduced by John Brooke in 1996 to asses perceived *Usability* of a system, product, or service, prized for its flexibility and broad applicability. Known for being straightforward, quick to administer, and reliable in delivering a clear *Usability* overview, the SUS has become one of the most widely used post-study questionnaires in industrial *Usability* research [154] and is increasingly applied in educational technology design [307]. The SUS comprises a 10-item questionnaire completed by users after system use, with each item rated on a 5-point Likert scale from "Strongly Disagree" to "Strongly Agree." This tool offers a single, overall *Usability* score, where an average of 68 or above indicates good *Usability* [155]. Specific score ranges provide further insight, with 85+ often regarded as excellent and scores below 50 signaling critical *Usability* issues. The SUS is especially valuable in user-centered design processes, providing a practical benchmark for *Usability* and helping to identify areas for improvement, which is why it was extensively used in this thesis.

TASK LOAD The NASA Task Load Index (N-TLX) is a questionnaire used to assess the subjective mental load or stress associated with performing a specific task. First detailed by Hart and Staveland in 1988 [105], it has become one of the most widely used tools for measuring *Subjective Workload* in human factors research. The N-TLX assesses aspects such as mental demand, physical demand, temporal demand, performance, effort, and frustration, each rated on a scale from 0 to 20. In this thesis, the N-TLX was employed to gauge the subjective perception of task load, particularly in relation to interaction concepts. Analyses focused on general task load indicators, specifically concerning these interaction concepts, which is why the unweighted N-TLX score (Raw TLX) was chosen. This approach offers a straightforward summation of ratings without additional weighting, providing an initial, comprehensive assessment of perceived task load, well-aligned with the focus on overall task load perception rather than specific, weighted insights [104].

MENTAL ROTATION ABILITY Spatial understanding and mental rotation ability are considered predictors of intelligence and are therefore frequently applied as measures in the medical school application process. The Mental Rotation Test (MRT), originally developed by Steven G. Vandenberg and Allan R. Kuse in 1978 [304], is a standardized psychological instrument used to assess mental rotation skills. Since parts of the experiments in this thesis involve understanding spatial relationships (such as organ positioning within the body), this test was selected as a relevant measure. To assess rotational ability, the MRT was used in the variant by Ganis and Kievit [89], which incorporates 3D views and improves the representation of spatial orientation through enhanced shading and depth perception. The test comprises 96 pairs of cube figures, with each pair requiring a decision within 7 seconds on whether the figures are identical or different. An instructional text and 10 sample items are provided at the beginning, and the test takes approximately 10 minutes. Excerpts from this test application are included in Section 9.4.

2.5 CHAPTER SUMMARY

This chapter established the use of the term MR for the experiments presented throughout this thesis and emphasized the primary technologies involved—namely, HMD-based systems—as well as input and output devices, interaction methods within VE, and design features central to this work, framing the fidelity characteristics of MR systems. Based on theoretical foundations, it was demonstrated that integrating MR into medical education holds significant potential for enhancing learning outcomes and practical skills. MR supports improved visualization, engagement, and virtually unlimited access to information while facilitating blended-learning approaches that promote both collaboration and individual knowledge construction. Additionally, the chapter examined the HCD process, highlighting key evaluation methods for MR-based VE and emphasizing the relationship between *Presence* and UX, a theme that will continue throughout this thesis. These foundations set the stage for the next chapter, which focuses on related work on MR for educational purposes.

RELATED WORK

ch. 3

SYNOPSIS *This chapter provides an overview of related work in medical education and medical simulations, focusing on the impact of manipulating visual and interaction fidelity in immersive environments, exploring different didactic approaches such as collaborative settings, and examining the use of MR technologies in anatomy education. It also highlights two distinct focal areas examined in this thesis: the development of the embryonic heart and training for liver surgery.*

ABOUT THIS CHAPTER *Selected passages in this section draw on previously published peer-reviewed work by the author included in the Core publications ([Core1] - [Core7]). These passages have been adapted and incorporated to provide a coherent and comprehensive background for the related work discussed in this chapter. Reuse of content complies with the respective publishers' policies: articles published by IEEE are reused with permission in accordance with IEEE's thesis reuse policy; open access publications under Creative Commons licenses are reused in compliance with CC BY 4.0; the Elsevier article is reused in accordance with Elsevier's thesis reuse policy; and the ACM article is reused in accordance with ACM's open access guidelines.*

3.1 EFFECTS OF MANIPULATING REALISM

As highlighted in Chapter 2, realism describes the degree to which a simulation resembles the reality we perceive daily [128]. The manipulation of visual, auditory, haptic, and olfactory characteristics aims to achieve the highest possible fidelity in a VE, striving for maximum realism. Various studies have demonstrated that fidelity aspects of MR can influence UX [95], perception [326], and *Presence* [128]. Therefore, thi section provides an overview of methods primarily describing visual environmental characteristics, followed by a discussion on the effects of manipulating input modalities.

3.1.1 *The Role of Realism in Virtual Environments*

The design of VEs depends on numerous individual components to achieve a 'realistic'-appearing setting. For example, realistic lighting, along with dynamic shadows and reflections of moving objects, enhances these effects and has a positive impact on VF and, consequently, *Presence* [167, 187, 331].

The effects of geometric realism, i.e., polygon count and texture resolution, were investigated by Hvass et al. [112] in the field of MR video games. Higher degrees of realism evoked higher subjective *Presence* ratings as well as stronger physiological responses.

Newman et al. [191] conducted two experiments to examine the effects of environmental representation on perception. The first study found that MR experiences elicited more positive responses than watching videos, but were less immersive than real-life observations. In their second experiment, they examined the impact of VE realism on stress recovery and found that higher realism enhanced the process and increased the *Sense of Presence*.

In a study on the effects of visual realism in virtual acrophobia therapy, Schmied-Kowarziki and Paelke [248] observed that individuals without acrophobia reported increased *Presence* in more realistically rendered environments. However, the influence of environmental realism on individuals with acrophobia was less pronounced, leading to the conclusion that less expensive prototypes might be adequate for therapeutic purposes.

A different approach was taken by Ragan et al. [218], who explored the impact of visual complexity on performance during a scanning task. Their findings indicated that task performance deteriorated as visual realism increased. Despite this, they advocate for higher levels of visual complexity, arguing that it is necessary for the effective transfer of learned strategies to real-world scenarios.

A comparative study by Mizuho et al. [181] investigated the effects of switching between virtual and real environments on memory. It was found that the visual quality of VEs had no impact on context-dependent forgetting and source-monitoring errors. Additionally, the study found that a high-fidelity VE significantly enhanced the feeling of presence.

Gonçalves et al. [95] most recently presented a review article on studies investigating the impact of realism in MR. They determined, that in general, realism positively influences UX. They also categorized other research results in several independent variables that have been manipulated to achieve different fidelity levels. The resulting categories were avatar visual, environment visual, audio, haptic, and olfactory content variables, as well as audio, haptic, interaction, camera, lights, and physics system variables.

3.1.2 *Advanced Input and Feedback*

MR experiences are also influenced by the realistic nature and availability of input modalities. For instance, controllers are commonly used for interaction but lack flexibility and tactile feedback, making their use less realistic. Another potential way to improve the realism of simulations is to represent interactive virtual objects with tangible physical mock-ups, e.g., by 3D-printing them and using trackers to determine their position and rotation in MR space [295]. Hand tracking can also lead to a more positive perception of the experience [308]. The combination of MR visualization and tangible physical objects was shown to be more advantageous with respect to realism and enjoyment compared to either method alone [180]. Different senses can also be addressed by multiple modalities in MR [231]. McMahan et al. [171] compared a natural interaction technique using tracked hand-held devices to traditional mouse and keyboard input in a MR game and additionally varied between a stereoscopic 360° Cave Automatic Virtual Environment (CAVE) display and a monoscopic single-wall display of that CAVE. They showed that the combination of low visual and low interaction fidelity performed comparably to the combination of high visual and high interaction fidelity, and that both outperformed the other two factor level combinations. Natural and more realistic interaction paradigms were also shown to be beneficial for learning technical skills in virtual training compared to gamepad input in MR and mouse and keyboard input on a monitor [127]. However, the study showed that the latter was better for procedural knowledge transfer.

In the field of prototype *Usability* testing, Zhou and Rau [337] compared a tangible physical mock-up of the prototype to a condition without haptic feedback. Visual output was created using either immersive MR or a monitor. The physical mock-up could improve performance, and MR generally evoked more positive subjective feedback. However, in the MR condition, using the tangible object did not improve *Involvement*.

In the medical field, Plümer et al. explored the potential of MR for prototyping applications aimed at enhancing visual representations for controlling a mobile medical imaging robot [106]. The study examined two visual techniques—visualizing hidden processes and controlling a virtual surrogate—to mitigate the perceived latency between user input and robot activation. They demonstrated that MR prototyping could achieve relative validity and provided insights into how different visualization techniques can impact interaction efficiency.

In interventional use cases, Van Nguyen et al. [303] present a training system for performing biopsies in MR. Snarby et al. [284] presented a system that enables training of medical procedures by incorporating real-time image data. The use of immersive MR in training environments can also improve the preparation of assisting personnel during surgeries [98]. Efforts also extend to enhancing patient comfort by using MR to reduce anxiety and claustrophobia during Magnetic Resonance Imaging (MRI) scans, simulating the experience and educating patients to decrease cancellations due to discomfort [41]. The results of a study by Nakarada-Kordic et al. [189] suggest that a VE has the potential to improve patients' experience and prepare them for examinations compared with a simulated MRI scan. They also emphasize the benefits of MR in MRI examinations, as it can be a cost-effective and low-risk tool for knowledge transfer to both patients and physicians. However, this study does not use a mock-up or tangible objects in their VE. The *Sense of Presence* could be improved by including tactile feedback from furniture and interactive elements. In addition, the study does not compare the effects of different level of VF and does not provide a fully immersive simulation.

3.2 IMMERSIVE LEARNING EXPERIENCES IN GENERAL

Immersive learning experiences and the application of Mixed Reality (MR) in education are instrumental in increasing student engagement and retention. De Freitas et al. [63] propose an evaluation methodology to support the development of learning activities within VLEs. They argue that increasing interactivity in virtual and hybrid spaces not only fosters new experiential learning opportunities but also cultivates complex social interactions, thereby enhancing learner ownership. The framework for assessing these environments is built around four main components: learner preferences, pedagogy, presentation, and context. These lead to key factors such as learner characteristics, pedagogical approach, environmental interactivity, immersion, and contextual elements, with a central focus on the role of presence in immersion.

Lee et al. delve into the qualitative aspects of VLEs, identifying both key and confounding factors that affect learning outcomes [148]. Their findings emphasize the critical role of the learning experience—shaped by presence, motivation, cognitive benefits, and reflective thinking—in influencing outcomes within an MR-based desktop environment. The study also explores how student characteristics like spatial ability and learning styles moderate these effects. As VLE technology continues to evolve, allowing for fully immersive applications, the challenge remains on how to effectively design and benchmark these environments.

Moreover, the implementation of MR in higher education faces significant challenges due to the need to accommodate the diverse viewpoints and requirements of various stakeholders [124, 232]. This necessitates a careful consideration of design strategies that can meet these varied needs and ensure the effective integration of MR technologies into educational settings.

3.3 ADVANTAGES OF MR IN ANATOMY EDUCATION

The use of MR in medical contexts significantly facilitates the acquisition of anatomical knowledge and the training of medical skills, effectively minimizing risks to patients [195, 213, 214, 231]. Digital tools such as online videos, workshops, and learning apps are already enhancing medical education [203], but MR applications bring a new dimension of immersive simulations that allow for repeated practice and introduce fewer ethical dilemmas than traditional methods [91, 294]. These technologies notably improve the understanding of complex spatial relationships and show positive trends in learner engagement and performance outcomes [102, 296], though their implementation requires careful consideration of both learner and educator needs [291].

In anatomy education specifically, MR supports detailed studies in both regional anatomy, which focuses on specific body parts, and surgical anatomy, which applies anatomical knowledge to guide surgical interventions and prevent complications [213]. This dual application underscores the versatility of MR in handling diverse educational requirements. A comprehensive overview by Rashidian et al. [219] further identifies these branches as crucial for training that integrates cognitive knowledge with psychomotor skills, traditionally acquired through supervised practical sessions in operating rooms or through simulation-based training [48, 196, 198, 306].

Johnson et al. [125] advocate for a shift in medical education from passive, didactic approaches to more interactive and clinically relevant curriculum over the past decade. VLEs are increasingly viewed as favorable tools for anatomy education, serving as viable alternatives to cadaver-based learning and potentially enhancing test scores through more dynamic and engaging instructional techniques [138, 271].

The results of a study by Kadri et al. suggest that the immersive and interactive nature of MR systems can significantly enhance learning experiences and efficiency compared to traditional methods [131]. The VLEs developed in these studies enable interactive 3D exploration of anatomical structures using hand gestures and gaze control, incorporating game-like elements to make learning both effective and engaging. Students using these VLEs not only improved their anatomical vocabulary but also required less time to complete tasks, showcasing the profound impact of MR on modern medical education.

Building on these advancements, simulating medical education through MR holds the potential to further accelerate clinical training [245]. Beyond advanced MR simulators [17], immersive MR has already been evaluated as a powerful tool for learning human anatomy [78]. For instance, Weyhe et al. [322] introduced a virtual 3D anatomy atlas, demonstrating in a user study that their MR application not only enables rapid learning but also achieves higher user satisfaction compared to conventional methods.

In addition to academic research, a number of commercially available solutions for exploring human anatomy are offered by companies specializing in immersive medical technologies. For instance, Meta Quest[1] features applications such as Sharecare[2], which offers real-time simulations of the human body, its organs, and their natural functions, and Anatomy Explorer[3], which provides an interactive anatomical learning experience. Similarly, the Pico platform[4] features

1 Meta Inc.: https://www.meta.com/quest/
2 Sharecare, Inc., https://www.meta.com/en-gb/experiences/pcvr/sharecare-vr-2017/1656800021020362/
3 Virtual Medicine, s.r.o., https://www.meta.com/en-gb/experiences/pcvr/anatomy-explorer-2020/3878356485536748/
4 PICO Immersive Pte. Ltd., https://www.picoxr.com

Human Anatomy VR[5], enabling detailed anatomical exploration. All of these applications are designed as single-user experiences, allowing individual exploration and interaction with controllers.

3.4 COLLABORATIVE LEARNING

Collaborative learning in medical education is supported by various MR systems that allow students and professionals to interact with virtual anatomy in shared environments, either in co-located or remote settings [76, 210, 266]. This section reviews the technological approaches and benefits of integrating collaboration and MR in medical education.

Richardson et al.'s [230] presented a system, which enables multiple students to study gross anatomy together in a shared Second Life environment. While limited to virtual avatars and simple interactions, this approach shows the potential for collaborative anatomy exploration.

Collaborative environments can be co-located or remote, which significantly influences accessibility and communication possibilities. Research shows that when small groups of students explore complex anatomical structures together, learning outcomes improve—whether they interact in the same room or remotely [40, 118, 119, 182]. Typically, one student leads the exploration while others observe, but technical setups for such environments remain costly and are often unavailable in many medical faculties.

For individualized learning, one-on-one virtual teaching has also shown promise. Moorman [184] presented a video-conferencing solution where instructors guide students remotely. In contrast, Saalfeld et al. [240] created a one-on-one tutoring system focused on the human skull base, using a stereoscopic display for the tutor and an HMD for the student within a scaled-up skull model, enhancing spatial learning through immersive MR.

Advanced systems leverage MR to improve hands-on skills. Pedram et al. [205] studied a MR system for medical training, finding significant improvements in clinical skills and safety practices among 44 students, though knowledge retention was comparable to traditional methods. Similarly, Bork et al. [34] demonstrated that MR systems improve students' 3D understanding in anatomy and radiology. In another application, Mehta et al. [173] presented a mobile MR platform allowing multiple users to manipulate patient-specific 3D heart models, though practical surgical planning applications remain underexplored.

Recent innovations also highlight the value of real-time interactive environments. Zhang et al. [334] demonstrated a multi-user setup

5 Virtual Medicine, https://store-global.picoxr.com/de/detail/1/7179344687503785989

that combines 3D autostereoscopic visualizations and gesture interaction, allowing students to compare morphological differences between healthy and diseased states. This system provides a dynamic learning space based on interactive medical images.

To support vascular surgery planning, Wang et al. [311] developed an MR platform enabling surgeons to collaborate synchronously from different locations. This system allows real-time interaction with virtual models and near-instant blood flow simulation feedback, facilitating precise planning for vascular interventions.

Chheang et al. [49–51, 54] have made multiple contributions to collaborative MR systems for medical training. Their work includes a system for liver surgery planning that supports collaborative virtual resections on 3D organ models alongside 2D image slices [51]. Another project addresses laparoscopic procedure planning, integrating laparoscopic input devices and various user roles, further extended with an anesthesia simulation to create an interprofessional training environment [49, 50]. More recently, Chheang et al. [54] explored the use of a generative AI virtual assistant within a VLE for anatomy education. This assistant, available in avatar- and screen-based modes, enabled interactive responses to questions of varying complexity. The study found differences in student performance based on question type and assistant configuration, providing insights into how generative AI can support adaptive, interactive learning in medical education.

In addition to academic research, a number of commercially available solutions for medical training and imaging are offered by companies specializing in immersive and AI-enhanced medical technologies. For example, VRAIn Medical[6] provides an advanced 3D bioimaging platform that enhances the visualization and analysis of medical images, enabling more detailed insights for clinicians. ORamaVR[7] offers tools for educators to create, record, and publish medical XR training simulations, complete with objective metrics and performance analytics, supported by an AI-based co-tutor. Luxsonic[8] delivers solutions such as a virtual reading room for radiologists and a DICOM viewer optimized for the latest HMDs, including Apple's Vision Pro[9].

3.5 LEARNING EMBRYOLOGY

Since Chapter 5 and Chapter 6 focus on MR applications for anatomical education with a specific emphasis on embryonic heart development, this section provides an overview of related work and methodologies in this field.

6 VRAIn Medical, https://vrain-medical.com
7 ORamaVR, https://oramavr.com
8 Luxsonic Technologies Inc., https://luxsonic.ca
9 Apple Inc.: https://www.apple.com/apple-vision-pro/

Falah et al. [79] introduced an interactive VE and 3D visualization system which provides self-directed learning and assessment of adult human cardiac anatomy using semi-immersive stereoscopic displays and projections. A comparative study between the MR heart anatomy system and traditional medical teaching modalities (physical heart model) showed that MR-based learning improved student understanding of heart anatomy by offering an enhanced experience [10]. It also demonstrated the usefulness of the system by showing a higher satisfaction rate regarding structure and visualization. Maresky et al. [166] visualized Computer Tomography (CT) and MRI data of normal and near-normal adult hearts in "The Body VR: Anatomy Viewer" in combination with anatomical models in "Sharecare VR" to create a unique cardiac virtual environment. A pilot study with undergraduate medical students for evaluation demonstrated the viability and the effectiveness of MR in teaching cardiac anatomy. Anderson et al. [13] commented on the pedagogical strengths of Maresky et al.'s study but raised concerns about the anatomical accuracy of the models, highlighting the difficulties in developing virtual learning materials that meet medical standards. The creation and rearrangement of cardiovascular structures in virtual spaces supports spatial understanding of complex anatomy through various interaction methods, such as 3D sketching, deformation, and puzzles [229, 238, 239].

Understanding anatomical relationships in the early stages of formation of the organ system is both challenging and crucial, as it forms the foundation for explaining subsequent pathological conditions. Conventionally, embryonic development is taught with two-dimensional illustrations and physical models. However, embryonic development involves a rapid and simultaneous growth process, characterized by complex changes in both shape and position, within a brief period of time [45]. The formation of organ systems during embryonic development is particularly challenging due to the absence of fixation points for 3D orientation [201]. Especially, understanding the formation of the human heart is difficult to both learn and teach. These deformations are precisely the causes of many pathologies, making it crucial for students and future cardiologists to understand these processes to diagnose and treat congenital heart defects accurately. Conventional digital and analog media cannot fully reflect these processes. For this reason, researchers have been working for years on innovative teaching methods to establish immersive technologies as a supplementary medium in medical education. Various types of technical approaches, ranging from 3D autostereoscopic visualizations with gesture interaction in multi-user settings to mobile applications, attempt to address these challenges. These methods demonstrate that students can benefit from multidimensional representations, thereby improving learning outcomes [99, 334].

There are currently few examples in the literature of how these dynamic transitions can be made tangible for students. However, there are approaches under the increasing use of multimedia content to illustrate stages of development.

Visualizing The Developing Brain[10] is a project of the Center of Anatomy and Human Identification at the University of Dundee to help students better understand the morphological development process and in particular the difficult-to-visualize 3D folding of the early human brain. An interactive animation video and 3D model show the growth of the embryo and, in particular, the brain and its individual components.

There is a wide variety of approaches ranging from physical models of embryonic organ systems or cardiac malformations, created using 3D printing techniques to match the teaching content, in comparison to conventional media and 3D digital representations [45, 283]. Bakker et al. [18] realized a 3D atlas of human development that depicts the temporal development of all organ systems. This atlas allows students to explore different embryo structures using a PDF file with interactive 3D models. This tool lacks in the performance to be used fluently: while it depicts the different stages, it does not show the transition from one stage to another. Hull et al. [111] evaluated the use of a visualization of craniocaudal folding at the beginning of embryogenesis, which represented three time points of development. Buttons could be used to switch between the stages, which were displayed on screen using classical 3D viewers. Individual organ systems could be faded in and out individually. Even though the application provides an interactive understanding of temporal folding formation, this approach lacked the substantial immersion which could be provided by Mixed Reality Approaches.

Gustilo et al. [99] developed an interactive mobile application called *Embryonic Virtual Heart Application* that contains a series of anatomical 3D models of the embryonic heart with and without congenital heart defects which already shows that the students benefit from a 3D Model. Further indications of the general usefulness of an immersive concept were presented by Tait et al. [293] who used a 3D reconstruction of sheep embryos in conjunction with images of the corresponding histological slides, which could be viewed in a mobile application for Android tablets in a classic 3D viewer and in handheld MR. The results of the evaluation suggested that the use of a 3D modality such as the presented MR application significantly improves the understanding of slide alignment compared to current methods. In addition, the application was considered more interesting, useful, and user-friendly than current histology tools.

In the area of integrating new visualization approaches into the curricula in use, two works are particularly noteworthy. First, a case

10 https://visualisingthedevelopingbrain.co.uk

study where Meguid et al. [1] describe how they integrated the *Human Developmental Biology Resource* atlas[11] and the 3D *Atlas of Human Embryology*[12] into the curriculum at Newcastle University and provide a perspective on how these new learning resources might impact teaching in the future. The *Human Developmental Biology Resource* atlas is a database of embryological tissue samples from human embryos donated to research for educational use and includes rendered animations of rotating organs, 3D models of segmented organ systems, and histological embryo sections [93]. Second, Moraes et al. [185] provided students with multimedia materials of embryos, e.g., clinical histories, autopsy images, ultrasound images, movies, and animations. The teaching material was used by the students in class and subsequently accompanied by a knowledge examination and interviews. The multimodal use proved to be useful and was able to reveal knowledge gaps between basic sciences and clinical disciplines for medical students. This demonstrates the potential of a multimodal, immersive learning experience for teaching such a difficult topic.

Immersive technologies are also being explored in clinical approaches to facilitate visualization of complex anatomy and support periprocedural pain management, rehabilitation, and patient education, thereby enhancing the treatment process. This underscores the growing significance of these technologies in both medical education and patient care, emphasizing the need for methodologically robust studies on how to tailor technology to specific educational and clinical needs [160].

3.6 LEARNING LIVER ANATOMY

In Chapter 7, an explorative VE for liver anatomy education is presented; accordingly, this section provides an overview of related applications and approaches designed to address surgery-relevant educational needs, specifically preparing students for clinical scenarios.

A more 'passive' way to perceive medical information is through videos, which are an easily accessible source that can even be used in the operating room without intrusive hardware [243]. When tailored to educational purposes, videos can lead to positive learning outcomes. Nobuoka et al. [192] introduced a system that leverages this approach, using a multi-layer, three-dimensional liver anatomy atlas created by filming a real dissection layer by layer from different anatomical perspectives. This method enables students to replicate procedural steps by studying the recorded images and videos, thereby enhancing their understanding of anatomical structures and the flow of surgical procedures. Building on real surgical footage, Fung et al. [86] integrated 3D animations to develop an online video

11 https://hdbratlas.org/
12 https://www.3dembryoatlas.com

atlas that covers a range of procedures involving the liver, pancreas, and transplant surgery. This resource has been applied to visualize rare and complex cases [215].

More interactive options for liver anatomy education are offered through web-based applications. Furcea et al. [87] describe an e-learning platform that integrates tools for pre-operative planning with laparoscopic liver surgery training. This platform is accessible remotely via a web browser, though it lacks broader support for MR devices and advanced interactive features. When real data sets are included, they are typically based on CT scans as 2D representations, requiring students to mentally convert them into 3D structures—a challenging task for those with limited experience. Crossingham et al.[59] created an interactive website featuring 3D liver model reconstructions, although these models cannot be connected back to their original 2D datasets. This limitation was addressed by Birr et al. [29] with the *LiverAnatomyExplorer*, another web-based tool that combines 2D images, 3D models, surgical videos, and assessment features. A similar tool was developed by Mönch et al. [183], incorporating volumetric models and specialized interaction techniques to allow individual resections, further enhancing the realism of the training experience.

There are further MR solutions dedicated to surgery training, planning, and education. *IMHOTEP*[13] is one such MR framework, consolidating treatment data, multi-modal patient data, 2D images, 3D volumetric models, and 3D surfaces into organized workspaces. Similarly, the *LiverPlanner* [228] supports pre-operative planning for complex liver surgeries within a semi-immersive environment using a stereoscopic large-screen projection system. For input, it utilizes a combination of a tablet and a six Degree of Freedom (DOF) controller. Unlike the approaches in this thesis, these applications focus specifically on preparing and planning individual surgeries

Hack et al. [101] conducted a survey of semi-immersive and immersive anatomy education systems, identifying 38 platforms that use either shutter glasses, passive glasses, or autostereoscopic displays. Their findings highlight several advantages of these systems, particularly in enhancing spatial understanding—benefits that are especially pronounced for complex vascular structures [2], which are highly relevant for liver anatomy.

HMD-based anatomy education systems, designed as 3D puzzles, were introduced by Messier et al. [175] and Pohlandt et al. [209]. Messier et al. compared a 2D monitor, a stereo monitor, and the Oculus Rift, reporting positive initial results, while Pohlandt et al. utilized an HTC Vive. In these systems, students can select from various anatomical structures (such as the skull or foot) and freely scale them. The MR puzzle format also allows students to disassemble a

13 http://imhotep-medical.org/

completed puzzle in a specific sequence, simulating the order of dissection steps commonly followed by medical students.

3.7 CHAPTER SUMMARY

This chapter provided an overview of related work in medical education and simulation, focusing on how immersive technologies enhance anatomical and medical training. Examples illustrated how variations in realism and interactivity influence learning experiences, covering scenarios such as collaborative learning environments and individual exploration. Additionally, relevant applications of MR in anatomy education were discussed, with a focus on two specific areas examined in this thesis—embryonic heart development and liver anatomy—to showcase the transformative potential of MR technologies in medical education. These insights set the stage for the following chapter, which delves deeper into the impact of fidelity levels in MR through the first experiment discussed in this thesis within medical task simulations.

EFFECTS OF VISUAL AND INTERACTION FIDELITY ON MEDICAL TRAINING

ch. 4

SYNOPSIS *This chapter investigates the effects of different fidelity levels in MR, focusing on user performance, UX, and presence within medical task simulations. Specifically, it examines how well these simulations replicate real-world interactions through varied levels of VF and interaction modalities. The presented experiment evaluates these aspects, aiming to provide design guidelines for future MR-based medical training tools, and emphasizes the nuanced effects of fidelity on user engagement and performance. Building on these findings, the chapter concludes with an outlook that outlines plans for refining these methods.*

ABOUT THIS CHAPTER *Parts of this chapter have been published in Schott et al. "Is this the vReal Life? Manipulating Visual Fidelity of Immersive Environments for Medical Task Simulation" [Core5] and have been reused for this thesis. The article is open access and published under a Creative Commons Attribution 4.0 International License. Additionally, some of the methods described in this chapter were developed as part of Ms. Lara Stallmeister's Master's thesis, titled "Exploration von Qualitätsgraden der Visualisierung und Interaktion in Virtual Reality" The master thesis project was supervised by me and F. Heinrich.*

MY CONTRIBUTION *I developed the main research idea and was responsible for the conceptualization and methodological approach, refining the research concept initially formulated in Ms. Stallmeister's Master's project by expanding the theoretical foundations and conducting a comprehensive literature review. Additionally, I co-managed the research project and oversaw the study design. Data collection and analysis were performed collaboratively. I took primary responsibility for drafting, reviewing, and editing the manuscript of the original paper. Furthermore, I supervised and actively participated in the creation of graphics, producing new and additional photos, tables, illustrations, and other content specifically for this thesis.*

4.1 INTRODUCTION

As discussed in Chapter 3, high fidelity in MR settings has been shown to positively affect user engagement, which in turn can influence learning and performance in medical training environments. However, applying these insights to specific medical use cases, such as MRI-guided interventions, poses unique challenges. Fidelity in such applications must balance the demands for realism with the practical constraints of technology, system complexity, and accessibility.

Medical MR simulations, especially for tasks like MRI-guided interventions, face specific obstacles. Ethical considerations, limited access to MRI machines, and electromagnetic interference complicate HCI research in actual MRI environments. Consequently, alternative methods are needed to facilitate the development and evaluation of medical tools within controlled, accessible settings. MR simulations using physical mock-ups and virtual radiology suites offer solutions by enabling realistic, repeatable testing environments. This approach allows researchers to test early prototypes and develop interaction techniques without requiring clinical MRI system access [206, Further8]. Such MR prototypes also support the exploration of innovative interaction techniques, with insights that can be applied in practical contexts [226, 332]. The use of MR prototypes to investigate usability and

UX aspects is presented in my own research [Further5], establishing a classification framework that supports HCD within these simulated environments.

This chapter focuses on how MR can enhance medical simulations, specifically examining the importance of visual and interaction fidelity in such contexts. As highlighted in Chapter 3, recent advances in MR technology have enabled realistic simulations that support learning and practical applications in medical settings. Designing these simulations requires balancing numerous factors, as explored in Section 2.1.4.5, where it is shown that creating realistic VE involves multiple sensory stimuli beyond visual perception alone. The level of graphical detail, for instance, may shift developer resources away from optimizing interactions and core simulation content, while also increasing hardware demands. For example, Harman et al. [103] found that minimal environments may be adequate for memory recall tasks, and Schmied-Kowarziki and Paelke [248] demonstrated that environmental detail had limited impact on acrophobia patients in virtual exposure therapy. Expanding on findings in Chapter 3, Gonçalves et al. [95] showed that realism in VR enhances UX across fidelity dimensions such as visual, audio, haptic, and olfactory components. Focusing on the visual aspects of VE and IM, this review is particularly relevant to the research objectives discussed here.

In practice, a key question centers around determining the optimal level of fidelity required to achieve desired effects on *Presence*, UX, and other factors within medical simulations—forming the foundation for this thesis's first research question:

> **RQ1** | What are the critical visual and interaction fidelity factors that contribute to creating engaging and effective medical task simulations in MR?

This study aims to optimize the balance between VF and input modalities in medical task simulations, using the placement of a radiological coil in an MRI-guided, needle-based procedure as a practical use case. The simulation emphasizes workflow and interaction, providing a realistic context for evaluation. An experiment was conducted to manipulate VF and IMs, laying the groundwork for design recommendations in future medical task simulation research. Additionally, the developed simulation serves as a versatile tool, functioning both as a testing environment for prototypical evaluations and as a training platform for medical education scenarios. Three levels of VF were defined, and a tangible object-based interaction approach was compared to traditional controller input. The study assessed the effects of these manipulations on *Presence* and UX—key quality measures for MR simulations—while also evaluating the impact of VF on user performance.

A video demonstration of the application is linked in the Appendix (see Section 9.1).

4.2 EXPERIMENT 1

This section presents the primary experiment of this chapter, which focuses on the use of a VE in MRI-guided interventions.

4.2.1 *Materials*

To provide an overview of the medical and technical considerations, as well as the implementation of the VE, this subsection clarifies these aspects. It highlights the rationale behind selecting the interaction task and its alignment with the research objectives. Additionally, it covers the relevant medical background, outlines the procedural steps of the MRI-guided intervention, and details the chosen interaction task for implementation.

4.2.1.1 *Medical background*

MRI as an imaging modality allows a high-contrast visualization of soft tissues in the human body and reveals the boundaries of target structures (e.g., tumors) [56]. This is done by establishing a magnetic field and then acquiring a radio frequency signal. Unlike other modalities, it does not emit radiation harmful to the human body [227]. MRI can help perform minimally invasive procedures that allow controlled interventions [4] (e.g., needle-based procedures, such as biopsies) and precise navigation [21]. A biopsy involves taking a sample from a specific tissue for examination [264]. Image guidance in interventions offers advantages, such as reducing the risk of accidentally penetrating or damaging critical blood vessels or surrounding tissues, and increases the likelihood of successfully targeting the tissue of interest. This has led to a rise in the number of MRI-guided interventions in clinical practice [319]. To minimize tissue damage, it is crucial to plan the entry point and pathway with as much spatial resolution as possible before intervention, which is achieved through the use of gradients [268]. Additionally, specific MRI coils are often required depending on the application. These locally placed radiofrequency coils near the target area improve the signal-to-noise ratio across the field of view, resulting in clearer images [318]. This is particularly important for achieving clear images in fast, real-time interventional sequences, where the local intervention coil plays an even more critical role than in planning datasets.

INTERVENTIONAL PROCEDURE The individuals involved in each step of the procedure have different skills and responsibilities. Pannicke [199] gives an overview of the team as follows: The interventional radiologist, a trained specialist, is responsible for carrying out the procedure. They are supported by a radiologic technologist, who assists in patient preparation and manages the operation of the medical equipment during the intervention. An anesthesiologist can play a crucial role, especially in therapeutic interventions, by administering general anesthesia and monitoring the patient's vital signs. This professional is also prepared to manage respiratory arrest if necessary. Additionally, an anesthesia nurse provides specialized support, ensuring the patient's safety and comfort throughout the procedure.

MRI-guided interventional treatments are not yet widely implemented, but Barkhausen et al. [21, 22] provide a comprehensive overview of the methods, current state of development, and potential applications. The workflow presented here offers a broad summary of these procedures and is based on the aforementioned literature, the insights of Pannicke [199], clinical expertise from a radiologist involved in this project, and an observational analysis conducted by a project partner. Eight steps were identified at the core of the process, as illustrated in Figure 5 and described as follows:

1. **Patient Admission:** The process begins with the patient's admission, involving administrative tasks and evaluating the patient's suitability for MRI to ensure there are no contraindications such as metal implants or severe claustrophobia.

2. **Patient Preparation:** Initially, the patient is positioned on the MRI table, and a surface coil is placed for preliminary imaging to assess the target area.

3. **Puncture Planning:** After the initial scans, a loop coil may be positioned to enhance the imaging of the target area. This coil adjustment ensures images with a high signal-to-noise ratio, which is crucial for precise puncture planning. Proper placement of the loop coil is essential for optimal imaging results.

4. **Locating the Puncture Site:** The entry point on the patient's skin is identified using techniques such as live MRI [237] or the fingertip method [82], where the exact location is marked on the skin for precision.

5. **Needle Placement:** The procedure moves into a sterile phase where the intervention site is disinfected and covered with sterile drapes. The needle is then carefully inserted under live imaging guidance, ensuring it follows the planned trajectory to the target.

6. **Therapy/Biopsy:** Once the needle is correctly positioned, tissue samples are collected or therapeutic interventions like tumor

ablation are performed. Real-time MRI helps monitor the procedure and adjust the approach if necessary.

7. **Post-procedure Patient Monitoring:** After the intervention, the patient undergoes further MRI scans to check for any immediate complications and to ensure the effectiveness of the treatment. This step is critical for confirming that all target tissues have been adequately addressed or removed.

8. **Patient Discharge:** The patient is monitored until stable and then discharged with instructions for home care and follow-up appointments. This step concludes the workflow of the MRI-guided biopsy.

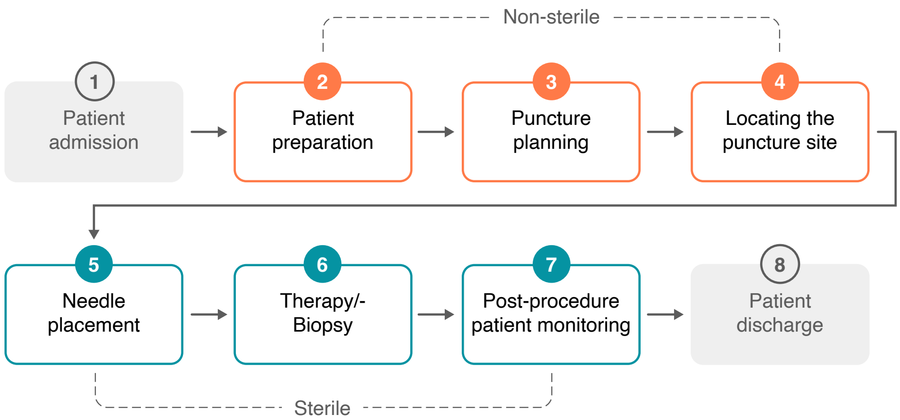


Figure 5: Workflow steps of an MRI-guided intervention adapted from Pannicke (2021) [199]. Licensed under CC BY 4.0.

SELECTED EXEMPLARY TASK The workflow described in the previous section is streamlined, focusing on the essential tasks primarily carried out by the radiologist. It is not considered necessary to replicate the entire workflow for the research objectives because doing so would involve significant effort and could introduce errors.

For MR training or simulation purposes, implementing steps 2, 3, 4, and 5 shows particular promise (Recap: see Figure 5). These steps involve substantial interaction with the MRI machine and demand precise execution, making them ideally suited for MR-based skills training. However, despite the potential benefits, it has been decided not to pursue these steps within this project. This decision is based on the fact that the user, typically a radiologist, remains primarily stationary in one area of the room, thus experiencing limited engagement with the broader environment.

The placement of the coil was selected as an exemplary task from the interventional workflow to be reproduced in MR. This task was chosen because it involves a wide range of motion and requires direct interactions with tangible objects at different locations within the room. As per Skarbez et al.'s observation [273], physical coherence is largely influenced by the level of interaction from the participants, leading us to infer that each user engages with both real and virtual objects. The coil is stored in the MRI room, either in a cabinet or on a coil cart, and can only be accessed by walking around the room. Subsequent placement of the coil requires direct interaction. A loop coil was chosen for its specific imaging capabilities and ergonomic handling. Furthermore, the coil must be positioned manually in a relatively confined and precise location, making it invaluable for training purposes because trainees can learn how to optimally position the coil for both imaging and procedural access. The utilized coil is based on the Noras iLoop Interventional Coil[1], which was provided as a real model and accompanied by training during a workshop conducted by the company.

As connecting the coil to the MRI machine is also an interactive step, this is also included in the virtual task, making the whole process more immersive and practical. The input modality for the task is through natural hand interactions with tangible objects, enabling interaction with a variety of objects with different ergonomics and testing various motion sequences during the experiment. The task, as replicated in reality and its virtual translation, is illustrated in Figure 8. The selected task was reduced to elementary steps, which may not correspond to 100% of the detailed sequence and intermediate steps and are not necessarily carried out by the radiologist themselves. However, the planned evaluation was assessed by an expert as realistic. A description of the specific task implemented for evaluation purposes can be found in Section 4.2.2.3.

4.2.1.2 *The Virtual Interventional MRI-suite*

VEs for real-time applications can be reconstructed in different ways. For example, it is possible to use 360-degree panoramas of the real world as environments for virtual space [233] and to enable interactions in 360-degree videos [301]. Reconstructing environments using 3D scans or photogrammetric methods is also a common method [80]. 360-degree panoramas can also be combined with 3D models in real-time applications, but the full individualization of the environment is lost, and interaction with objects is limited [297]. A particular obstacle in the context of this research is the magnetic field in MRI examination rooms, which severely limits the use of these technologies. While 3D scanning can be applied to individual objects not used in the examination room, the manipulation of material, texture, surface, color, and object properties is limited and requires substantial effort. To address these challenges and leverage the benefits of realistic texturing, free placement, and interactivity, traditional 3D mod-

1 NORAS MRI products GmbH, Germany, https://noras.de/en/mri-produkte/iloop-interventional-coil-055t/

eling techniques were employed. The virtual replica, created for use in MR, was based on photographs and inventory lists from a real MRI intervention room at our partner university and was developed in consultation with medical procedure experts.

FIDELITY LEVELS Research has identified two major areas affecting the visual design of MR applications: the lighting system [85, 246, 280, 331, 339] and object design [94]. These factors significantly impact user perception in the virtual world. However, existing studies rarely address the medical context specifically. A literature review indicates a lack of comprehensive investigation into the combined effects of previously studied design components. It was hypothesized that a combination of these factors might either mitigate or amplify their individual impacts. To test this hypothesis, three fidelity characteristics were developed, each blending abstract and realistic visual components. Consequently, visual quality was enhanced to reflect the development effort of such VEs, and three levels of gradation were established:

1. Essential Detail Level (EDL) – This level prioritizes minimal development effort, focusing solely on task-relevant objects in the environment while representing the rest schematically with 3D primitives.

2. Standard Detail Level (SDL) – At this level, the typical effort involved in creating 3D environments, including basic shapes, materials, and lighting, is followed.

3. Advanced Detail Level (ADL) – The primary emphasis at this level is on achieving optimal graphic implementation, with a strong focus on attaining a high level of visual accuracy to closely resemble realism.

Figure 6 gives an overview of the environment and its visual fidelity characteristics. A detailed technical realization of the three gradations can be seen in Table 1.

To give users the feeling of being in a real operating room (OR), audio recordings of the pumping sound from the helium compression system of an MRI scanner were added[2]. While this contributes to the ambient noise, the soundscape in an MRI room—particularly during image acquisition due to the gradient coils—creates a loud environment overall, often necessitating the use of headphones and making verbal communication challenging. Therefore, 2D audio in SDL and 3D (Spatial) audio in ADL were implemented. To maintain minimal development effort in EDL, the audio was completely omitted. Additionally, different qualities of representation for male and

2 (CC BY 3.0): `https://freesound.org/people/solidphase/sounds/442831/`

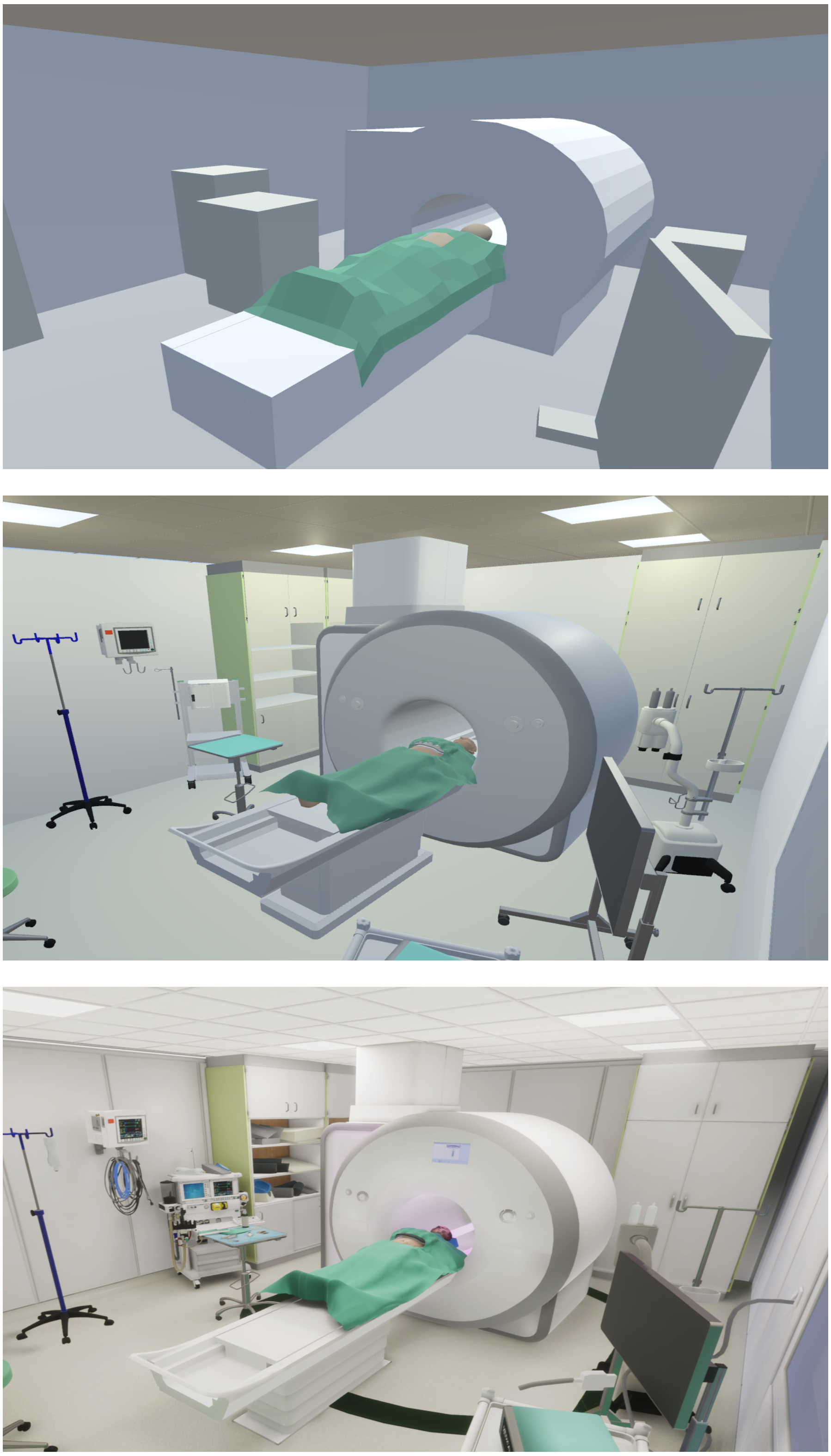


Figure 6: Virtual environment for evaluating an interventional procedure at different VFs. From top to bottom: lowest fidelity (EDL), medium fidelity (SDL), and highest fidelity (ADL).

Table 1: Breakdown of the different visual fidelities (ADL, SDL, EDL) and their technical implementation. Categories are numbered for reference: 1) No. of polygons, 2) Object features, 3) Material, 4) Light & Shadow, 5) Room architecture, 6) Interior. Adapted from Schott et al. (2023a) [Core5]. Reused in accordance with ACM's policy for open access articles.

#	ADL	SDL	EDL
1	>100.000 faces	>10.000 faces	<10.000 faces
2	Realistic shape replication, High-poly models, Natural irregularities	Essential geometric details (bevels / fillets), Medium-poly models	Basic geometric shapes, Low-poly models, No details (e.g. fillets)
3	HDR-materials, High resolution texture maps, Realistic properties & colors	Monochrome, Reflections, No texture maps	Monochrome, No reflections, No texture maps
4	HQ shadows, Baked lightmap, Ambient occlusion	No shadows, Default light setup (6 Sources)	No shadows, Two light sources
5	High-detail OR-facilities, (e.g. sockets, baseboards), Natural irregularities	Essential OR-facilities (e.g. doors, windows)	Waiver of facilities (no doors, windows)
6	Full OR-furniture & Accessories	Basic OR-furniture, (no small-scale equipment)	Only task requirements (no additional furniture)

female hands were implemented, with photo-realistic hand models and forearms included in ADL, while EDL omitted the gender feature to streamline development

INTERACTION MODALITIES As a realistic input modality, non-contact hand tracking is combined with real tracked objects to provide haptic and visual hand feedback while offering many degrees of freedom. This form of interaction, which has been shown by Schrom-Feiertag et al. [261] to enhance the user's *Sense of Presence* and interaction fidelity, is similar to the findings of Luong et al. [159], who noted that hand tracking significantly increases the naturalness of user interactions within VE. A study by Bolder et al. [32] shows that implementing a form of interaction with tracked hands and real objects for interaction can achieve similar *Usability* as in the real world. This *Usability* is further supported by Adkins et al. [5], who found that despite the intuitive nature of hand tracking, controllers can sometimes enhance task performance by reducing physical strain, particularly in longer sessions. The controller represents a second input modality that allows the user to grasp and move virtual objects by pressing buttons. This modality is particularly effective in situations where prolonged interaction might lead to user fatigue, as noted by Luong et al. [159]. The implemented input modalities (controllers/hands-only) allow for the interaction with virtual replicas of tangible objects (coil, adapter, door handle) to perform the medical task at different VFs. An illustration of the implementation of the respective input modalities is shown in Figure 7.

Figure 7: Input modalities and realized Visual Fidelity. Top: Controller and hands only input. Bottom: Virtual replications of tangible objects (coil, adapter, door handle).

4.2.2 *Evaluation*

A study was conducted to evaluate the levels of VF and the IMs of an exemplary medical task in a virtual OR in terms of their impact on *Presence*, UX, and user performance. The methodology of the study is described in the following. The experiment followed a two-way within-subjects design that combined the different VFs (EDL, SDL, ADL), with the two IMs *controller* and *tangible objects*. Because effects of different fidelity alterations on a holistic task simulation were to be evaluated, multiple parameters were varied at the same time. In this way, more general answers were sought initially.

4.2.2.1 *Hypotheses*

This study aimed to investigate several hypotheses. It is generally presumed that interacting with tangible objects rather than controllers is more natural, and, thus, more realistic. First of all, it had to be ensured that the manipulations were perceived as intended by the subjects. Therefore, the first two hypotheses are:

- H1.1: Higher VF evoke higher subjective realism.
- H1.2: Interacting with tangible objects is perceived as more realistic than controller-based interaction.

Based on related work [112, 191], it is speculated that VF also has a positive effect on *Presence* in this case. The following hypotheses are made to verify this assumption:

- H2.1: Presence is positively affected by increasing VF.
- H2.2: Tangible objects elicit a greater *Sense of Presence* than controller interaction.

In their literature review, it was shown by Gonçalves et al. [95] that higher degrees of realism most often also had positive effects on UX. However, since not all reported studies could demonstrate this effect, this aspect is intended to be investigated as well:

- H3.1: Higher VF are associated with a better UX.
- H3.2: Tangible objects improve UX compared to controller input.interaction.

A moderate correlation between *Presence* and user performance was found by Stevens et al. [290]. It was questioned whether such a relationship could also be demonstrated in the experiment. The final hypotheses are:

- H4.1: Better user performance can be observed for higher realism.
- H4.2: Participants perform better using the tangible objects than using the controller.

Figure 8: Realization of subtasks in reality including custom 3D printed tangible objects (left) vs. the analog realization in MR with hand interaction in ADL (right). Adapted from Schott et al. (2023a) [Core5]. Reused in accordance with ACM's policy for open access articles.

4.2.2.2 *Sample design*

Professional expertise was not required for the study, as it was deemed unnecessary for performing the experimental tasks. However, to ensure that subjects had some basic medical background knowledge and were familiar with medical environments, only participants who were studying human medicine from their second semester onward were invited. Nineteen medical students aged 21-31 years (Median (Mdn) = 24) were recruited from our university. Of these, fourteen reported their gender as female, while five reported as male. Each individual was compensated 30 euros. A rating scale from 1 (no experience) to 5 (very experienced) was used to assess technological experience, which resulted in the following distribution: technology affinity (Mdn = 3), gaming experience (Mdn = 2), MR experience (Mdn = 2). More than half of the participants reported having a medium technological affinity. A high technological affinity was reported by one participant. The majority reported having little to no (>60%) experience with gaming and MR, with none rating their experience at the highest level (5). Ten participants reported using glasses to correct their vision. A color vision deficiency was not reported by any participant. All participants reported being in their 5th semester or higher of medical school. See Appendix (Chapter 9) for the demographic data collection sheet.

4.2.2.3 *Task*

A radiological coil placement task was considered for this study. Participants were placed in an interventional MRI suite and had to move to three randomized positions in the room (Figure 8 (1)). Once there, a cabinet needed to be located and opened (Figure 8 (2)). Participants found a flexible loop coil inside of it and needed to take it (Figure 8 (3)). The coil then had to be placed on top of an already prepared virtual patient (Figure 8 (4)). A mark on the patient's skin showed where the coil needed to be positioned. Participants were instructed to align the center of the circle-shaped coil with this marking. Finally, the coil needed to be connected to a plug positioned at the MRI couch to end the task (Figure 8 (5)).

4.2.2.4 *Variables*

Two independent variables were investigated in this experiment. Both were given by the factorial design of the study. The first manipulated independent variable was the VF of the VE. The second one was the respective IM. Seven dependent variables were observed in the experiment. First, the Task Completion Time (TCT) was measured. Measurements began when participants opened the closet door and ended when the coil was plugged in. In addition, the Placement Deviation (PD) of the final coil position was assessed. This was the distance between the center of the circular part of the coil and a target marked on the virtual patient. The IPQ was used to estimate *Presence*-related qualia. All sub-scales (*General Presence*, *Spatial Presence*, *Involvement*, *Experienced Realism*) were considered as dependent variables for this study. *Experienced Realism* primarily served the purpose of a manipulation check, to assess whether the intended manipulations of realism were successful. Additionally, the short version of the UEQ was used to get an estimate on the users' subjectively perceived UX quality. Finally, a few control measures were collected; participants were asked to rank their experiences with video games and with MR, as well as their technical affinity, as detailed in Section 4.2.2.2. Furthermore, the time participants spent in MR before starting the actual coil placement task was measured. This *Adjustment Period* was the time required to walk to all predefined floor markers.

4.2.2.5 *Apparatus*

The study took place in a *Usability* lab with a total size of $\approx 50\ \text{m}^2$ at our university, where a mock-up resembling a real MRI device was present [Further13]. The MRI mock-up measured approximately 2 m in height, 2.3 m in width, and 1.7 m in length, with an 80 cm bore and a patient bed (approximately 2.5 m $\times$ 0.5 m). A loop coil, a closet handle, and a coil connector were 3D printed. The coil was placed inside a closet in one corner of the room and the printed handle was glued to one door of that closet (see Figure 8). A short section of a plastic tube was attached to the coil connector to mimic the cable that would take its place in real MRI suites. A human torso dummy made of Styrofoam represented an average patient in size and position and was placed on the MRI bench. Velcro on this dummy was used to attach the 3D printed loop coil replica, which also had Velcro dots on its bottom. The coil connector was placed on top of the dummy to be near the user when needed. To ensure the greatest possible freedom of movement, a ceiling-mounted cable management system was used for the MR system. Tape was placed on the floor to mark the target position and individual interaction points analogous to the performance of the study task in VR. The laboratory had a size of about $50\ \text{m}^2$. The virtual tracking space, in which the participant could move, had a size of approx. 5x5 m. The VE represents a faithful replica in terms of room size, as well as placement of furniture and objects in the room.

The setup included a Valve Index HMD[3] equipped with the Stereo IR 170 Evaluation Kit[4] for hand tracking. The system also utilized four Valve Basestations (2.0) and corresponding controllers. Enhanced precision in tracking position and rotation was achieved

---

3 Valve Corporation, `https://www.valvesoftware.com/en/index`
4 Ultraleap Ltd., `https://www.ultraleap.com/product/stereo-ir-170/`

using Vive Trackers 3.0[5] mounted on the closet door, the loop coil, and the connector. Calibration of the MR handhelds was performed manually, using visual feedback to accurately align the virtual and physical objects.

The PC used featured the following specifications: an Intel Core i7-8700K 3.70GHz 6-core CPU[6], an NVIDIA GeForce RTX 3090 Ti[7], 32 GB RAM, and a 512 GB PCIE 3.0 SSD. Tangible objects were 3D printed using Ultimaker PLA and TPU 95A (flexible materials)[8]. Software applications were run on Unity 2020.3.31 in C# with Microsoft Windows 10 Pro (build 19044)[9] and Unity[10]. Hand tracking was implemented with Ultraleap's Hand Tracking v5.2.0 and Leap Motion Core Plugin v5.5.0 for Unity. Rendering of EDL and SDL was conducted using Unity's built-in render pipeline, whereas in ADL, the High Definition Render Pipeline (HDRP) version 10.8.1 was employed with materials from the Sample Scene (Measured Material Library for HDRP[11]). Realistic female and male hand models, later assigned to subjects according to gender, were sourced from the Unity Assetstore[12]. Changes to the material in SDL were made, while standard low poly models were utilized in EDL. Further details regarding the technical implementation of the rendering can be found in Table 1.

4.2.2.6 *Procedure*

After welcoming the study participants, informed consent was obtained and demographic data were collected. Subsequently, the subject received a brief introduction to the medical field of application and an overview of the study procedure. As training, each participant first started performing the task in reality (without MR equipment), which was demonstrated once by the study investigator. For this purpose, the interaction objects and the spatial laboratory conditions (mock-up as a model for the interventional MRI) were addressed. The participant was reminded of the correct order in the interaction task: specifically, the order of coil placement and plug insertion was emphasized. The coil should be placed centrally over the target marker on the patient. Care should be taken to ensure that the Velcro side of the coil was faced down. The participant should perform the task at a reasonable pace and was not motivated to complete the task as quickly as possible. After training, the experimental task described in Section 4.2.2.3 was performed three times for each combination

5 HTC Corporation, https://www.vive.com
6 Intel Corporation, https://www.intel.com
7 NVIDIA Corporation, https://www.nvidia.com
8 Ultimaker B.V., https://www.ultimaker.com
9 Microsoft Corporation, https://www.microsoft.com
10 Unity Technologies, https://unity.com
11 https://github.com/Unity-Technologies/MeasuredMaterialLibraryHDRP
12 https://assetstore.unity.com/packages/3d/characters/humanoids/leap-motion-realistic-female-hands-211090

Table 2: Summary of the ANOVAs' results on *spatial presence (SP)*, *involvement (INV)*, and *experienced realism (REAL)*; as well as robust ANOVAs' results on *general presence (G)*, *UX*, *time*, and *placement deviation (Dev.)* ($\alpha < .05$). Test statistic F and effect size η^2 are reported for ANOVAs and test statistic Q and effect size δ_t are reported for robust ANOVAs. Adapted from Schott et al. (2023a) [Core5]. Reused in accordance with ACM's policy for open access articles.

Variable / Eff. Typ.	Factor	DFn	DFd	F/Q	p	Sig.	η^2/δ_t	Effect
SP								
Main effects	*VF*	2	34	0.030	0.971		0.001	-
	IM	1	17	0.002	0.964		0.000	-
Interaction effect	*VF * IM*	2	34	0.106	0.900		0.001	-
INV								
Main effects	*VF*	2	34	16.541	<0.001	*	0.158	L
	IM	1	17	1.005	0.330		0.008	-
Interaction effect	*VF * IM*	2	34	2.089	0.139		0.013	-
REAL								
Main effects	*VF*	2	34	35.048	<0.001	*	0.372	L
	IM	1	17	0.223	0.643		0.002	-
Interaction effect	*VF * IM*	2	34	4.747	0.015	*	0.053	S
G								
Main effects	*VF*	2		10.581	<0.001	*	0.648	M
	IM	1		0.141	0.707		0.038	-
Interaction effect	*VF * IM*	2		5.736	0.003	*	-	-
UX								
Main effects	*VF*	2		17.924	<0.001	*	0.595	M
	IM	1		0.117	0.733		-0.029	-
Interaction effect	*VF * IM*	2		0.517	0.596		-	-
Time								
Main effects	*VF*	2		0.183	0.833		0.093	-
	IM	1		0.161	0.688		0.078	-
Interaction effect	*VF * IM*	2		0.344	0.709		-	-
Dev.								
Main effects	*VF*	2		9.845	<0.001	*	0.294	S
	IM	1		4.984	0.025	*	0.466	S
Interaction effect	*VF * IM*	2		38.532	<0.001	*	-	-

of VF and IM factor level combination in MR. The order of the six resulting experimental conditions was partially randomized. VF was randomly arranged for each participant. Both IMs were tested one after the other for each respective VF. The order of the modalities was alternated between two participants. In SDL and ADL, corresponding male or female hand models were set, taking into account the gender indicated by the participants. During this period, participants needed to walk to four specific markings on the floor in a given order. Three different but comparable routes were predefined. This phase was included to allow the subjects to become accustomed to, and aware of, the environment before the actual task began. It was also performed during the training phase in reality, where tape was used to mark the positions on the floor.

Table 3: Summary of descriptive results for the dependent variables related to presence ($n = 19$). All entries are in the format: mean value [standard deviation]. SP - *spatial presence*, INV - *involvement*, REAL - *experienced realism*, G - *general presence*. Adapted from Schott et al. (2023a) [Core5]. Reused in accordance with ACM's policy for open access articles.

Variable	SP	INV	REAL	G
Accumulated	4.63 [0.60]	4.43 [1.22]	4.02 [1.22]	5.27 [1.25]
ADL	4.65 [0.58]	4.93 [1.12]	4.79 [0.87]	5.97 [0.88]
SDL	4.62 [0.55]	4.58 [1.03]	4.22 [1.01]	5.25 [1.25]
EDL	4.63 [0.67]	3.78 [1.24]	3.04 [1.07]	4.58 [1.20]
Tangible	4.63 [0.62]	4.33 [1.28]	4.06 [1.34]	5.31 [1.24]
ADL	4.64 [0.58]	4.86 [1.26]	4.86 [0.88]	5.94 [0.73]
SDL	4.63 [0.60]	4.61 [1.03]	4.51 [1.04]	5.67 [0.97]
EDL	4.61 [0.71]	3.51 [1.17]	2.79 [1.08]	4.33 [1.33]
Controller	4.63 [0.58]	4.53 [1.16]	3.98 [1.10]	5.22 [1.27]
ADL	4.66 [0.61]	5.00 [1.00]	4.72 [0.88]	6.00 [1.03]
SDL	4.60 [0.51]	4.54 [1.06]	3.92 [0.90]	4.83 [1.38]
EDL	4.64 [0.64]	4.06 [1.28]	3.29 [1.04]	4.83 [1.04]

4.2.2.7 *Statistical analysis*

All statistical analyses were conducted using R (version 4.2.2)[13]. The data for each dependent measure was first checked for homogeneity of variances with Levene's tests. The test's implementation in the 'car' R package [83] was used for this. Next, normality assumptions were verified. To this end, the data of each variable was fitted to a linear model using the *ez_aov* function of the 'afex' package [270]. Shapiro-Wilk tests were then conducted using the respective linear models' residuals to check for normality. The 'stats' package [217] was used for this purpose. In case homogeneity and normality assumptions were met, two-way repeated measures Analysis of Variances (ANOVAs) were conducted to analyse the data. This was also done using the *ez_aov* function of the 'afex' package [270]. For variables violating one of the assumptions, robust two-way ANOVAs for within-subject designs based on trimmed means were calculated to evaluate main and interaction effects (also see [327]). The function *wwtrim* of Wilcox' [327] R implementation was used here. The δ_t estimate proposed by Algina et al. [11] was interpreted as effect size for main effects assessed by robust ANOVAs. It was calculated using the *akp.effect* function of the 'WRS2' package [161]. Afterwards, post-hoc tests on statistically significant VF main effects were conducted using the *pairwise_comparisons* function of the 'ggstatsplot' package [200]. Pairwise paired t-tests with Bonferroni adjustments were applied to variables having met the normality assumption and robust Yuen's trimmed means tests with p-value adjustments using Hochberg's method were performed otherwise [327].

13 R Foundation for Statistical Computing, https://www.R-project.org/

Table 4: Summary of descriptive results for the dependent variables related to user experience and performance ($n = 19$). All entries are in the format: mean value [standard deviation]. UX - *user experience*, Time [s] - *time*, Dev. [mm] - *deviation*. Adapted from Schott et al. (2023a) [Core5]. Reused in accordance with ACM's policy for open access articles.

Variable	UX	Time [s]	Dev. [mm]
Accumulated	5.45 [1.17]	20.71 [6.79]	10.53 [6.32]
ADL	5.97 [0.95]	20.92 [6.87]	10.56 [5.57]
SDL	5.56 [1.00]	20.55 [6.24]	09.00 [5.44]
EDL	4.83 [1.25]	20.67 [7.40]	12.03 [7.53]
Tangible	5.43 [1.21]	20.74 [6.88]	11.37 [6.47]
ADL	5.96 [1.03]	20.58 [6.71]	09.60 [3.89]
SDL	5.61 [1.02]	20.91 [6.80]	07.96 [3.47]
EDL	4.72 [1.26]	20.72 [7.48]	16.54 [7.71]
Controller	5.47 [1.13]	20.69 [6.77]	09.70 [6.11]
ADL	5.97 [0.89]	21.27 [7.20]	11.51 [6.84]
SDL	5.51 [1.01]	20.18 [5.80]	10.05 [6.82]
EDL	4.93 [1.27]	20.61 [7.54]	07.52 [3.78]

4.2.3 *Results*

This section presents the complete set of experimental results. In terms of statistical outcomes, the mean values were calculated to aggregate PD and TCT data under identical experimental conditions. Then, final *IPQ* sub-scale and UEQ scores were calculated according to the questionnaires instructions. Table 3 and Table 4 summarize all resulting descriptive results. No conducted Levene's test showed significant results. Thus, it was assumed that the homogeneity assumption held true for all dependent variables. However, significant Shapiro-Wilk test results on *General Presence*, UX, TCT, and PD suggested that this data would not be normally distributed. Therefore, robust ANOVAs were conducted on these variables. *Spatial presence*, *Involvement*, and *Experienced Realism* were evaluated using conventional two-way repeated measures ANOVAs. Results of all these analyses are summarized in Table 2. Statistically significant VF main effects on *Experienced Realism*, *Involvement*, *General Presence*, and UX were revealed. These effects are visualized in Figure 9. Pairwise comparison results are depicted in these plots. PD showed significant main effects for both factors. However, these are challenged by a significant interaction effect on that variable. PD results are visualized in Figure 11. Significant interaction effects were also shown for the *Experienced Realism* and UX variables. These effects are visualized in Figure 10.

4.2.3.1 *Control measures*

The participants' answers regarding video game experience, MR experience, and technological affinity were not evenly distributed. For

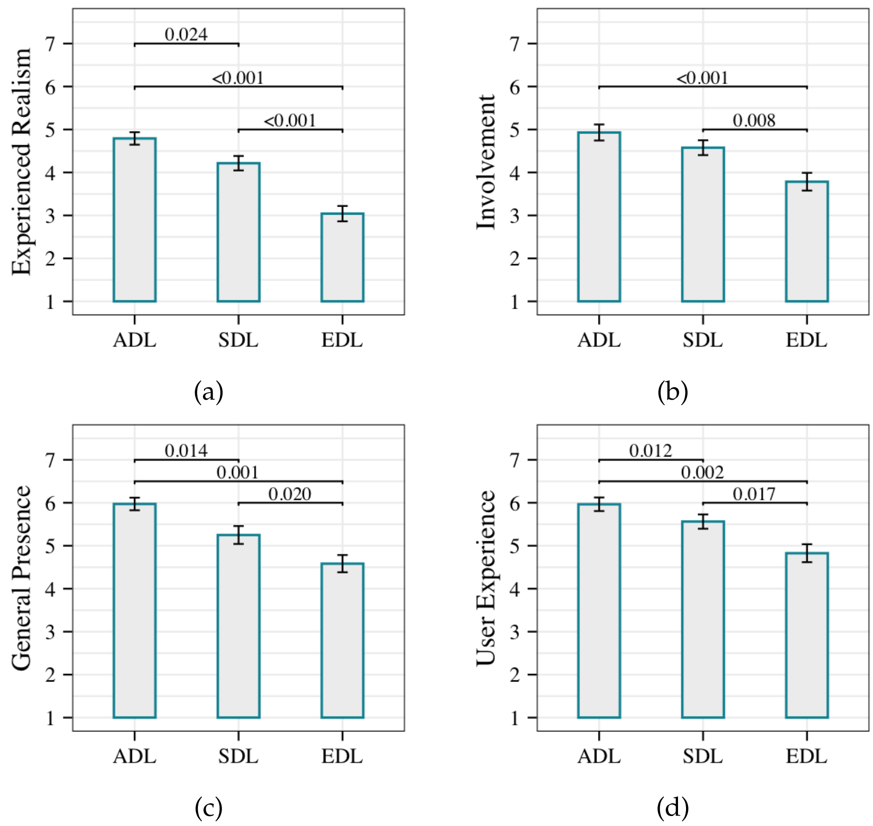
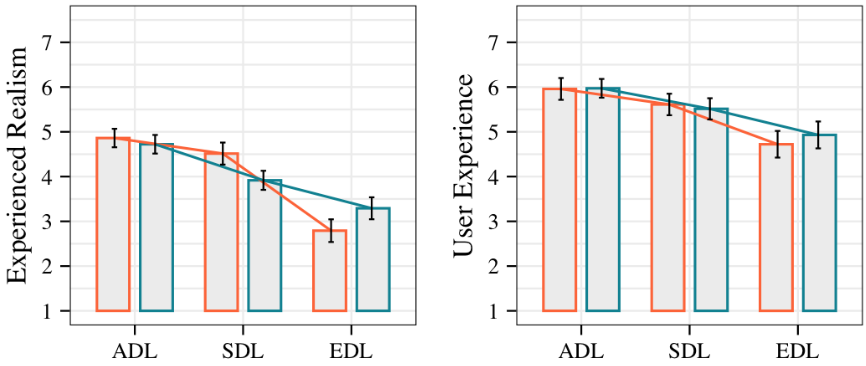


Figure 9: Significant main effects of the VF factor on: (a) *Experienced Realism*, (b) *Involvement*, (c) *General Presence*, and (d) UX. Bars represent mean values and error bars represent standard errors. Significant post-hoc test results are highlighted with brackets. Adapted from Schott et al. (2023a) [Core5]. Reused in accordance with ACM's policy for open access articles.

example, 10 of 19 participants reported having a medium level of technical affinity, 12 of 19 participants had very little to little experience with video games and 11 of 19 participants said they had very little to little experience with MR. Because of this uneven distribution of group sizes, extensive statistical analyses of the effect of these measures did not seem viable. However, as exemplary analyses, Pearson's correlation tests between technical affinity and *Experienced Realism* were conducted for each combination of factor levels. No test returned significant results. Therefore, these control measures were not considered any further. Regarding the *Adjustment Period*, a Shapiro-Wilk test revealed a violation of normality. Therefore, this measure was analyzed with a robust two-way repeated measures ANOVA. This test showed a statistically significant VF main effect ($Q = 16.94, p < 0.001$). Follow-up pairwise comparisons using robust Yuen's trimmed means tests showed that participants spent significantly more time in the adjustment phase in ADL ($M = 7.75s, SD = 1.01s$) compared to



(a) (b)

Figure 10: Significant interaction effects: (a) *Experienced Realism*, and (b) UX. Bars represent mean values and error bars represent standard errors. ■ represents the IM 'Tangible', while ■ represent 'Controller'. Lines connect VF mean values with respect to each IM factor to visualize the interaction effects. Adapted from Schott et al. (2023a) [Core5]. Reused in accordance with ACM's policy for open access articles.

both, SDL ($p = 0.001$; $M = 6.93s, SD = 0.85s$) and EDL ($p = 0.001$; $M = 6.89s, SD = 0.80s$).

4.2.3.2 *Interpretation of results*

The following attempts to interpret the identified effects and to find reasons for their occurrence.

MANIPULATION CHECK The *Experienced Realism* sub-scale of the *IPQ* questionnaire was considered as a manipulation check to ensure the conditions had the intended effects. Statistically significant differences between factor levels of the VF variable confirmed a clear ranking between conditions (see Figure 9). Therefore, H1.1 can be accepted. However, no significant differences could be detected regarding the IM factor. Observations during the study and inspection of the raw data revealed that whether tangible devices or controllers were perceived more realistically seemed rather user dependent. Thus, H1.2 cannot be accepted.

In addition, a significant interaction effect was shown regarding the *Experienced Realism* variable (see Figure 10). The tangible IM seemed more realistic than the controller modality in SDL. However, this assessment was reversed in EDL, which is why this significant interaction effect occurred. Because the higher *Experienced Realism* assessment in EDL (using the controller) was still below the smaller value in SDL (also using the controller), it is argued that the significant VF main effect is valid despite the interaction effect and does not influence the acceptance of H1.1.

A reason for the interaction may be found in the visualization of the interaction devices. The EDL showed very abstract hand models without textures, while SDL showed already quite realistic ones. At the same time, the controller's appearance did not change much across conditions. It may be plausible that participants focused more on their hands when operating tangible objects than when using the controller. Therefore, the more realistic-looking controller evoked a higher degree of *Experienced Realism* than the abstract hands in EDL. Simultaneously, the more realistic hands in SDL may have been perceived as more natural and, thus, more realistic, than the controller. Interestingly, a comparable effect could not be shown for the ADL condition. Here, both IMs ranked very similarly. This may have been because the inherent VF of the VE caused participants to focus more on their surroundings and less on their hands. Thus, when filling in the *IPQ* questionnaire after MR exposure, they ranked both IMs similarly for this VF. An indicator for this can be found in the *Adjustment Period* control measure. Participants required significantly more time to navigate all the ground markings in ADL. It can be argued that this was caused by the participants spending more time observing their surroundings and, as a result, concentrating less on the actual task.

FIDELITY FACTOR Consistent results of the VF factor show a clear ranking and relationship between the degree of visual quality and MR-related qualia. Higher VF resulted in a higher sense of "being there", caused participants to feel more involved in the VE and to devote more attention to it, and to have a generally better UX (Figure 9). Therefore, H2.1 and H3.1 are considered as accepted.

Regarding the latter, half of the items of the UEQ are related to the hedonic quality of the system. Hence, it was less surprising that improving the visual quality and attractiveness of the VE was also reflected in this measure. The significant interaction effect on UX is similar to the one on *Experienced Realism*. Likewise, it can be explained analogously and was probably caused by the controller being rendered similarly in EDL and SDL, while the respective hand models showed visual differences. These observed differences are comparably low and the overall trend between VF conditions seems unaffected by this interaction.

The increased *Involvement* ratings constitute a more interesting finding. Participants were less aware of their surrounding real world in ADL. This indicates that the increasing match between the VE and the real world blurred the boundaries between both realities. However, it is not known if this was caused by visual quality improvements or by the increased amount of observable items in the VE at higher VF. The *General Presence* item of the *IPQ* questionnaire was answered significantly different between *visual fidelities*. It is said to be closely related to *Spatial Presence* [262]. However, this sub-scale did not seem to be affected by the VF factor in this study, which partly diminishes the acceptance of H2.1. The *General Presence* item asked participants if they had a sense of "being there". This question could have been interpreted ambiguously and other aspects, e.g., *Experienced Realism* and *Involvement*, could have also affected responses. *Spatial Presence*, as in the sense of being *physically* present in the VE, may be more susceptible to other factors (such as display-related properties) that were not manipulated in this study.

INTERACTION MODALITIES FACTOR Except for PD, which will be discussed separately, no statistically significant differences were found between both IMs in this work. Since the descriptive data also does not reveal any meaningful insights, it can be argued that, for the selected task, the choice of IM should be rather user preference-driven. Therefore, hypotheses H2.2 and H2.3 can be rejected. It was expected that interacting with haptic tangible objects would evoke a more positive UX response. Its absence may have been due to implementation reasons. All tangible objects were tracked using only one tracker. In addition, some parts were flexible, which could not be translated to MR. Visuoproprioceptive mismatches between the displayed virtual objects and their real world counterparts caused by tracking or registration errors may have negatively affected participants' perception of this modality.

USER PERFORMANCE Two user performance measures were considered. The TCT variable showed no significant differences. In contrast, two significant main effects were found on coil PD. However, both appear to be strongly affected by a significant interaction effect on this variable. In SDL and ADL, participants were more accurate using the tangible objects. In contrast, subjects could place the coil with a similar accuracy using the controller in the EDL but performed worst using the tangible objects in this VF. Therefore, H4.1 and H4.2 are not able to be accepted. Considering a loop coil diameter of 18.5 cm, observed differences in deviations below 1 cm seem only minor. Nonetheless, the statistically significant nature of these differences is noteworthy. Perhaps participants were the most accurate overall using the controller in EDL, because this factor level combination introduced the least amount of distracting stimuli, thus providing an environment for more concentrated work. In the more realistic VEs, participants may have benefited from the more natural tangible IM creating an overall workflow close to reality that facilitated more concentrated performance.

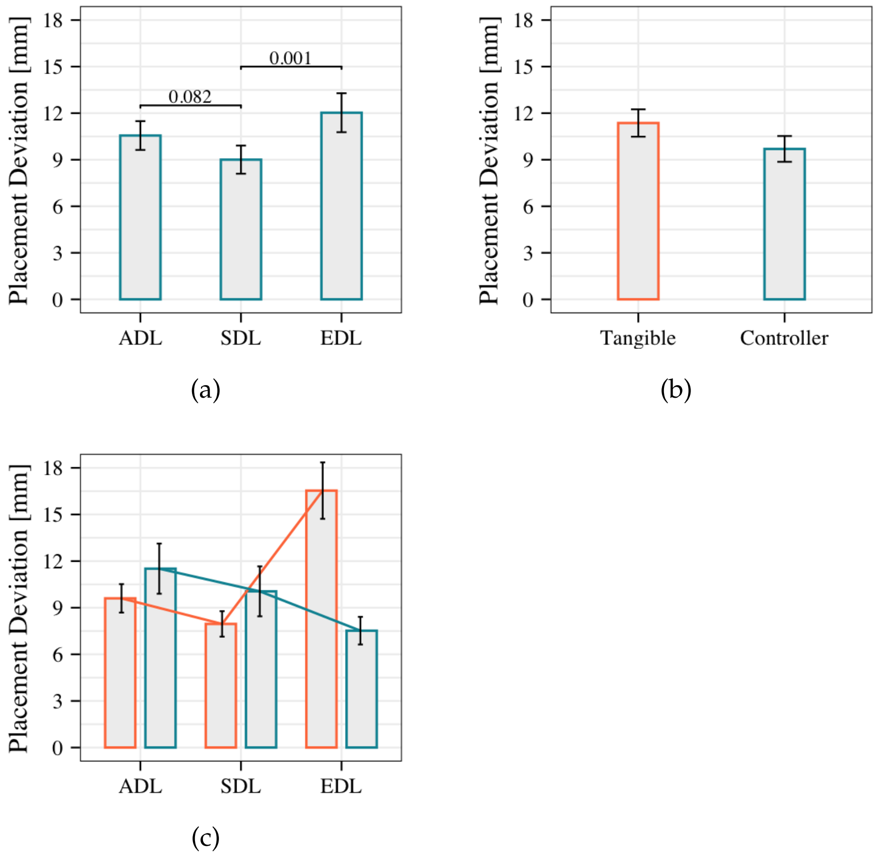


Figure 11: Significant main and interaction effects on PD. (a) *visual fidelity main effect*, (b) *IM main effect*, and (c) interaction effect. Bars represent mean values and error bars represent standard errors. ■ represents the IM 'Tangible', while ■ represent 'Controller'. Significant post-hoc test results are highlighted with brackets. Lines connect VF mean values with respect to each IM factor to visualize the interaction effect. Adapted from Schott et al. (2023a) [Core5]. Reused in accordance with ACM's policy for open access articles.

4.2.4 *General discussion*

Significant effects from the degree of VF on several subjective measures were identified. This research finding is in line with the work of Newman et al. [191] and Mizuho et al. [181], who also found that a high-quality VE created a greater *Sense of Presence*. The IM did not seem to affect *Presence*-related qualia in this study, which is similar to the work of Zhou and Rau [337]. There, using a tangible object in MR did also not improve *Involvement* compared to a standard condition.

DIVERSITY AND INCLUSION A higher proportion of female participants was observed in this evaluation. The selection process prioritized availability and willingness to participate, with no gender-specific criteria being influential. Knowledgeable medical students were recruited generally, and a random sampling approach was employed, which could potentially impact the generalizability of the findings. As a result, future research should aim to extend these findings to broader populations to enhance the external validity of the study and ensure inclusion and diversity.

VIRTUAL BODY OWNERSHIP AND HAND VISUALIZATION In this study, the visual rendering of the user's virtual hands varied across different fidelity levels (VF), which may have influenced participants' Virtual Body Ownership (VBO) and, consequently, affected results. The literature suggests that VBO is influenced by both sensorimotor coherence (bottom-up factors), such as the movement congruence between real and virtual hands, and cognitive congruence (top-down factors), such as the visual resemblance between the virtual and actual hands [302]. All conditions in this study implemented consistent hand tracking mechanisms, maintaining coherence on the bottom-up layer. Studies by Maselli & Slater [168] and Lugrin et al. [158] have shown that even mannequin-like hands or non-human avatars can evoke vivid VBO without highly realistic appearances, while Latoschik et al. [143] found that more realistic avatars enhance body ownership. This leaves the potential effects of varying hand appearance in our study open for further investigation.

For hand visualization, gender-matched 3D hand models were used in SDL and ADL, providing personalized visual representations with fully depicted arms, known to enhance VBO [129]. However, EDL and SDL lacked forearm connections, which might have influenced VBO, especially during dynamic interactions [298]. While the hand models varied in size by gender, individual differences in hand size and skin tone were not accounted for, potentially affecting interaction. Although the study sample was relatively homogenous, future designs should consider a range of skin tones to enhance inclusivity and reduce potential bias. In clinical settings, gloves are often worn for sterility, but they were omitted here to maintain a consistent VBO effect [100].

CONGRUENCE AND PLAUSIBILITY A novel theoretical model is proposed by Latoschik & Wienrich [144] suggesting that congruence and plausibility are the two essential conditions of MR experiences. Congruence is referred to as the coherence of processed and expected cues on two bottom-up (sensation and perception) and one top-down (cognition) layer. This congruence then leads to a condition of plausibility which affects other qualia, such as *Presence* and VBO. In this study, changing the degree of visual fidelity may be interpreted as a manipulation on the cognition layer, as the coherence between the processed virtual OR and the expectation of a real one was altered. Different resulting congruence levels resulted in different degrees of plausibility. This may then explain the observed effects on *involve-*

ment and *general presence*. At the same time, the two investigated IMs may have had less effects on the observed measures because their expected and processed behavior was equally congruent, thus leading to similar plausibility.

UNCANNY VALLEY Interaction effects on VF, UX, and PD were observed. EDL was rated higher and was seen to perform better when paired with controller interaction. ADL received higher ratings and exhibited lower PDs when tangible objects were utilized. The reversed factor level combinations may have then performed worse because the mismatch caused an uncanny valley-like effect [186]. This theory stems from robotics and describes the phenomenon that, after a certain degree of human likeness, the human observer's emotional response will not increase anymore, but instead decrease. Then, at a very high level of human likeness, the emotional response will increase again, to even higher degrees than experienced before. With respect to VR, this theory has been applied mostly to virtual avatars [143]. Howard et al. [108] also extended the theory to simulation and VEs and concluded that the quality degrees of the environment and the options to control it need to match. If they do not, an *uncanny valley* effect may occur. Applying this to the results, it could explain the observed interaction effects on *visual fidelity*, *user experience*, and *placement deviation*. McMahan et al. [172] investigated whether such an effect also applies to interaction in MR. They theorize that, as with robots, more natural interaction paradigms will feel and perform worse after a certain degree and will only improve once a high resemblance to real world interaction is achieved. Both the controller and the tangible IMs performed very similarly in this study. The tangible condition was designed to feel more natural and to be, thus, more realistic. However, the final prototypical implementation may have just not been good enough. Slight tracking and registration errors may have caused a feeling of eeriness that is associated with the uncanny valley effect.

EXPERIENCED REALISM In general, lower ratings were received for *Experienced Realism* in ADL, as would have been expected ($M = 4.65$ on a scale from 1 to 7). This suggests that there is still some room for improvement. Future work could investigate if different HMDs have an impact on this measure. For example, the Varjo XR-3[14] was evaluated to provide a very high visual acuity [133]. Moreover, the visual realism of the developed VE could be improved by including social companions. Previous research showed that the quality of experience can be positively affected by co-locating an increasing amount of such virtual agents [142]. Therefore, having radiology technologists, anesthesiologists or surgical nurses joining the VE could en-

14 Varjo Technologies Oy, https://www.varjo.com

hance the *Experienced Realism*. Another potential way to increase the visual fidelity is the inclusion of multi-sensory stimuli. So far, only visual and auditory output as well as haptic feedback of the tangible objects have been considered. Doukakis et al. [72] showed that users prioritize visual quality over other sources. However, if more budget is available, a more balanced distribution of auditory, visual and olfactory stimuli may be preferred. In that regard, the inclusion of olfactory and tactile feedback, e.g., wind, elicited a greater *Sense of Presence* in an experiment of Jung et al. [130]. Thus, it would be interesting to also consider MRI intervention specific smell and tactile sensations.

VISUALIZATION AND PRECISION Following Fink et al. [81], obstacles in the form of primitive 3D objects were used as placeholders where real objects are located in EDL, which constrained locomotion in virtual space. This resulted in the inability to define specific medical devices, but it allowed participants to be focused on the execution of the task.

In addition, it remains uncertain if different tasks would have resulted in different outcomes. No differences were identified between IMs in this study. However, tasks requiring more precision could potentially benefit from the more natural tangible interaction paradigm. Future work could therefore repeat the experiment with a different task design.

STUDY SAMPLE The sample size in this experiment was relatively small. A sensitivity power analysis indicated that, with a β of 0.8, effects up to a Cohen's f effect size of 0.245 were likely identified with reasonable probability, corresponding to medium and large-sized effects. Therefore, it cannot be ensured that small-sized effects were not missed.

Additionally, the sample design included only medical students because no interventional expertise was required. However, the digital twin of the MRI room may be experienced differently by clinicians working in ORs or radiology suites on a daily basis due to their prior knowledge. Since they are already accustomed to the environment, they could potentially focus less on their surroundings at higher VF. At the same time, they could notice errors or mismatches that medical students would not see, which in turn could cause distractions. Hence, conducting a similar study with subjects from the expert domain would also be a meaningful continuation of this project.

USAGE OF SOUND A realistic MRI background noise was used as ambient noise in SDL and ADL levels to enhance the feeling of being in an OR. To maintain minimal development requirements, the integration of sound in EDL was decided against. However, this decision may have caused EDL to score less well, as emphasized by Dinh et al.[66], who noted that the use of sound has a greater impact on *Presence* than VF. The effect of contrast between the *Presence* and absence of sound at different visual levels is also supported by Poeschl et al.[208]. Since the audio used was very monotone and no additional audio effects (such as those occurring when interacting with objects) were included, it is suspected that the influence of audio use on *Presence* in this study will be very small.

IMPLEMENTATION A major limitation of this study is that meaningful conclusions on which specific VF aspect was responsible for the observed effects cannot be drawn (see Table 1). For this first experiment, a decision was made to compare holistic impressions instead. Multiple follow-up studies would be required to determine which individual components had stronger effects.

4.2.5 *Conclusion*

This work investigated the effects of three different gradations of VF on *Presence* and UX for a medical task in MR. Tow IMs were investigated: The first one was based on natural hand interaction with tangible, 3D-printed objects and the second paradigm used conventional MR controllers. A control measure confirmed the successful manipulation of *Experienced Realism* with the three investigated *visual fidelities*. The results revealed a strong connection between VF and the dependent measures of *General Presence*, *Involvement*, and UX, while *Spatial Presence* and user performance were less affected. Furthermore, no differences were observed between the two IMs for the examined task. Examination of the raw data indicated that the perception of realism between tangible devices and controllers appeared to depend more on individual user preference.Future work should examine whether these results can be reproduced in tasks that require more precision. In addition, clinical experts should be included in follow-up studies and individual components of the considered VFs could be investigated separately. The identified advantages of high VF in VEs have practical implications for the development of future systems. However, the consideration of development effort in future work might become redundant, as the manipulation of the level of detail using technologies such as AI-based filters and 3D model conversion could be accomplished without additional effort. Investigations of interaction paradigms in MR seem less dependent on the visual quality of the VE and do not seem to require high standards. However, experiments focused on MR related qualia, such as *Presence*, can benefit from extensive efforts towards realism.

4.3 ONGOING WORK

In this final section, an extended experimental setup is introduced, and a concept is outlined that provides a forward-looking perspective on future experiments, building on previous research. This addition aligns with the chapter's primary goal of exploring the fundamentals of design and illustrating potential future directions.

For a planned study, the VE discussed in this chapter has been redesigned, and the use case expanded. This will allow for the collection of additional requirements and insights into the development processes of medical task simulation. Some potential research questions to be explored include:

- Which combination of VF and IM is comparable to the real-world condition?
- What effect do the respective factors VF and IM have on users' performance scores when completing the task?
- To what extent is performance influenced by secondary factors such as UX, *Presence*, and *Subjective Workload*?

4.3.1 *Interaction and Visualization Rationale*

As concluded earlier in the previous experiment, it was observed that VF led to a greater *Sense of Presence*, while user performance seemed to be less affected. It can be speculated that this may have been due to the basic procedural task itself, as it required less precision, which is why the upcoming experiment have to consider an interaction that requires more precision. Within the workflow described in Section 4.2.2.6, the task of needle insertion was extracted for the follow-up experiment because it requires precise instrument handling and critical haptic feedback, essential for training practitioners to perform these delicate and technically demanding procedures safely and effectively (Figure 5). In addition, the step of sterilizing the puncture site was added as a preparatory measure to introduce an additional interactive element and extend the time spent in MR, ensuring that the user has sufficient time to fully engage with and experience the simulation.

INPUT CHOICE Also, in the previous experiment, no effects of the IMs on the evaluated factors were found. This could be due to the comparison of two tactile input methods (controller vs. tangible objects). For this reason, the future experiment will focus on the comparison between tangible and virtual objects.

FIDELITY DEGREES According to the findings from the first experiment, the impact on variables such as user performance, *Presence*, and

UX was more pronounced when comparing SDL to EDL than to ADL. Specifically, participants experienced a clearer distinction in terms of realism, involvement, and *Presence* between EDL and SDL, while the differences between SDL and ADL were relatively smaller. In the follow-up study, the focus will be on the extremes of the VFs, which is why only SDL and ADL will be examined. Focusing on the extremes provides clearer insights into how different levels of VF impact users, particularly in terms of *Presence*, UX, and task performance. This could help identify the minimum level of VF required to achieve optimal results in user performance and experience without complicating the design process. It will also allow a better understanding of the effects of both low and high VF on user performance and experience.

GLOVES In the upcoming study, the visualization of hands will be implemented with surgical gloves, addressing both of the aforementioned points. By using gloves, we aim to minimize the impact of individual variations in hand size, shape, and skin tone, thus enhancing inclusivity and reducing potential biases in VBO. This approach will also maintain the realism expected in medical environments, where wearing gloves is a standard practice for maintaining sterility.

TASK EXPANSION The task will be as follows: Within this apparatus, we implemented a navigated needle insertion task. A custom laser system, mounted on a ring for full mobility and adjustable at various angles, is used to mark the entry site on the abdominal area of the virtual patient. The user must first grab the swab on the nearby table and sterilize the marked area, repeating this process three times. Afterward, the swab is placed back in the iodine bowl, and the needle is picked up. The user aligns the needle tip and end with the laser, ensuring the correct insertion angle. The needle must then be inserted into the body until a specified depth is reached.

Once again, emphasis was placed on the rapid prototyping approach and cost efficiency. To make these environments feasible with conventional MR hardware, 3D-printed and tracked instruments will also be utilized. Figure 12 shows screenshots of the already implemented features, the execution of the task including real and virtual replicas, as well as the visualization levels.

4.3.2 *Apparatus Design*

The setup will be located in the same lab as described in Section 4.2.2.5. However, the room has since been subdivided into an examination room (approximately 33 m^2), a control room (approximately 8 m^2), and an observation room (approximately 9 m^2). The latter two rooms are connected to the examination one by doors and half-mirrored windows. The experiment will also be conducted using

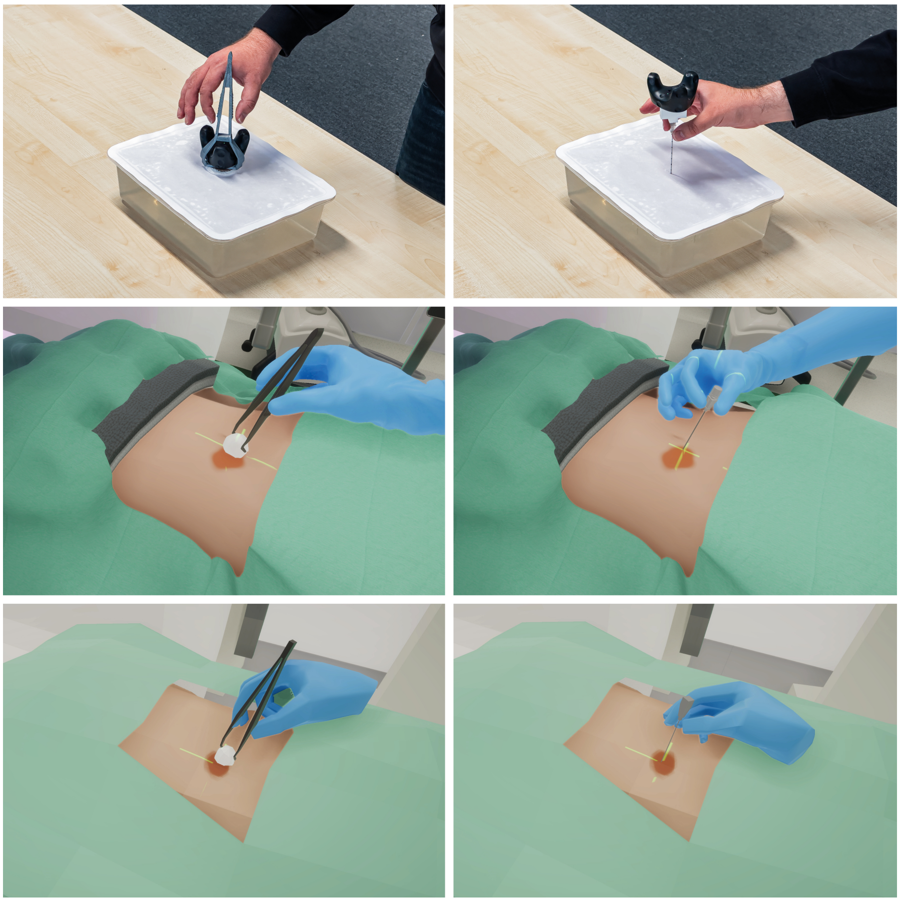
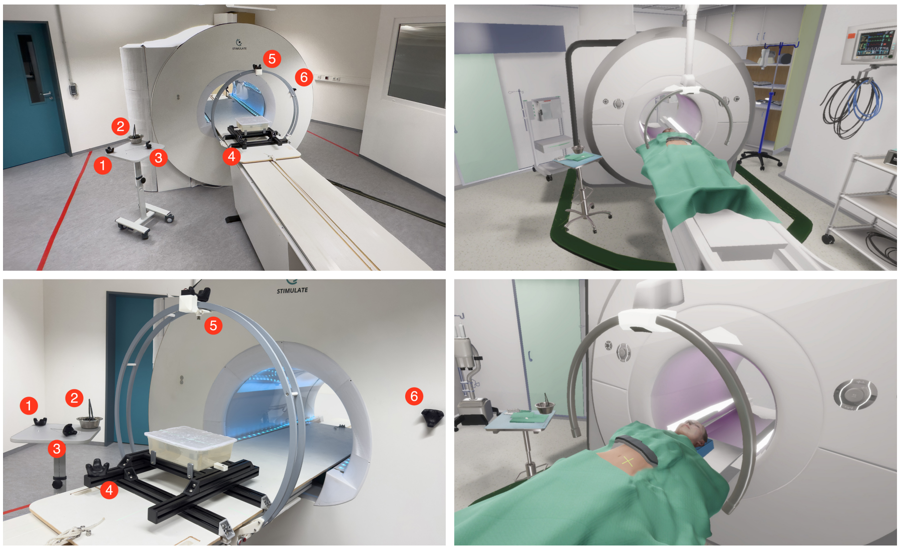


Figure 12: Executed Needle Insertion Task. The user first sterilizes the laser-marked area (left) and then inserts a needle at a predefined angle and depth towards the target structure (right). The top row displays the 3D-printed tangible objects equipped with trackers. The middle row shows the task in high-fidelity MR, while the bottom row depicts it in low-fidelity MR.

the same mock-up described in Section 4.2.2.5; however, it has since been reinforced for added stability and enhanced with a movable patient bed operated by an electric pulley. Specifically for the following experiment, the mock-up was enhanced with a custom-built laser navigation unit, flexibly mounted on a 90 cm ring profile. This system, serving as a navigation aid, is inspired by a mobile system for image-guided interventions[15] and is designed to help participants find the needle insertion point and the corresponding orientation. This allows for an expansion of the sample size, as it enables participation by individuals other than just experienced interventionists. In total, the setup has been expanded with six additional trackers to correctly align the position of several physical objects with their virtual twins. The laboratory is equipped with five Valve Basestations[16], creating a tracking space that covers the entire MRI working area and the control room.

15 ATLAS medical Technologies GmbH, `https://www.atlas4d.de`
16 Valve Corporation, `https://www.valvesoftware.com/en/index`



REAL VIRTUAL

Figure 13: Overview of the apparatus. Physical mock-up of the MRI device and interventional instruments equipped with HTC Vive trackers (left) and corresponding virtual replica of the environment (right). The tracked objects are: (1) instrument table, (2) swab in a fixed bowl, (3) biopsy needle, (4) patient phantom, (5) laser system, and (6) MRI device.

Figure 13 gives an overview of the system by comparing the real apparatus with the virtual twin.

4.3.3 *Planned Variables*

Finally, the study will be expanded to include a real-world condition, which will not be conducted in MR. Instead, the tracked instruments will be used within a real interventional mock-up to perform the same tasks, serving as a ground-truth measurement. The study design has already received ethical approval from our Medical Faculty's Ethics Committee. The study will include the following measurements to address the questions mentioned above:

- Demographic data (age, gender, prior experience)
- UX (UEQ)
- *Subjective Workload* (N-TLX)
- *Sense of Presence* (IPQ)
- Subjective evaluation of visualization levels and IMs (rating)

The performance metrics to be determined include the measurement of TCT (start, stop, task duration) as well as the puncture error (deviation of the achieved puncture depth and angle from the predefined path).

4.4 CHAPTER SUMMARY

This chapter explored the impact of manipulating various visual and interactive factors on user performance and engagement within a medical task simulation. The findings revealed a significant relationship between VF and the measured outcomes, while no differences were observed between the IMs. These results prompted the development of a follow-up evaluation concept to refine the interaction approach further. Although the described environment centers on a medical task simulation, its adaptability extends to broader training applications. Thus, the developed simulation serves as a flexible tool, functioning as both a testing environment for prototypical evaluations and a training platform for medical education scenarios. This chapter enhances the understanding of choosing different components in the development of VEs using MR technology and paves the way for the subsequent chapters of this work, which focus on practical learning applications. The next chapter centers on the design and evaluation of applications aimed at individual knowledge construction in the field of anatomy education.

INDIVIDUAL LEARNING ENVIRONMENTS FOR EMBRYONIC HEART EDUCATION

ch. 5

SYNOPSIS *This chapter introduces a VLE aimed at supporting early-stage medical education, helping students better understand the dynamic morphological changes that occur during embryonic heart development. By leveraging the interactivity and immersive capabilities of MR technology, the VLE seeks to make these complex processes more accessible. The chapter begins with an overview of the medical background, current teaching methods, and the transition to 4D models, followed by an exploration of the application's key features and evaluation. A subsequent experiment further develops the application and methods, assessing its effectiveness and investigating questions related to its impact on learning outcomes and influential factors in MR system characteristics.*

ABOUT THIS CHAPTER *Parts of this chapter were previously published in Schott et al., "Cardiogenesis4D: Interactive Morphological Transitions of Embryonic Heart Development in a Virtual Learning Environment" [Core6] and have been incorporated into this thesis. The article is open access under a Creative Commons Attribution 4.0 International License. Additionally, some of the methods described in this chapter were developed as part of Mr. Tom Wunderling's Master's thesis, titled "Interaktive Visualisierung der embryonalen Herzentwicklung in einer Virtual-Reality-Lernapplikation," which was supervised by myself and M. Kunz. Furthermore, sections in "Experiment 3" were previously published in Kunz and Schott et al. (shared first authorship), "Embryonic Heart Development as an Immersive Experience: Unveiling Learning Effects and Influential Factors in Virtual Learning Environments" [Core3], and are reused here in accordance with the copyright policy of Elsevier. ©2025 The Authors. Published by Elsevier Ltd.*

MY CONTRIBUTION *For "Experiment 2," I developed the primary research idea and refined the initial concept formulated in Mr. Wunderling's Master's project by expanding its theoretical foundation and conducting an extensive literature review. I also co-managed the project, co-designed the study, and took primary responsibility for drafting, reviewing, and editing the original manuscript. I supervised and actively contributed to the creation of graphics, including new photos, tables, illustrations, and additional content specifically tailored for this thesis. For "Experiment 3," I co-created both the research concept and the implementation of the methodological approach, including software prototyping, interface design, and visualization. I co-authored the manuscript, contributing to drafting, reviewing, and editing, and collaborated on data collection and analysis. I was also responsible for creating new graphics, photos, tables, illustrations, and other thesis-specific content.*

5.1 INTRODUCTION

The previous chapter provided insights into the impact of manipulating visual and interactive factors on user engagement in VEs within a more general medical context. This discussion contributes to a deeper understanding of how different components can be selected for developing MR applications, paving the way for the next chapter, which focuses on a practical learning application. In particular, this chapter delves into the development of a system for anatomical education, using the example of embryonic heart development.

Embryology, the branch of anatomy that studies the human development in the womb, offers insights into the form and function of anatomical structures as they develop. This knowledge forms an essential foundation of medical education and is key to understanding congenital diseases. Comprehending embryology—especially the complexities of embryogenesis and organogenesis—is challenging due to the lack of three-dimensional orientation [201]. This difficulty arises from the simultaneous growth processes and rapid shape changes that occur within a short timeframe [1, 45]. For example, the embryonic heart undergoes complex morphological transitions, reshaping multiple times in a few days. Understanding these processes is crucial for students and future cardiologists to accurately diagnose and treat congenital heart defects, as the origins of many pathologies lie within these transitional stages and their vulnerabilities.

As highlighted in Chapter 2, traditional learning tools, such as 2D illustrations, videos, and static 3D models, are commonly used in embryology education. However, new teaching methods are required to effectively capture the dynamic internal and external developments inherent in this process, as discussed in Section 3.3 and further explored in Section 3.5.

As also previously discussed, the high level of immersion and enhanced depth perception achievable with MR can create engaging and impactful learning experiences—well-suited to the study of embryology. While anatomy is a fundamental subject that students find both exciting and practical, particularly during dissection courses, embryology lacks similar popularity due to its abstract nature, driven by the small scale and limited visualization formats of the developing structures. Additionally, embryology occupies only a small portion of the curriculum, with limited emphasis on the cardiovascular system in particular.

Therefore, this chapter presents the development and evaluation of VLEs designed to enhance the understanding of embryonic heart development and create an engaging experience for this complex subject. The MR application developed is referred to as *CardioGenesis4D*, named after the original paper [Core6]. Its development was shaped by interdisciplinary collaboration with anatomy experts and cardiologists from our university, who provided interviews, application-specific feedback, and guidance throughout the process.

In the first part of this chapter—*Experiment 2*—the foundational VLE is introduced, based on the original paper as previously mentioned. This initial phase established the conceptual and technical groundwork, addressing individual learning styles by implementing various features. Through advanced interactions and dynamic visualizations, users can explore 4D morphological changes via deformable organ models, thereby gaining a deeper understanding of temporal evolution and supporting individual knowledge construction.

In this experiment, creating a user-friendly foundation was essential before investigating the actual effects on learning behavior. Consequently, this initial study focuses primarily on aspects related to *Usability* and *Presence*. Therefore, both experienced and inexperienced users of MR were included as well as the target group of medical students. Learning aspects were considered only secondarily in a smaller sample group.

To summarize, this experiment addresses the following research question:

> **RQ2** | How can suitable visualizations and interactions in MR be designed to effectively represent embryonic heart development?

The second part of this chapter presents a further development of *CardioGenesis4D*, referred to as *Experiment 3*. Based on the findings from *Experiment 2*, the application was significantly improved, with expanded content and enhanced functionality.

The goal of the study in *Experiment 3* was to investigate how a single exposure to the VLE might influence knowledge retention of a previously learned topic. Consequently, the study examined long-term knowledge retention, UX aspects, and their impact on learning effectiveness. This evaluation included the development of a specialized knowledge test, modeled on real exams, to assess these factors. Additionally, the use of the MR application during the exam period for medical students was explored. Within this context, the following research question is addressed:

> **RQ3** | Are there measurable learning effects when using MR to understand embryonic heart development, and which factors influence these outcomes?

A video demonstration of the applications is linked in the Appendix (see Section 9.1).

5.2 EXPERIMENT 2

This section presents the first experiment exploring the use of an MR application to enhance the learning experience in anatomy education, specifically focusing on the development of the embryonic human heart.

5.2.1 *Material & Methods*

In this part, the development process of the VLE is outlined based on user research. The context of use is discussed, and key requirements are highlighted. Finally, the resulting prototype, user interactions, technical specifications, and defined objectives for a learning scenario are presented.

5.2.1.1 *User Research*

As discussed in Chapter 2, the foundation of HCD lies in analyzing the context, ergonomic considerations, and identifying specific user requirements. Consequently, requirements for teaching and learning were identified based on three semi-structured expert interviews with lecturers and physicians from our medical faculty (Institute of Anatomy and Clinic for Cardiology and Angiology, University of Magdeburg), which informed the design of a prototype with an appropriate interaction concept. Additionally, a survey of medical students was conducted to determine which aspects of cardiac development were particularly challenging to understand. Followed by an analysis of the current teaching situation at our university and an assessment of the students' difficulties, ideas and approaches for technical implementation were designed by an interdisciplinary team of software developers, designers, and medical professionals, based in these requirements.

5.2.1.2 *Context of Use*

The term embryo refers to a stage from fertilization to the end of the eighth week of pregnancy (day 56). The cardiovascular system is the first functional system in the embryo, with the heartbeat detectable by ultrasound as early as the 4th week of development. Heart development (cardiogenesis) can be roughly divided into two stages: formation of the cardiac loop (1) and formation of the interior of the heart (2).

At our university, embryology is taught in the first year of the human medicine program, both through lectures and as a seminar within the dissection course. During lectures, images and videos from various sources are presented, with references to further resources on the intranet or recommended literature for self-study. The embryology of the heart is addressed in small groups of 5-10 students during a half-hour breakout session in a dissection room. Feedback from a survey of medical students - which was conducted as part of this research project – indicated that the formation of the primary heart tube and cardiac looping are perceived as particularly challenging topics. In the teaching of embryology, two aspects are considered especially important by interviewed lecturers:

- **(1)** Understanding complex, time-dependent, 3D morphological changes in the developing embryo.

- **(2)** Attention to the clinical implications of congenital organ defects that arise when these morphological processes deviate from the norm.

5.2.1.3 *Concept*

The focus was on spatial understanding of shape transformation. The formation of the heart loop involves a 3D overlay, making it challenging to comprehend in two dimensions. As a result, lecturers have developed a method using different colored modeling clay to simplify the fusion into the primitive heart tube and subsequent loop formation (see Figure 14). This method also helps convey the changing positional relationships after each formation step. Drawing inspiration from this practice, the advantages of MR were leveraged to display the shape changes immersively. Unlike existing static models, the resulting 4D concept is designed to animate shape and positional changes through temporal transitions between developmental stages, creating an interactive experience.

5.2.1.4 *Virtual Learning Environment*

Based on the previously described concept and in continuous collaboration with cardiologists, a virtual learning application was developed through an iterative design process, with a focus on interactive visualization for learning the temporal evolution of the human heart. A virtual seminar room is proposed, allowing users to explore heart development and manipulate transitions using hand interactions. In this VLE, deformable 3D models of the early heart can be examined at different stages to enhance spatial understanding of the developmental process. Three phases of early heart development have been implemented:

1. Formation of the primitive heart tube (day 20 - 21)
2. Formation of the cardiac loop (day 22 - 23)
3. Completion of the S-shaped heart loop (day 24 - 35)

INTERACTIVE MODES In addition to the 3D models of the different stages of the heart, the VLE includes four modes. The first mode, *Training Mode*, is designed to familiarize users with MR interaction techniques, particularly to help inexperienced students get started easily. Basic 2D and 3D elements such as sliders, buttons, and toggle switches are presented here. The second mode, *Exploration Mode*, displays segmented 3D models of corresponding embryos. In addition to the isolated heart, other anatomical structures and organs can be toggled on and off to provide contextual visualization. The central

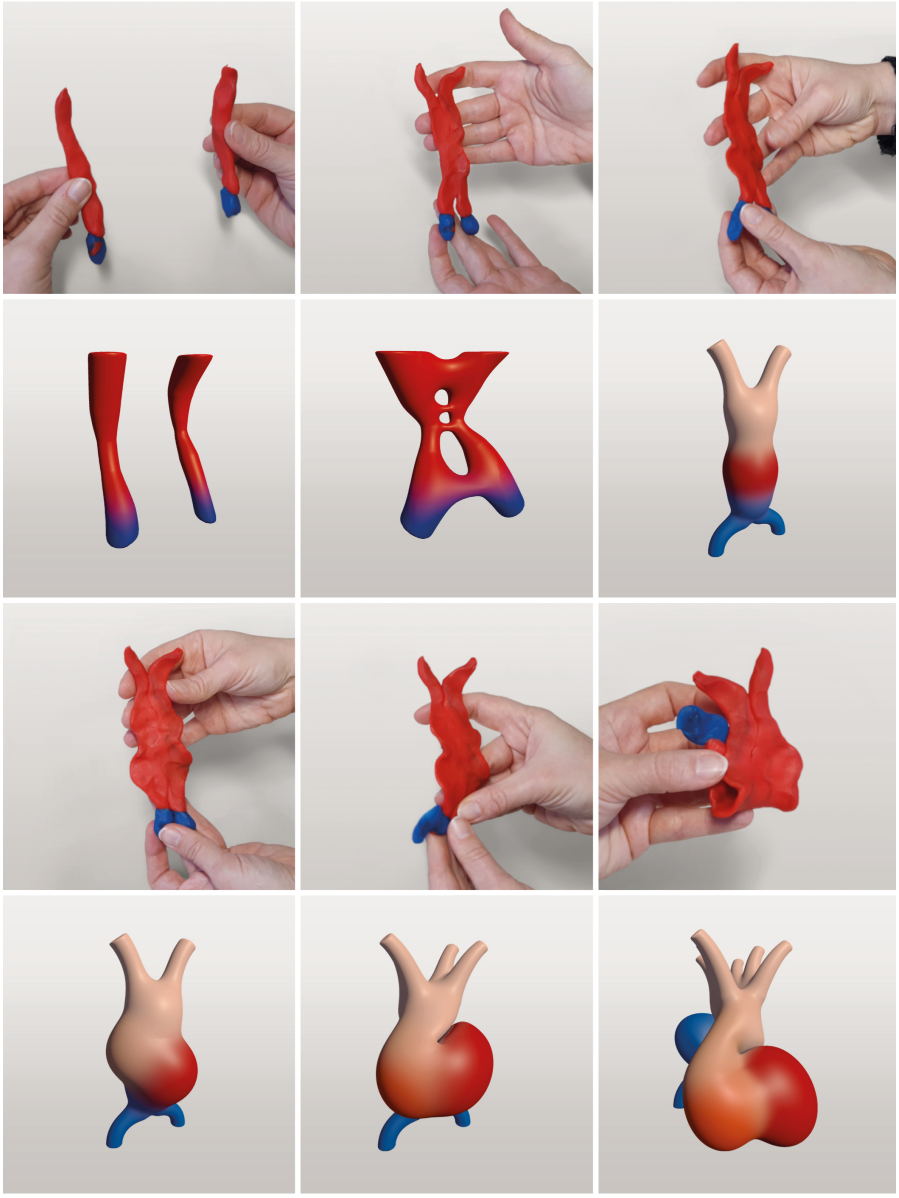


Figure 14: Comparative view of the physical and virtual representations of the three stages of heart development as implemented in the VLE. Rows 1 and 3: Modeling clay representations used in the classroom. Rows 2 and 4: Corresponding virtual 3D models derived from the clay models. Adapted from Schott et al. (2023b) [Core6], rearranged for this thesis. Licensed under CC BY 4.0.

element is the third mode, *Learning Mode*, intended as a training unit for interactively simulating 4D morphological development and promoting spatial understanding. To support this, 2D diagrams with cues and labels are provided, along with various interaction elements to manipulate the heart model. An overview of *Learning Mode* is shown in Figure 17. A final fourth mode, *Modeling Mode*, has been implemented to allow free modeling of a heart tube (see Figure 16 (a)).

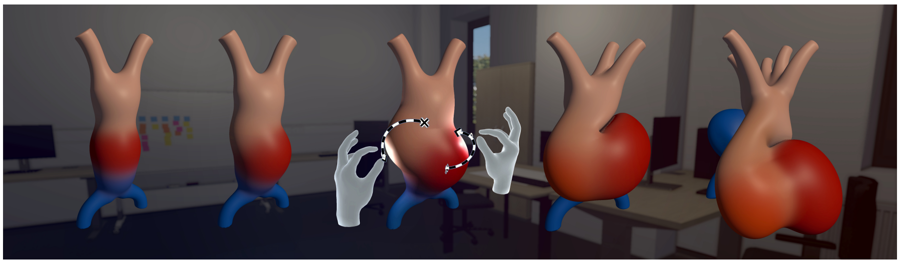
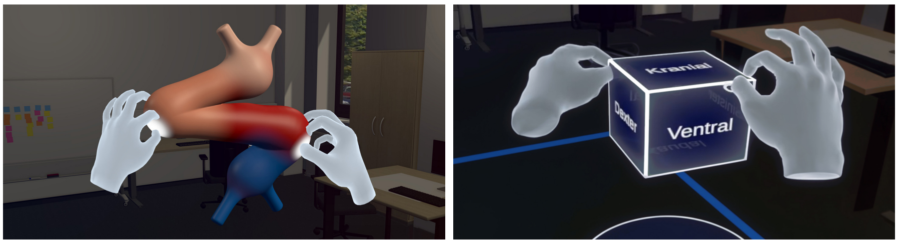


Figure 15: A collage of multiple stages of the transformation of the heart tube during embryological cardiogenesis as seen in the VLE. The center heart shows the transformation in progress: the tracked hands of the user follow the arrows indicating growth directions. Reprinted from Schott et al. (2023b) [Core6]. Licensed under CC BY 4.0.



(a) (b)

Figure 16: The modeling mode allows the user to deform the heart tube freely and without restrictions (a). Reprinted from Schott et al. (2023b) [Core6]. Licensed under CC BY 4.0. The *Navigation Cube* widget is used to rotate and scale the model in alignment with the anatomical position (b).

USER INTERFACE To assist the user in transitioning from the initial state (A) to the final state (B) of morphological change within a developmental step, several orientation and explanatory tools are provided, as shown in Figure 17. On the user's left side in the VE, an illustration[1] describes the current development stage. The two states (A + B) are marked with the respective developmental day and anatomical labels. Below the 3D model in the center, a playback control (play | pause | toggle) activates the change as an animated motion.

Simultaneously, a progress bar displays the current stage of development over time as a percentage (0–100) and can also be adjusted directly by the user, who interacts with a handle. Additionally, the playback speed of the animation can be modified in discrete steps using a toggle switch, and the animation can be paused at any point, allowing the user time to scale, rotate, or move around the 3D model. The 3D models can be rotated and scaled via a widget—a *Navigation Cube*—that displays the anatomical orientation (see Figure 16 (b)). Rotation is restricted to the cranial axis, while scaling is unrestricted.

1 Used under CC BY 4.0: `https://openstax.org/books/anatomy-and-physiology/pages/19-5-development-of-the-heart`

When multiple 3D models are displayed, manipulation of this cube affects all models synchronously. A button advances the user to the next of the three developmental stages.

To focus attention on the learning content, the entire UI is separated from the environment. The background is dimmed and draws less attention due to a semi-transparent box placed around the user, with the controls positioned in a recessed area.

HAND INTERACTION To make the interaction as natural as possible, the model and corresponding structures can be deformed by users with their hands. This is done by grabbing the virtual heart with both hands at points marked with a cross. Additional navigational aids are provided to facilitate guided exploration in MR, emphasizing feasible interaction with the 3D model [46]. Two black and white dashed arrows on the virtual heart indicate the required movement to transform the object from one state to the next (see Figure 15).

Analogous to the animation, only a linear deformation can be performed. Thus, free interaction is imitated by guided movement, preventing unwanted deformations and ensuring that the real physiological development is accurately represented. The interaction is designed to be bi-manual, as each hand mimics a part of the complex movement at specific points on the model, reflecting the parallel processes that occur in reality. This guided linear control of the interaction was selected because it aligns with physiological development. This motorized movement is intended to support learning and consolidate knowledge. According to the theory of embodied cognition, motor movements and accompanying sensory experiences can influence cognitive processes, thereby making abstract concepts more understandable [3]. Additionally, a fourth mode for free deformation was implemented. This experimental function imposes no restriction on the degrees of freedom, allowing any kind of virtual shaping to be performed as a kind of creative approach.

ACTIVE AND PASSIVE FEATURES Although active interaction with 3D models in a VE leads to more efficient object viewing [120], individual learning styles are not accounted for. To address this, two features with different interaction modalities were implemented to cater to the target group's individual learning preferences and to explore the advantages and disadvantages of these features. Consequently, the *Learning Mode* can be completed in both a *passive*-explicit and an *active*-implicit variant. The distinction lies in the UI: in *passive* training, classic UI elements such as buttons and sliders are available to control the animation. In *active* training, the layout remains identical, but playback control buttons are hidden, and interaction with the progress bar is disabled. For manipulation, only direct hand interac-

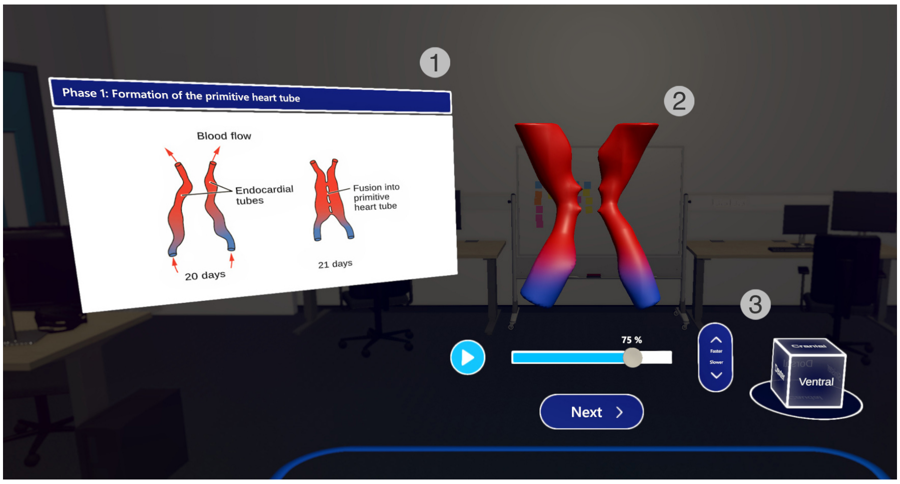


Figure 17: VLE in *Learning Mode* with *passive* features. (1) 2D graphical scheme with descriptions, (2) 3D model in current transition, (3) UI elements for manipulating the animation: buttons, slider and *Navigation Cube* (right). Reprinted from Schott et al. (2023b) [Core6]. Licensed under CC BY 4.0.

tion with the 3D model is possible. Once the task in *active* training is completed, the animation can be reset or replayed using a button.

TECHNICAL SETUP Software development was conducted using Unity 2020.3.35f1[2] in C# on Microsoft Windows 10 Pro (build 19044)[3]. The Mixed Reality Toolkit (MRTK) for Unity (v2.7.2)[4] was used to implement hand tracking, which was integrated via Ultraleap Hand Tracking v5.2.0[5] and the Leap Motion Core Unity Plugin v4.9.1[6]. Interface elements were created using MRTK prebuilt UI components.

The interactive models of the embryonic heart were created using MudBun v1.3.29[7]. MudBun employs signed distance fields for procedural volumetric mesh generation, enabling real-time modeling of organic shapes directly within the Unity editor and during run-time. For polygonizing the 3D scalar field, several meshing algorithms were evaluated. The Marching Cubes algorithm delivered the best balance between visual quality and performance. To maintain an acceptable frame rate (approximately 90 FPS) for MR, spatial hashing and sparse voxel trees were employed.

On the hardware side, the prototype was implemented using a Valve Index HMD[8], extended at the front with Ultraleap's Stereo IR

2 Unity Technologies, `https://unity.com`
3 Microsoft Corporation, `https://www.microsoft.com`
4 Microsoft Corporation, `https://learn.microsoft.com/windows/mixed-reality/mrtk-unity`
5 Ultraleap Ltd., `https://www.ultraleap.com`
6 Ultraleap Ltd., `https://developer.leapmotion.com/unity`
7 Long Bunny Labs, `http://longbunnylabs.com/mudbun`
8 Valve Corporation, `https://www.valvesoftware.com`

170 Evaluation Kit[9] mounted via a custom 3D-printed adapter to enable touchless interaction. The PC used featured the following specifications: Intel Core i7-9700 3.00 GHz 8-core CPU[10], NVIDIA GeForce RTX 2080 Ti[11], 32 GB RAM, and a 512 GB PCIE 3.0 SSD.

5.2.1.5 *Learning Objectives*

The VLE was designed to enable students to explore theoretical knowledge about the development of the early human heart in an exploratory manner based on a problem-based learning approach [109]. This MR application is intended for self-study and individual knowledge construction, for example, following initial lectures on heart development. The VLE was created as a controlled space in which users can focus on essential learning content without external influences. The learning experience is motivated by a sense of embodiment [156], emphasizing natural hand interactions. Literature and interview insights indicated that hands-on interaction, such as palpating organs, enhances spatial understanding and comprehension of positional relationships. Additionally, the design aimed to accommodate different learning preferences [149], enabling both *active* and *passive* learning of 3D objects [120] through various features and offering guided exploration with an appropriate didactic structure [46].

5.2.2 *Evaluation*

To accommodate different types of learners and explore potential learning contexts, various functionalities were implemented in the prototype to facilitate exploration of the early human heart within a VLE. This VLE was evaluated in terms of *Usability*, *Subjective Workload*, and *Presence* through quantitative measurements and qualitative feedback from the potential target audience: medical students. Spatial awareness and knowledge gain were also assessed using a self-developed test. Additionally, the prototype was presented to a group of non-medical professionals with technical backgrounds and a high affinity for MR to evaluate the VLEs *Usability* and *Presence* for non-domain experts; this group was not questioned on the medical quality of the VLE. Finally, the application was presented to three domain experts—two practicing cardiologists and an anatomy lecturer uninvolved in the VLE development—to obtain feedback on the test, advice on the appropriateness of use scenarios, and recommendations for integrating the VLE into a future curriculum. The evaluation, based primarily on qualitative data collection, thus adopts an exploratory approach, enabling iterative improvement of the application based on stakeholder feedback.

9 Ultraleap Ltd., https://www.ultraleap.com
10 Intel Corporation, https://www.intel.com
11 NVIDIA Corporation, https://www.nvidia.com

5.2.2.1 *Study Design*

The group of medical students was divided into two test groups: Group Active (A) began the experiment with the *active* feature as the initial condition, while Group Passive (P) started with the *passive* feature in *Learning Mode*. Since Modes 2–4 serve as the central learning tools, the investigation primarily focused on these elements within the application. *Usability* and *Presence* of the VLE were evaluated using the SUS and IPQ. To assess *Subjective Workload* of the interaction concepts, the overall unweighted N-TLX score (Raw-N-TLX) was analyzed. Based on the TAP, participants were asked to articulate their thoughts while interacting with the VLE. Additionally, a semi-structured interview was conducted at the conclusion of the experimental series. To assess spatial understanding and rotational ability within the sample, the MRT described in Section 2.4.2.2 was utilized (also see an excerpt in Section 9.4).

ANATOMY KNOWLEDGE TEST To establish an indicator of knowledge transfer for the VLE and assess its effectiveness as a learning tool, an Anatomy Knowledge Test (AKT) was developed in collaboration with anatomy experts, consisting of 30 questions. The test was divided into three categories, each focusing on aspects critical to understanding embryonic heart development. In the *Shape* category (15 questions), participants were tasked with identifying illustrations that did not correspond to the anatomical structures from the training phase. Discrepancies were introduced by mirroring, omitting, or distorting sections. In the *Time* category (4 questions), participants arranged six illustrations in the correct chronological order. In the final category, *Location* (11 questions), participants identified a cube (analogous to the *Navigation Cube* from the application) from four possible options that matched the illustration shown. Each category included both colored and colorless images. The test is provided in Section 9.3.

5.2.2.2 *Participants*

To evaluate the two implemented modes, 19 participants were recruited. Twelve medical students (8 women), aged 24–30 years (Mdn = 27), were recruited from our university and compensated with 20 euros per hour. All medical participants had previously taken anatomy courses; 11 were in their 6th to 15th semesters of study (Mdn = 10), and one was a graduate student. Seven individuals reported having corrected visual impairments. A rating scale from 1 (no experience) to 5 (very experienced) was used to assess the following: technical affinity (Mdn = 3), MR experience (Mdn = 2), gaming experience (Mdn = 2.5), and prior knowledge of embryology (Mdn = 2.5). Three participants reported no prior MR experience, and none reported color vision deficiency. Most participants indicated that their last contact with embryology had been several years ago.

The second group included seven unpaid non-medical students with technical backgrounds (3 women), aged 22–30 years (Mdn = 27.5), also recruited from our university. Four participants in this group reported having corrected visual impairments. Using the same rating scale, the following ratings were obtained: technical affinity (Mdn = 4), MR experience (Mdn = 3), gaming experience (Mdn = 4), and prior knowledge of embryology (Mdn = 1). Five participants rated their technical affinity as 4 or higher, and one participant reported no prior MR experience.

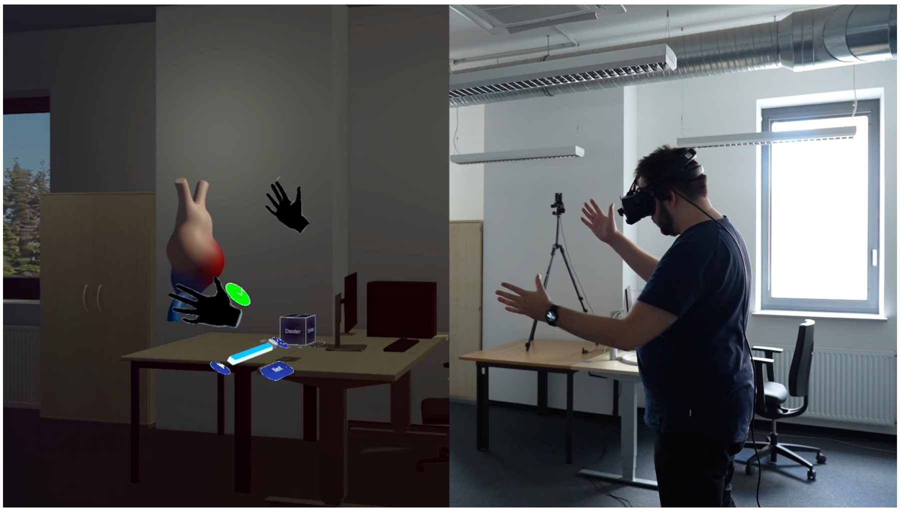


Figure 18: User interacting in the VLE as seen in virtual replica of our laboratory (left) and real lab (right).

5.2.2.3 *Setup*

The study was conducted in a 70 m^2 MR lab at our university. During each data collection session, one participant and one investigator were present in the room. A designated area was set up for the participant, equipped with a monitor, keyboard, and HMD, positioned next to the experimenter's recording station. Inspired by Voit et al. [309], who investigated UX and *Presence* in both physical in-situ and virtual lab environments, reporting comparable results regarding user insights and feedback, a virtual replica of the laboratory was created (see Figure 18). This safe and controlled space served as an orientation point for users in the real world and was designed to function as a proxy for a seminar room. The virtual space, allowing user movement, measured approximately $3 \text{ m} \times 4 \text{ m}$ and was bounded by a virtual blue line.

5.2.2.4 *Ethics Approval & Hygiene Policy*

The experiment was conducted during the global SARS-CoV-2 pandemic. Therefore, a strict hygiene policy was followed. Specific ethics approval was not required by our institution. The occupational safety and health department approved all relevant measures.

Table 5: Summary of descriptive results with respect to demographics. All entities are in the format: mean value ± standard deviation (median). Reprinted from Schott et al. (2023b) [Core6]. Licensed under CC BY 4.0.

Variable	AKT	Raw-TLX	SUS
Active	0.63 ± 0.17 (0.68)	17.15 ± 11.42 (13.30)	90.00 ± 9.19 (92.00)
Technical	0.54 ± 0.26 (0.50)	13.21 ± 6.51 (10.80)	94.86 ± 4.14 (96.00)
Medical	0.68 ± 0.11 (0.71)	19.44 ± 13.22 (16.65)	87.17 ± 10.25 (89.00)
Passive	0.74 ± 0.05 (0.72)	12.89 ± 9.50 (10.00)	92.74 ± 5.59 (94.00)
Technical	0.74 ± 0.05 (0.73)	9.16 ± 5.78 (8.30)	95.43 ± 3.95 (94.00)
Medical	0.74 ± 0.05 (0.72)	15.07 ± 10.74 (11.25)	91.17 ± 5.94 (93.00)
First Training	-	14.99 ± 10.40 (10.80)	90.32 ± 8.88 (92.00)
Active	-	18.42 ± 10.17 (15.80)	87.56 ± 11.26 (90.00)
Passive	-	11.91 ± 10.10 (8.30)	92.80 ± 5.51 (94.00)
Second Training	-	15.05 ± 11.05 (11.70)	92.42 ± 6.20 (94.00)
Active	-	16.00 ± 12.88 (12.05)	92.20 ± 6.70 (95.00)
Passive	-	13.99 ± 9.26 (11.70)	92.67 ± 6.00 (94.00)

5.2.2.5 *Procedure*

The user study lasted approximately 90 minutes. At the start, all participants were informed about data protection, hygiene regulations, and the study procedure. Demographic data were collected, followed by a mental rotation test on the participant's screen using keyboard input, with the screen content mirrored on the investigator's monitor. Once the participant entered a marked area in the center of the tracking space, the HMD was placed on and individually adjusted, followed by a brief introduction to the VE. A reading test with small text in MR was conducted to assess visual performance.

Participants were then immersed in the virtual lab and initially encountered a tutorial comprising elements from the MRTK framework (*Training Mode*). By interacting with 2D and 3D interfaces such as sliders and buttons, they familiarized themselves with hand-tracking functions. After completing this introduction, a button directed them to *Exploration Mode*, where they could examine the developmental stages of embryos, with a primary focus on the cardiovascular system. By pressing a start button, participants then initialized the central *Learning Mode*. Test groups A and B first engaged with the assigned variant of *Learning Mode* and were asked to comment on their activities. The task involved moving the virtual object from the initial state (A) to the final state (B) using a specified interaction. The task was marked as complete when 100% of the required movement was achieved, with a green checkmark and audio signal indicating successful completion. In this mode, a reset button was available for interactive tasks, while the "Next" button allowed progression to the next developmental stage.

After completing the initial training phase, participants were asked to fill out the N-TLX and SUS questionnaires and complete the AKT. They then repeated the training with the alternate variant, followed by completion of the questionnaires and a second variant of the AKT, with the same tasks presented in a different order. The experimental *Modelling Mode* for free-form deformation was then made available for exploration. Upon finishing all training phases, participants completed the IPQ, and data collection concluded with a final semi-structured interview, allowing participants to provide feedback on the VLE experience.

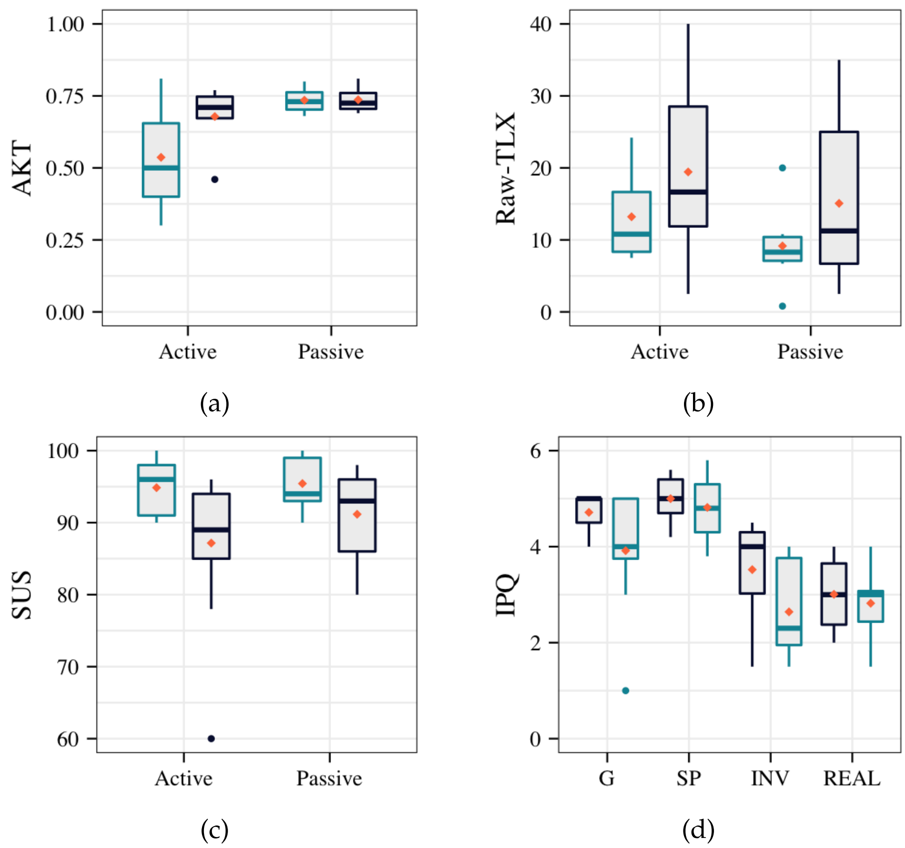


Figure 19: Descriptive results of *active* and *passive* features on AKT, Raw-N-TLX, and SUS scores and descriptive IPQ results with respect to participant background. ■ represents Technical background, while ■ represents Medical background. ♦ indicate mean values. The scales of the Raw-N-TLX and SUS scores were truncated after 40 (max. 100) and before 60 (min. 0) respectively for better visibility of differences. Adapted from Schott et al. (2023b) [Core6]. Licensed under CC BY 4.0.

5.2.2.6 *Deviations from the Protocol*

The same methods and procedures were used to collect data from participants with a technical background. Due to the lack of medical qualifications within this group, data analysis focused on the technical aspects of the VLE. As a result, qualitative feedback was not collected or analyzed. The three experts invited for interviews did not follow the study design but were instead asked to explore the application and its functions independently and to describe their impressions. The experts' statements were subsequently discussed, documented, and evaluated. Their feedback was interpreted to assess the suitability of the VLE within an educational context.

5.2.2.7 *Data Analysis*

Individual statements from study participants with medical backgrounds were summarized into specific categories in tabular form. These categories included feedback on the UI, interaction with the 3D models, cardiac development, anatomical knowledge, and the VLE in general. For the descriptive analyses, mean values and medians were calculated for all measured variables, differentiating between participants from technical and medical domains. For the MRT, the error rate (number of errors divided by the total number of items) was multiplied by the average response time for each participant to derive a single mental rotation measure, with a lower score indicating better performance. Finally, post-hoc analyses were performed to examine the correlation between performance and mental rotation capabilities.

5.2.3 *Results*

The primary goal of this evaluation was to assess the *Usability* and *Presence* of the VLE. The results were intended to provide insights into the effectiveness of the implemented functions and the *active* and *passive* interactions. Therefore, participants were divided into two groups. In the next step, this subdivision was used to analyze the groups' learning behavior, as the long-term goal is to establish a learning platform. To this end, cognitive aspects were examined through the mental rotation test and the anatomy knowledge test. This section presents descriptive analyses of the measured data, along with findings from the final participant interviews and expert interviews.

5.2.3.1 *Descriptive Analyses*

A descriptive analysis of the data was performed with respect to the background of the study participants. The summarized results are presented in Table 5 and visualized in Figure 19. Regarding the AKT, no major differences were found between the user groups. Participants with a medical background performed similarly with both the *active* (Average (Avg) = 0.68, Standard Deviation (SD) = 0.11) and *passive* (Avg = 0.74, SD = 0.05) features. Participants with technical backgrounds achieved noticeably lower scores on average compared to the other test group when using the *active* feature (Avg = 0.54, SD = 0.26). However, they showed similar AKT scores with the *passive* feature compared to the medical group (Avg = 0.74, SD = 0.05). Overall, the VLE achieved low *Subjective Workload* scores on N-TLX (Avg = 15.02, SD = 10.58). On average, the medical student group rated the VLE as more demanding (Avg = 17.26, SD = 11.99) compared to the technical group (Avg = 11.19, SD = 6.28). The *passive* feature was associated with a lower *Subjective Workload* (Avg = 12.89, SD = 9.50) and higher *Usability* (Avg = 92.74, SD = 5.59) compared to the *active* feature (Raw-N-TLX: Avg = 17.15, SD = 13.30; SUS: Avg = 90.00, SD = 9.19).

The SUS questionnaire data consistently showed high scores (Avg = 91.37, SD = 7.63) for the VLE across all user groups, features, and training phases. The technical background group reported higher *Usability* ratings (Avg = 95.14, SD = 3.90) compared to the medical group (Avg = 89.17, SD = 8.44).

Concerning the IPQ data, high scores were observed in General (Avg = 4.21, SD = 0.64) and Spatial Presence (Avg = 4.88, SD = 0.64), while lower scores were noted in Involvement (Avg = 2.97, SD = 1.06) and Experienced Realism (Avg = 2.89, SD = 0.74). Users with a technical background consistently reported higher scores compared to the medical group (Technical vs. Medical: General Avg = 4.71, SD = 0.47 vs. Avg = 3.91, SD = 1.14; Spatial Presence Avg = 5.00, SD = 0.50 vs. Avg = 4.82, SD = 0.71; Involvement Avg = 3.52, SD = 1.05 vs. Avg = 2.64, SD = 0.94; Realism Avg = 3.01, SD = 0.77 vs. Avg = 2.82, SD = 0.72).

5.2.3.2 *Post-Hoc Analyses*

Interest centered on whether mental rotation capabilities correlated with performance outcomes. Therefore, post-hoc Pearson correlation tests were conducted between the MRT results and all other questionnaire data. Results are presented in Table 6. No significant correlation effects were detected, except for the AKT, where poorer performance in the MRT was associated with lower AKT results. This effect is illustrated in Figure 20.

In addition, a secondary review of the raw data revealed differences between participants' first and second training runs. Separate descriptive analyses were conducted for each training session (see Figure 21). Raw-N-TLX ratings for the *active* and *passive* features showed similar results in the second training. The *active* feature appeared to have a higher *Subjective Workload* in the first training compared to the second. Regarding the SUS score, similar results were observed across both sessions for the *passive* feature, while the *active* feature achieved higher scores in the second training.

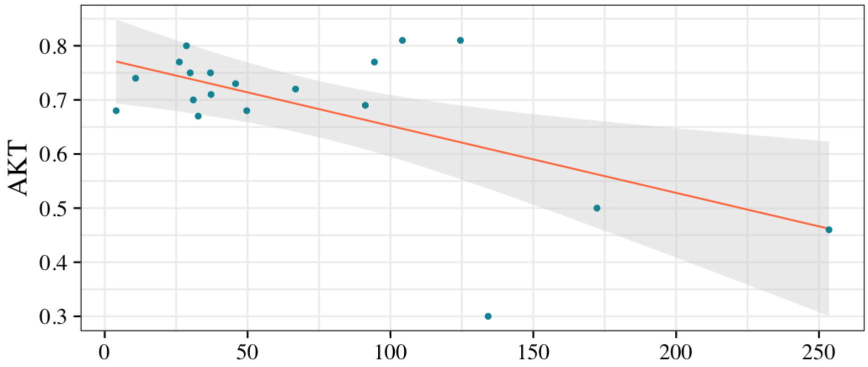


Figure 20: Correlation between MRT and AKT results. Adapted from Schott et al. (2023b) [Core6]. Licensed under CC BY 4.0.

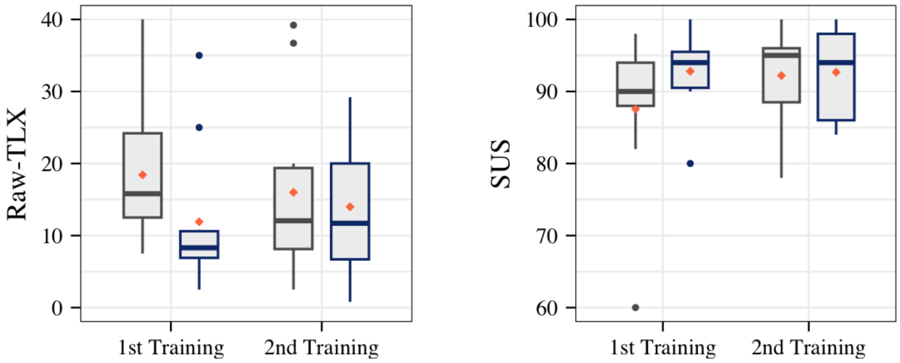
Table 6: Summary of MRT Pearson correlation test results ($\alpha < .05$). Reprinted from Schott et al. (2023b) [Core6]. Licensed under CC BY 4.0.

Training / Variable	df	p	r	Cor.
Complete Data				
AKT	17	0.006	-0.601	*
Raw-TLX	17	0.122	0.367	
SUS	17	0.179	-0.322	
IPQ-G	17	0.587	-0.133	
IPQ-SP	17	0.539	-0.150	
IPQ-INV	17	0.905	0.029	
IPQ-REAL	17	0.077	0.416	
First Training				
Raw-TLX	17	0.007	0.593	*
SUS	17	0.011	-0.569	*
Second Training				
Raw-TLX	17	0.733	0.084	
SUS	17	0.479	0.173	

Separate Pearson correlation tests were performed between the MRT and both questionnaires for the results of each training individually. These test results are presented in Table 6. No correlation was found for the second training; however, the Raw-N-TLX and SUS scores of the first training showed a significant correlation with the MRT results.

5.2.3.3 *Final Interview Results*

The audio recordings from the final interviews with medical students were transcribed into collective transcripts, resulting in a total of 195 individual statements. A total of 86 statements were summarized and assigned to five categories. Only statements that were consistent among at least three participants were included. This summary is presented in Table 7.



(a) (b)

Figure 21: Descriptive results of the individual first and second training on Raw-N-TLX and SUS scores with respect to the features tested in that training. ■ represents Active Feature, while ■ represents Passive Feature. ♦ indicate mean values. The scales of the Raw-N-TLX and SUS scores were truncated after 40 (max. 100) and before 60 (min. 0) respectively for better visibility of differences. Adapted from Schott et al. (2023b) [Core6]. Licensed under CC BY 4.0.

Table 7: Summary and number of statements (Quantity) sorted by category from the medical student survey. Adapted from Schott et al. (2023b) [Core6]. Licensed under CC BY 4.0.

Category	Statement	Quantity
3D Interactions	Trouble rotating the cube as intended	7
	Rotation around other axis desired (tilting)	5
	Connection of cube and model rotation is intuitive	4
	Walking around is easier than rotating via cube	3
	Two-handed rotation is easier than just one-handed	3
	Setting down the cube for larger rotations is awkward	3
	Rotation takes time to master and requires practice	3
	Desire to rotate the model directly by touching it	3
Active	Multiple attempts required to successfully trigger two-handed interaction	5
	Difficulty with hand tracking, coordination problems	3
	Trouble performing the two-handed grab and transform	3
	Better than passively watching the animation	3
Passive	Calm viewing experience improves perception of details	4
	Ability to precisely control the speed is rated positively	3
Free-form	Option to add more details is desired or expected	5
	Greater creative freedom is positive for learning success	4
	The level of detail of the modeling is more imprecise than guided interaction	3
General	Complexity and costs of MR are viewed as disadvantages	3
	Desire for an audio guide that leads through the content	3
	Annotations of anatomical structures should be displayed as an option	3
	More content needed for better chronological context	3
	3D representation is superior to classic illustrations from textbooks	3

5.2.3.4 *Feedback from Domain Experts*

Overall, the experts rated the VLE very positively. The advantages of 3D interaction were particularly emphasized, as the intuitive merging enhances the understanding of physiological deformation and is engaging for users. Experts noted that it resembles manual dissection on the adult heart, thereby promoting spatial awareness and contextual understanding. However, a limitation was noted in that the model cannot be viewed from all body axes, making it difficult to observe folding in the lower area and to view the interior. This issue could be addressed by adding a mode change button to toggle between deformation and rotation of the model.

Free modeling was regarded more as a creative feature for tracing, allowing students too many degrees of freedom, which may stimulate playfulness but is too time-consuming and does not accurately reflect physiological development. Movement constraints could be introduced by surrounding the heart tube with an adjacent pericardium.

The *active* feature was unanimously well-received, despite being initially more challenging to handle, as it stimulates motivation and may promote prolonged learning. Regarding the *passive* function, it was noted that reduced distraction allows for greater focus on details.

Given the emphasis on technical terms in medical education, it was suggested that these terms remain constantly visible and accurately assigned. Although the 2D display provides information on (developmental) day and stage as well as anatomical labels, medical terms should be directly displayed on respective structures within the dynamic development process. Additionally, the option to show and hide these annotations would allow learners to quiz themselves on the terminology.

The visualization of the heart was deemed sufficient in terms of color and shape, with no need for further realism. However, it was noted that the heart's context within the embryo is missing. It was recommended that the embryo and surrounding structures be made visible and adjustable through fade-in and fade-out options, with attention to color coding. The simplicity of the VE was also appreciated, as there are no distracting visual or auditory elements.

The custom AKT was rated as highly challenging; however, experts were uncertain whether it primarily tests logical thinking or actual knowledge. For example, the "Time" category was found relatively easy to solve. No differences were noted between abstract shapes and anatomical structures.

The application received broad approval for promoting 3D understanding and being well-prepared didactically. Although it shows only a small part of development and omits internal changes, it provides a clear insight into the complex development of the heart. For possible integration into the learning process and future curriculum, individualized training with a detailed auditory guide to support the user was recommended.

5.2.4 *Discussion*

The inclusion of a group with a technical background primarily served to assess the *Usability* of the VLE. This group rated the application slightly higher overall on the measurement tools, likely due to their familiarity with MR applications. However, it was found that users without technical expertise also rated the application as highly usable overall. The VLE received an overall very high *Usability* rating, with an average SUS score exceeding 90 [20]. It is speculated that the slightly lower ratings from physicians may stem from varying familiarity with MR applications, as this group generally has less experience with MR. This effect could also be attributed to the VLEs simplified interface and relatively limited scope. Nonetheless, the elements provided appear to meet user and expert requirements for the intended use case.

Due to the small sample size and the chosen study design, it was not possible to demonstrate a direct influence of the VLE on learning outcomes. However, recent studies indicate improved learning success in anatomy education through immersive MR as highlighted in Chapter 3.

Related work in embryology teaching [18, 45, 99, 111, 293] covers a broader range of developmental phases in static snapshots. This experiment distinguishes itself by visualizing the morphological transitions within each phase—though fewer in number—through interactive features for students.

In this study, the high IPQ score suggests that the VLE effectively induces a *Sense of Presence*. Particularly high scores in the sub-scales of "General Presence" and "Spatial Presence" indicate that users felt engaged within the VLE and could interact without distractions. The lower scores for "Immersion" and "Realism" are likely due to the use of abstract visualizations rather than realistic organ models.

No significant differences were observed between the two user groups in terms of *Usability* ratings; however, clear differences in *Subjective Workload* and IPQ scores emerged, suggesting possible cognitive overload. Participant feedback further indicated that certain learning types may benefit more from *passive* use of the VLE. For example, an animated 3D display that can be scaled and viewed without interaction may help reduce distractions.

The *passive* feature was associated with a lower *Subjective Workload*, likely because users performed fewer movements and interacted less overall, leading to reduced task demands.

Jang et al. [121] found that *passive* viewing, compared to interactive direct manipulation of 3D structures in a MR was less effective for learning in anatomy education. In this study's *active* variant, members of the technical group performed poorly on the AKT. This may be attributed to the small sample size, though it could also suggest a more effective learning experience in the *passive* VLE. It should also be noted that the AKT was not tailored to this group.

As this study serves as a pilot for broader testing and evaluation as part of our university's curriculum, identifying differences in learning styles is a priority. Lee et al. [149] demonstrated that VLEs can accommodate individual learning styles effectively.

Feedback from experts and medical students, along with collected data, suggests that the AKT may assess logical reasoning rather than learned anatomical knowledge. Consequently, the AKT could be considered a simplified version of an MRT, as the task design of both tests is quite similar, potentially explaining the correlation between their results. The MRT correlation tests between the first and second training sessions may indicate that prior experience positively impacts subsequent training.

As MR technology becomes increasingly common in medical education, the observed performance differences between users may diminish over time. Multi-user VLEs (as discussed in the next two chapters) are prominent in anatomy education literature, as they facilitate collaborative exchanges between students and instructors and support anatomical understanding through high levels of immersion [166]. This study adopted a controlled, single-user space to allow focused learning; however, the application is envisioned as adaptable for a seminar setting, where an instructor could either participate as a facilitator within the VLE or assume the role of active demonstrator. This transition is further investigated in Chapter 6.

5.2.5 *Limitations*

The implemented features, *active* and *passive*, differ only slightly in terms of actual user interaction. The primary distinction lies in the task of triggering a controllable animation. The ability to rotate the models or interface elements further complicates differentiation. The sample selected for evaluation demonstrated above-average mental rotation ability compared to the validation sample in the study by Ganis and Kievit [89], which may result in users with lower mental rotation abilities performing less effectively. The tested groups were not balanced, and the sample size was too small to address secondary questions. Due to this limited sample size, no significant findings were evident, and further statistical analyses (e.g., ANOVAs) were not conducted.

The option to freely model the heart tube initially appeared promising for enhancing interactivity and user motivation through a playful approach; however, it proved unsuitable, as it does not accurately reflect physiological development and thus contributes minimally to learning success. Providing pericardial constraints could potentially guide or restrict user movement, offering a compromise between free movement and realistic development. While the visualization meets the requirements, it does not enable an internal view of inner structures. Although smooth transitions are depicted within a phase, a fully seamless transition is not provided.

The developed AKT contains only images from the visualization itself and, therefore, is not representative of anatomical representations typically found in literature or real tests. As a result, an advanced knowledge test was developed, which will be explained in *Experiment 3*. Additionally, participants completed the same test twice in this experiment, with questions presented in a different order each time.

5.2.6 *Conclusion*

This experiment presented an MR-based system that focused on embryonic heart development. Designed with a HCD approach, the VLE aims to provide medical students with a novel means of understanding the dynamic morphological changes occurring in the early heart over several days. This objective was achieved by creating an immersive VLE featuring interactive, deformable 4D organ models. Various features were implemented to accommodate individual learning styles, offering insights into the unique challenges of such applications. Based on feedback from evaluations with students and professionals, further improvements for this VLE are envisioned.

Alternative learning scenarios using VST HMDs, projector-based setups, or handheld devices could also support collaborative learning and allow for direct integration into dissection courses. Despite enthusiasm for these technologies, future work should prioritize equity by addressing barriers such as device accessibility and individual student needs [124].

Collaborative learning approaches are further explored in Chapter 6, where physical presence and interaction align with the goals of anatomy education. A virtual replica of the examination room was deliberately used to provide users with a familiar yet adaptable environment, potentially enhancing the plausibility illusion and, consequently, the *Sense of Presence* and UX. In *Experiment 3*, a neutral VE was introduced, and in Chapter 6, students learn within an actual seminar room.

The importance of incorporating medical terminology directly onto anatomical models is emphasized in *Experiment 3*, where features such as annotations on specific structures play a crucial role in supporting terminology retention. Additionally, *Experiment 3* explores methods for displaying internal proportions and views of internal septation.

In summary, this experiment establishes a foundation for extending the VLE to other areas of embryonic development and highlights the transformative potential of MR in advancing medical education.

5.3 EXPERIMENT 3

This section presents the second experiment, which explores the use of the MR application *CardioGenesis4D*—enhanced based on prior findings and user feedback—to facilitate learning about embryonic heart development. Given the study's exploratory design, its initial focus on assessing application feasibility, and the limited sample size, no significant learning effects were observed. This highlights the need for a follow-up study to accurately evaluate the application's educational impact. This investigation focuses on the following sub research questions questions:

- Sub-RQ1: Can a VLE enhance understanding of complex anatomical transitions, and does it have a direct effect on students' knowledge acquisition?
- Sub-RQ2: Does interaction with MR applications lead to lasting knowledge retention, thereby supporting successful transfer to long-term memory and fostering a deeper understanding?
- Sub-RQ3: Are factors such as *Subjective Workload*, *Presence*, and UX influencing learning effectiveness, and could these serve as indicators in the development and evaluation of VLEs?

5.3.1 *Material*

The technical foundation of the MR system is based on the application *CardioGenesis4D*, as outlined in *Experiment 2*, maintaining the same basic software and hardware configurations. To create a comprehensive application aligned with curriculum-relevant content, the learning material was expanded. The initial phases, which covered only the early days of development, were extended to include key stages of embryonic heart formation from approximately day 18 to day 37. New 3D models were developed for this purpose, incorporating complex structures such as the atrial appendages and offering internal perspectives through sectional plane depictions. This approach also enabled visualizations of internal septation (see Figure 22).

ANNOTATIONS In the original application, students and experts expressed a desire for annotations that explain anatomical structures with corresponding terms. Oeltze-Jafra and Preim [193] identified five different types of labels in the medical field that can be applied depending on the context: internal labels (displayed directly on struc-

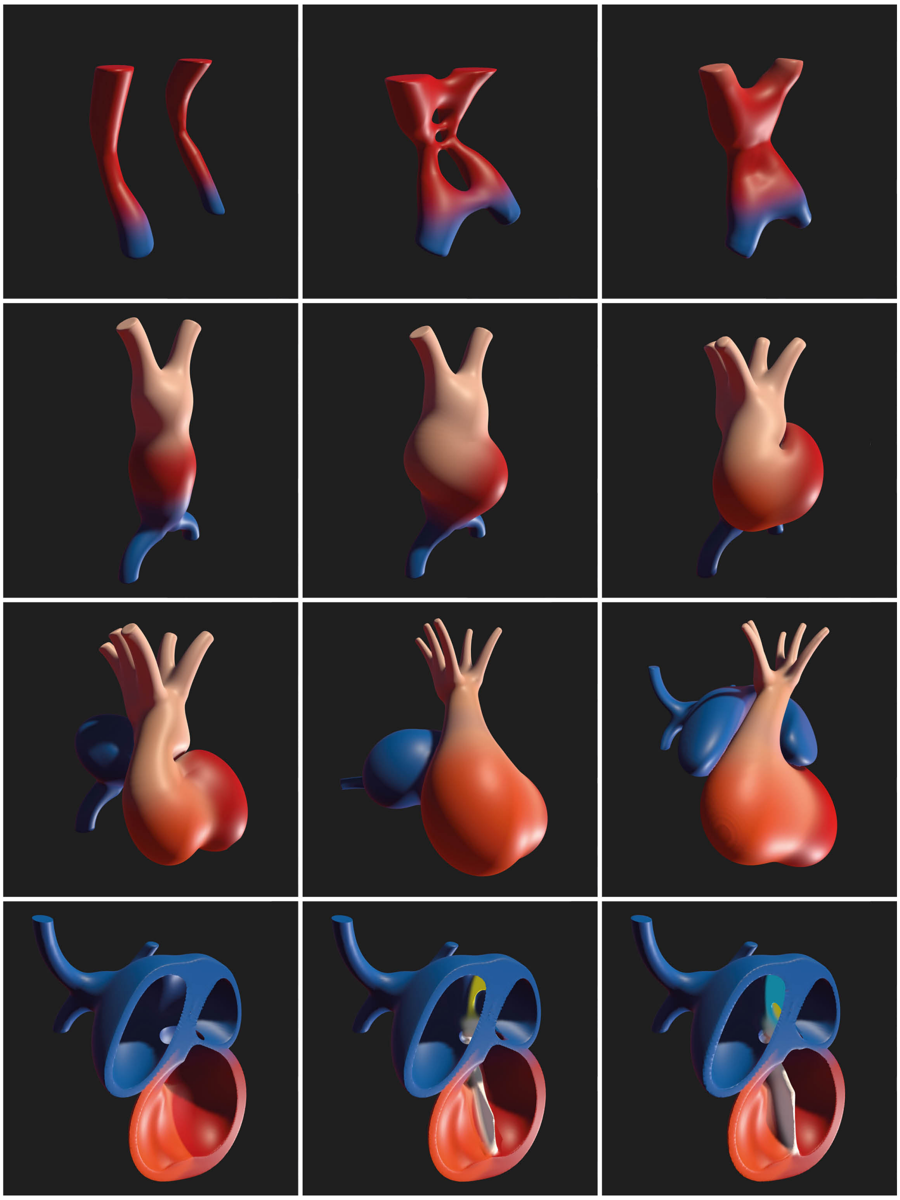


Figure 22: Overview of the total 3D model deformations. The enhanced representations include overlaps, outgrowths, and septations.

tures), external labels (positioned nearby with lines connecting them to structures), boundary labeling (arranged along the edges of a designated boundary area), excentric labeling (focused within a draggable region with labels outside the focus area), and necklace maps (labels placed around structures without lines, using color and proximity for association). In this application, external labels were chosen for their clarity and flexibility. Positioned near the structures with connecting lines, external labels allow detailed annotations without obstructing the view of the anatomical structures themselves. Therefore, dynamic labels were incorporated at appropriate points and aligned to the user's viewport, ensuring that the information remains accessible as the user explores different angles and perspectives.

AUDIO GUIDE To further enrich the learning experience, an audio guide was integrated, offering detailed insights into each developmental phase and guiding users through a step-by-step exploration of the application. This audio guide also clarifies controls and interactions, supporting independent, self-paced learning. Key details from the audio guide are summarized on an information board, conveniently positioned on the user's right-hand side.

UI OPTIMIZATION As identified in the previous experiment, providing learners with both *active* and *passive* interactions is essential. This insight informed the design of the UI, leading to refinements in interaction mechanics. The hand interaction system was enhanced to ensure smooth and intuitive control, including adjustments such as improved placement and scaling of colliders that register interactions between virtual objects and hand tracking, as well as general code optimizations. Advanced UI elements were incorporated, including a vertical timeline displaying the developmental stages to provide users with a clear overview of the process and their current position. Directly below the heart model in the center, animation controls are accessible, featuring a slider and play/pause button. Additional controls allow for audio management, rotation mode activation to prevent unintended manipulation of the model, toggling annotations, and resetting the model to its original position and shape. To guide user interactions, black-and-white crosses indicate starting points, while dashed arrows on the virtual heart display the direction of growth, illustrating the necessary movements to progress through each stage. These UI elements are fully movable, with a pinch gesture—triggered by a ray emitted from the user's hand—enabling convenient grabbing and repositioning of interface elements as needed.

NEW ENVIRONMENT Further adjustments to the VE have been made to place users in an open space that contextualizes the learning experience, creating a focused, immersive setting conducive to concentration and independent exploration. An illustration of the restructured application is provided in Figure 23.

5.3.2 *Evaluation*

The study aimed to evaluate the application's effectiveness in knowledge transfer, targeting medical students at their knowledge peak during the actual examination phase. Additionally, it examined the influence of factors such as *Subjective Workload*, UX, mental rotation ability, and demographic characteristics on learning outcomes. To as-

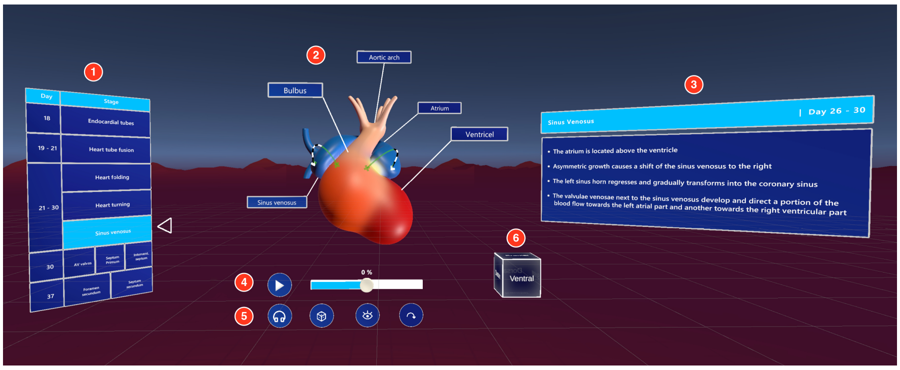


Figure 23: Enhanced application interface. (1) A timeline indicating the current developmental phase; (2) a 3D heart model with annotations and movement indicator arrows; (3) an information board summarizing the current phase; (4) a top toolbar featuring a slider and play-pause toggle for controlling the animation; (5) a bottom toolbar with controls for toggling the audio guide, rotation mode, labeling, and skipping transformations; (6) a 3D cube displaying anatomical orientation. Adapted from Kunz and Schott et al. (2025) [Core3], reused in accordance with Elsevier's author rights policy.

sess the educational effectiveness of the VLE, a knowledge test was administered, along with various questionnaires and the collection of subjective feedback.

5.3.2.1 *Ethical Approval*

This study received ethical approval (number 59/23) from our Medical Faculty's Ethics Committee. Ethical considerations, participant safety, and data protection measures were thoroughly addressed to ensure compliance with ethical standards and data safety regulations.

5.3.2.2 *Study Design*

A between-groups design was employed in this study, with participants randomly assigned to one of two groups. *Group A* completed an assessment of their embryology knowledge both before and after the MR session (pre-VLE), while *Group B* was assessed only after their MR session (post-VLE). The study was scheduled for the summer semester of 2023, just before the examination period, to ensure participants' knowledge was at its peak. To examine the retention of acquired knowledge, participants were also offered the opportunity to retake the knowledge test 14 days after their initial exposure. A detailed schematic of the study design is shown in Figure 24. The response rate was approximately 50%, with 38 participants in *Group A* and 34 participants in *Group B*.

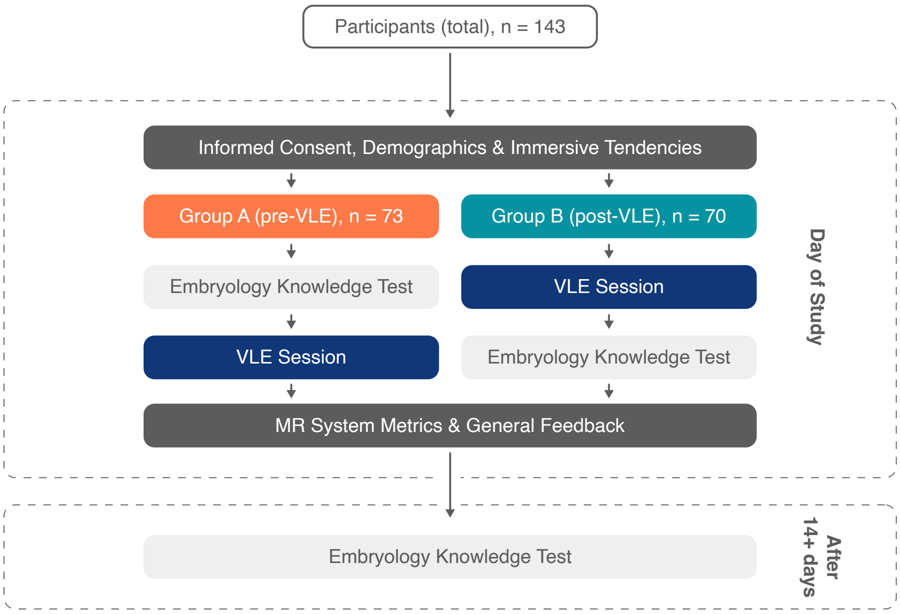


Figure 24: Study Design Flowchart: This diagram illustrates the randomized allocation of 1st and 2nd-year students into two groups, A (pre-VLE) and B (post-VLE), with a total of 142 participants. On the day of the study, all participants provided informed consent, were surveyed for demographics and immersive tendencies, and *Group A* took the EKT before the VLE session, while *Group B* experienced the VLE session prior to the knowledge test. The MR system´s metrics were assessed using the N-TLX, ITQ, and UEQ-s to evaluate the applicability of VE as a learning aid. Participants were retested on the embryology knowledge after 14+ days to measure long-term retention. Adapted from Kunz and Schott et al. (2025) [Core3], reused in accordance with Elsevier's author rights policy.

5.3.2.3 *Variables*

To thoroughly assess the effectiveness of the MR application as a learning tool, key variables with potential influence on learning outcomes were selected for examination. These variables were chosen to capture different aspects of UX and cognitive impact.

GENERAL As in the previous *Experiment 2*, spatial reasoning skills were to be measured using the MRT; however, due to a data collection error, these results were excluded from analysis. The N-TLX was used to gauge subjective mental load, with the unweighted Raw-N-TLX score providing an initial indication of *Subjective Workload* through a straightforward sum of ratings, as no specific weighting metrics were predefined [104]. Additionally, the UEQ-short was employed to assess UX, the IPQ to capture *Presence*-related qualities, and the ITQ to account for individual differences in predisposition toward immersion. A conscious decision was made not to measure *Usability* again using the SUS, as the basic application from *Experiment 2* had already received excellent *Usability* ratings.

EMBRYOLOGY KNOWLEDGE TEST A custom Embryology Knowledge Test (EKT) was developed in close collaboration with educators from the Institute of Anatomy at our university. This EKT differs from the AKT described in Section 5.2.2.1, as it not only evaluates the visual content presented in the application but is also closely aligned with actual anatomy examination formats. The EKT comprised 20 questions: 10 questions focused specifically on cardiac embryology content covered in the application (Heart EKT), while the remaining 10 assessed general embryology knowledge (General EKT). These general questions were included to ensure comparable baseline knowledge across groups and to identify any additional embryology learning that might have occurred outside the application. The test can be found in Section 9.5.

RATING AND QUALITATIVE FEEDBACK A 5-point rating scale was used to assess the application's effectiveness as a learning aid, ranging from 1 (strongly disagree) to 5 (strongly agree). Two items were included: "This application would facilitate my exam preparation" and "I endorse the inclusion of such an application in the curriculum." Regardless of group assignment or the timing of tests and questionnaires relative to MR session, each session concluded with a brief feedback discussion. This discussion focused on participants' experiences with the application, particularly its suitability as a learning tool, their interactions with it, and any positive or negative impressions. The study sessions averaged about one hour in duration.

5.3.2.4 *Participants*

A total of 143 medical students from the first (N=77) and second (N=66) years of study at our university participated in the study, each compensated with 20 EUR. Reflecting the gender distribution in the medical field, 72.7% of participants were female, closely aligning with the national average of 68.8% for female medical students in Germany[12]. Specifically, *Group A* consisted of 73 students, with a female representation of 77%, while *Group B* included 70 students, 69% of whom were female. The average age was approximately 22 years (*Group A*: Mean (M) = 21.9, Standard Error (SE) = .306; *Group B*: M = 22.3, SE = .361). Regarding visual impairments, most participants reported either no issues or corrected vision. A total of 11 participants across both groups reported general visual impairments or dyschromatopsia, with three cases present in *Group B*.

12 https://www.kbv.de/html/berufsmonitoring-medizinstudierende.php

Participants rated their technical affinity, MR experience, and gaming experience (see Chapter 9 for the demographic data collection sheet). Self-reported data indicated that both groups had moderately high technical affinity, with *Group A* averaging 3.03 (SE = .109) and *Group B* 3.29 (SE = .115). However, MR and gaming experience scores revealed limited familiarity with these technologies, with MR experience averages of 1.55 (SE = .105) in *Group A* and 1.63 (SE = .104) in *Group B*.

Additionally, participants completed a self-assessment on embryology knowledge, rating both their general understanding of embryology and their specific knowledge of heart embryology using Likert items from 1 (novice) to 5 (expert). They were also given the option to report their most recent examination grade on this subject, with scores ranging from 0 (failed) to 20 (excellent). A summary of the demographic data is provided in Table 8.

5.3.2.5 *Setup*

The study took place in a seminar room at our university's Institute of Anatomy. Each data collection session involved two participants and one investigator in the room. The application's automation and design, including an integrated audio guide offering step-by-step explanations, facilitated self-directed exploration. This guidance approach aimed to provide a smooth UX, reducing interruptions and supporting a focused research environment. A dedicated PC was provided for each participant, running both a web-based study questionnaire tool and the MR application. Additionally, separate tracking areas were arranged for each participant, ensuring a minimum free movement space of 3x3 meters, enabled by three Valve Base Stations (Valve Corp., USA).

5.3.2.6 *Procedure*

At the start of each session, participants received a comprehensive briefing on data protection protocols and a detailed overview of the study procedure, with informed consent obtained before beginning any activities. The 2nd-year students first completed a 10-minute MRT task on the PC (which, as noted earlier, could not be included in the analysis), followed by completing the required questionnaires in a browser. A red pop-up window then indicated when they were ready to proceed to the MR application.

Before the MR session, both *Group A* and *Group B* filled out the demographics questionnaire. *Group A* then completed the embryology test, while *Group B* completed it after the MR session. This timing prevented any interference between participants while using the MR application.

Table 8: Demographic Data Summary and student metrics. M = Median, SE = Standard Error; EKSA = Embryology Knowledge Self-Assessment; Last Grade = Last Embryology Exam Grade (Voluntary Disclosure). 1 = Glasses/Contact Lenses; 2 = General Vision impairment / Color vision deficiency; 3 = Assessment on 5-point Likert scale (1 = low/no experience/not knowledgable, 5 = high/highly experienced/expert); 4 = 20 point maximum; 5 = Immersive Tendencies Questionnaire Scoring. *** = Correlation is significant at 0.01 level (1-sided). Reprinted from Kunz and Schott et al. (2025) [Core3], reused in accordance with Elsevier's author rights policy.

Demographic Detail	***Type***	**Gr. A (pre-VLE)**	**Gr. B (post-VLE)**
Participants	*Number*	73	70
Gender	*Female*	56	48
	Male	17	22
Age	*M (SE)*	21.9 (0.31)	22.3 (0.36)
Year of Study	*1st*	38	39
	2nd	35	31
Visual Impairment	*None*	33	32
	Corrected[1]	32	35
	Misc[2]	8	3
Technical Affinity[3]	*M (SE)*	3.03 (0.11)	3.29 (0.12)
MR Experience[3]	*M (SE)*	1.55 (0.11)	1.63 (0.10)
Gaming Experience[3]	*M (SE)*	2.41 (0.15)	2.43 (0.17)
Last Grade[4]	*M (SE)*	18.27 (.267)	18.60 (.288)
EKSA[3]	*M (SE)*	2.74 (0.08)	2.69 (0.08)
Heart EKSA pre-VLE	*M (SE)*	2.59 (0.08)	2.59 (0.09)
Heart EKSA post-VLE	*M (SE)*	3.01 (0.11)***	3.26 (0.11)***
This application would facilitate my exam preparation	*M (SE)*	4.71 (0.06)	4.76 (0.07)
I endorse the inclusion of such an application in the curriculum	*M (SE)*	4.66 (0.08)	4.80 (0.06)
Time spent in VR	*M (SE)*	16.54 (0.83)	17.70 (0.92)
Involvement[5]	*M (SE)*	26.82 (0.82)	26.00 (0.89)
Focus[5]	*M (SE)*	31.68 (0.59)	31.37 (0.67)
Games[5]	*M (SE)*	4.18 (0.32)	4.63 (0.37)
Total[5]	*M (SE)*	71.26 (1.41)	70.51 (1.58)

Before entering the VLE, participants were briefed on safety aspects by their assigned study supervisors, and individual adjustments to the HMD were made. The MR experience started with an introductory tutorial, led by an investigator, that familiarized participants with essential UI elements and functionalities.

After familiarizing themselves with the VLE and its interactions, participants could independently initiate the application by pressing a button when ready. An audio guide then provided a step-by-step explanation of key app elements, including the timeline, text information, and interactions with the 3D model. Participants could explore each phase of development at their own pace, guided by the audio instructions, while an information panel summarized each developmental step. Their main task was to replicate the physiological movement of each stage using hand interactions. Upon successful completion of each step, a green checkmark confirmed the task, and participants had the option to play an automatic animation, interact with the model further, or reset the task using a dedicated button. Once they felt confident, they could progress to the next phase by pressing the "Next" button.

After reaching the final development phase, participants were informed of completion and had the option to repeat any steps. The duration of the MR experience was self-paced, averaging about 17 minutes. The session concluded with a brief opportunity for participants to provide general feedback on the VLE, wrapping up the study in approximately one hour.

5.3.2.7 *Data Analyses*

The required sample size was calculated using G*Power (version 3.1.9.7) based on the effect size reported in a recent meta-analysis [57], indicating a minimum of 64 participants. However, in accordance with the ethical approval, all students who expressed interest in participating were included, resulting in a final sample of 143 students. All statistical analyses were performed using SPSS (version 23) [13]. The EKT results were initially examined for homogeneity of variances using Levene's test and for normality using the Shapiro-Wilk test. The Shapiro-Wilk test indicated a significant result, suggesting that the data did not follow a normal distribution. Therefore, the Mann–Whitney U test was applied to assess significant differences in EKT scores between groups, and the Wilcoxon signed-rank test was used for within-group comparisons on paired samples.

5.3.3 *Results*

This section presents the results of our study examining the effectiveness of a VLE in enhancing embryological knowledge. Our primary focus was on evaluating both the immediate and long-term learning effects of a single VLE session. The table of results has been streamlined to highlight only variables with high correlations (see Table 9).

5.3.3.1 *Knowledge Tests*

Comparing Figure 25, a significant difference is observed in the *Heart Embryology* scores in *Group A* versus *Group B* (post-VLE), while no difference is noted in *General Embryology* knowledge. Results show that the learning effect in *Group A* (pre-VLE) remains after two weeks. A

13 IBM Corporation, USA

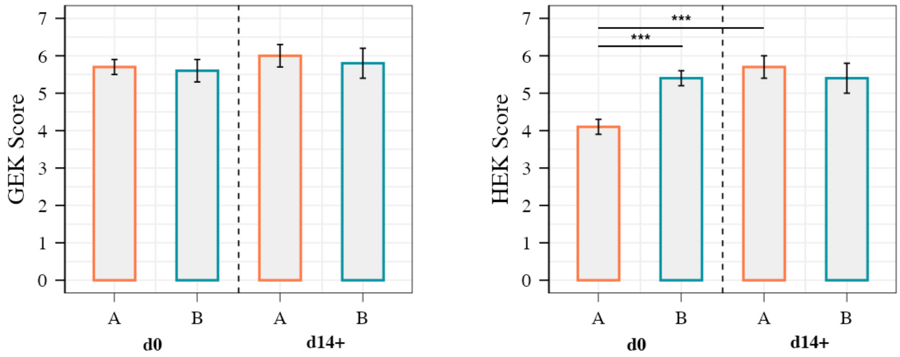
Table 9: ** = Correlation is significant at 0.01 level (1-sided); * = Correlation is significant at 0.05 level (1-sided); IPQ sub-scores: G = General Presence, INV = Involvement, SP = Spatial Presence, REAL = Realism, UX = User Experience Questionnaire Score. TLX = Raw NASA Task Load Index. ITQ = Immersive Tendencies Questionnaire Score. LE = Learning Effect; PCC = Pearson Correlation Coefficient; Sig. = Significance; N = Number of Subjects. The intensity of colors represents the strength of the Pearson correlations: Darker shades of ■ indicate stronger positive correlations, while darker shades of ■ indicate stronger negative correlations. The gradation of colors corresponds to the respective Pearson r-value. Reprinted from Kunz and Schott et al. (2025) [Core3], reused in accordance with Elsevier's author rights policy.

		G	INV	SP	REAL	UX	TLX	ITQ	LE
G	*PCC*	1	.350**	.558**	.283**	.180	.077	.210*	.310*
	Sig.		.001	.000	.008	.064	.258	.037	.029
	N	73	73	73	73	73	73	73	38
INV	*PCC*	.350**	1	.359**	.238*	.083	-.048	.302**	.233
	Sig.	.001		.001	.021	.244	.342	.005	.080
	N	73	73	73	73	73	73	73	38
SP	*PCC*	.558**	.359**	1	.278**	.370**	-.207*	.302**	.328*
	Sig.	.000	.001		.009	.001	.040	.005	.022
	N	73	73	73	73	73	73	73	38
REAL	*PCC*	.283**	.238*	.278**	1	.134	-.180	.126	.305*
	Sig.	.008	.021	.009		.129	.064	.144	.031
	N	73	73	73	73	73	73	73	38
UX	*PCC*	.180	.083	.370**	.134	1	-.266*	.136	-.100
	Sig.	.064	.244	.001	.129		.011	.126	.276
	N	73	73	73	73	73	73	73	38
TLX	*PCC*	.077	-.048	-.207*	-.180	-.266*	1	-.032	.237
	Sig.	.258	.342	.040	.064	.011		.394	.076
	N	73	73	73	73	73	73	73	38
ITQ	*PCC*	.210*	.302**	.302**	.126	.136	-.032	1	-.002
	Sig.	.037	.005	.005	.144	.126	.394		.495
	N	73	73	73	73	73	73	73	38
LE	*PCC*	.310*	.233	.328*	.305*	-.100	.237	-.002	1
	Sig.	.029	.080	.022	.031	.276	.076	.495	
	N	38	38	38	38	38	38	38	38

slight decline is noted in the *Heart Embryology* scores for *Group B* (post-VLE) compared to *Group A*, though this is not statistically significant.

5.3.3.2 *General metrics*

The revised application used in *Experiment 3* achieved high scores across both groups in terms of UX (UEQ), *Presence* (IPQ), and low perceived Task Load (N-TLX) (see Figure 26 and Figure 27). The study results indicate no statistically significant differences between *Group A* (pre-VLE) and *Group B* (post-VLE) across the various scales and subscales. For the IPQ results:



(a) (b)

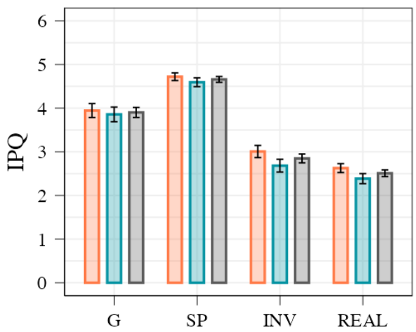
Figure 25: Comparison of EKT points scored (maximum 10) before and after VLE exposure. Figure (a) presents the evaluation of *General* EKT results for *Group A* (pre-VLE), represented by ■, and *Group B* (post-VLE), represented by ■. The left bars show the results on the day of the study (d0), and the right bars show the results after 14 days (d14+). Figure (b) follows the same scheme, displaying the results for the *Heart* EKT. Bars represent means, and error bars represent standard errors. Asterisks indicate significant differences: *** denotes $p < 0.001$. Adapted from Kunz and Schott et al. (2025) [Core3], reused in accordance with Elsevier's author rights policy.

- **General Presence**: M *Group A* = 3.95 (SE = 0.16), M *Group B* = 3.86 (SE = 0.17), Overall M = 3.90 (SE = 0.12)
- **Spatial Presence**: M *Group A* = 4.72 (SE = 0.09), M *Group B* = 4.59 (SE = 0.10), Overall M = 4.66 (SE = 0.07)
- **Involvement**: M *Group A* = 3.01 (SE = 0.14), M *Group B* = 2.68 (SE = 0.15), Overall M = 2.85 (SE = 0.10)
- **Realism**: M *Group A* = 2.63 (SE = 0.10), M *Group B* = 2.39 (SE = 0.12), Overall M = 2.51 (SE = 0.08)

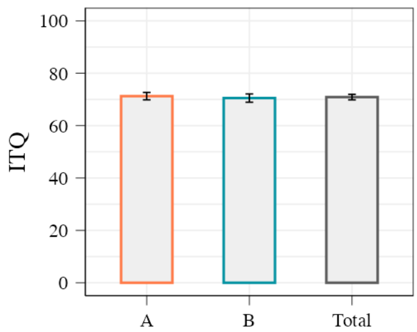
Similarly, the total ITQ score showed little difference between the groups, with *Group A* averaging 71.26 (SE = 1.41) and *Group B* averaging 70.51 (SE = 1.58), yielding a combined mean of 70.90 (SE = 1.05). The UEQ also displayed similar patterns: *Group A* scored a total of 2.42 (SE = 0.05), while *Group B* scored 2.48 (SE = 0.07), with an overall mean of 2.45 (SE = 0.04). The N-TLX results mirrored these patterns, with both groups showing comparable mean scores across the subscales:

- **Mental Demand**: M *Group A* = 9.38 (SE = 0.51), M *Group B* = 8.84 (SE = 0.54), Overall M = 9.12 (SE = 0.37)
- **Physical Demand**: M *Group A* = 3.41 (SE = 0.37), M *Group B* = 3.77 (SE = 0.39), Overall M = 3.59 (SE = 0.27)

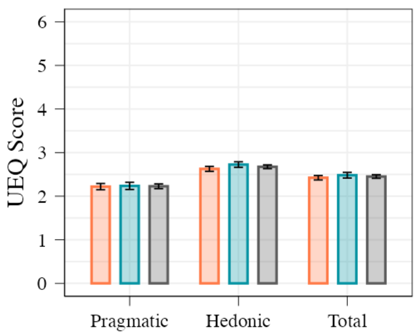
- **Temporal Demand**: M *Group A* = 5.10 (SE = 0.47), M *Group B* = 4.70 (SE = 0.49), Overall M = 4.90 (SE = 0.34)
- **Performance**: M *Group A* = 5.82 (SE = 0.48), M *Group B* = 4.84 (SE = 0.45), Overall M = 5.34 (SE = 0.33)
- **Effort**: M *Group A* = 4.48 (SE = 0.36), M *Group B* = 4.46 (SE = 0.39), Overall M = 4.47 (SE = 0.26)
- **Frustration**: M *Group A* = 3.56 (SE = 0.38), M *Group B* = 3.01 (SE = 0.37), Overall M = 3.29 (SE = 0.26)



(a)



(b)



(c)

Figure 26: Results of the questionnaires measuring *Sense of Presence* (a), immersive tendencies (b), and UX (c). ■ represents *Group A* (pre-VLE), ■ represents *Group B* (post-VLE), and ■ represents the total scores. Bars represent means, and error bars represent standard errors. Adapted from Kunz and Schott et al. (2025) [Core3], reused in accordance with Elsevier's author rights policy.

5.3.3.3 *Expert Feedback*

All students were given the opportunity to provide open feedback on their experience with the application. Participants shared their insights freely, and their responses were documented, summarized, and systematically categorized through qualitative content analysis. This

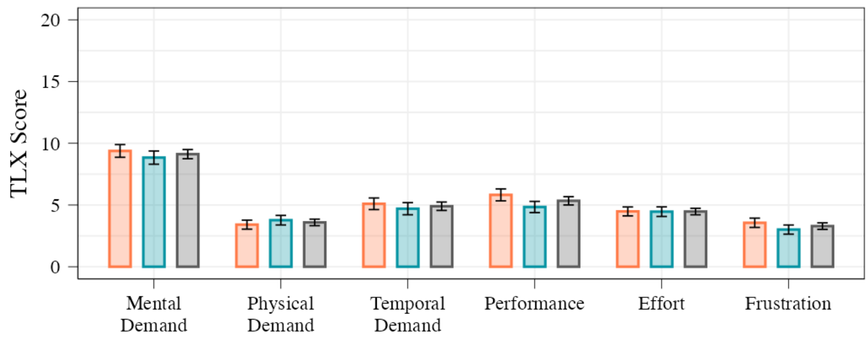


Figure 27: Results of the individual raw N-TLX items. ■ represents *Group A* (pre-VLE), ■ represents *Group B* (post-VLE), and ■ represents the total scores. Bars represent means, and error bars represent standard errors. Adapted from Kunz and Schott et al. (2025) [Core3], reused in accordance with Elsevier's author rights policy.

process helped to identify key themes and patterns in the feedback, providing a clearer understanding of users' perspectives and experiences. Table 10 presents a summary of this feedback.

Additionally, the application was demonstrated to three anatomy teachers (ATs), who explored it independently and provided feedback. All three experts expressed satisfaction, noting that the content was comprehensive, the detail level was adequate, and the application was well-suited for exam preparation. They highlighted its value as an additional learning tool, particularly due to the interactive features that facilitate understanding of embryonic structures and enhance theoretical knowledge.

However, some constructive feedback was provided: AT 1 recommended smoother transitions within the application and suggested including more foundational concepts, such as earlier developmental stages. AT 2 proposed refining the sequence of dissecting the heart. Despite these suggestions, the overall feedback was positive, with AT 3 praising the application's effectiveness and expressing interest in expanding it to cover additional developmental contexts and clinical extensions.

5.3.4 *Discussion*

This discussion evaluates the MR system´s effectiveness in supporting knowledge acquisition, long-term retention, and the influence of various factors on learning outcomes. Each sub-research question is addressed, examining the educational value of the VLE and the roles of *Presence*, *Subjective Workload*, and UX. Additionally, limitations and further insights are reviewed to guide future research and improvements in MR-based anatomy education.

5.3.4.1 *Sub-RQ1: Enhancing Knowledge Acquisition*

To address the first research question, this study assessed whether a VLE could enhance understanding of complex anatomical transitions in embryology. The study incorporated the EKT, with a *General Embryology* section to establish baseline equivalency and a *Heart Embryology* section to measure learning specific to the VLE session. Analysis of Figure 25 (do) showed a significant improvement in *Heart Embryology* scores before (*Group A*) versus after (*Group B*) the VLE session, with no notable difference in *General Embryology* scores. This suggests that even a single VLE session has a direct, positive effect on students' knowledge acquisition for complex topics like embryonic heart development.

5.3.4.2 *Sub-RQ2: Long-Term Knowledge Retention*

To explore long-term retention, participants' knowledge was reassessed after 14 days (see Figure 25, d14+). Findings indicate that the learning effect in *Group A* (pre-VLE) remained stable after two weeks, underscoring the VLE's potential for supporting long-term memory retention. Although a minor decrease in *Heart Embryology* scores was noted for *Group B* (post-VLE), this difference was not statistically significant, reinforcing the efficacy of a single VLE session in facilitating lasting knowledge acquisition. In contrast, general knowledge remained consistent across time and groups, suggesting no additional embryology learning occurred outside the VLE session.

As noted by [336], VLEs have demonstrated immediate educational benefits, though many studies lack the long-term perspective this research aimed to provide. This aligns with other MR studies suggesting that immersive sessions can significantly enhance retention [335]. While the minor decline in *Heart Embryology* scores for *Group B* (post-VLE) was not significant, future research should explore the effects of multiple sessions on retention, as suggested by [162].

5.3.4.3 *Sub-RQ3: Influence of Workload, Presence, and UX on Learning Effectiveness*

The third research question examined whether *Subjective Workload*, *Presence*, UX, and user traits affect learning outcomes and could guide VLE development. Correlation analysis indicated that *Presence*—particularly as measured by IPQ subscales—had a strong positive association with learning effect (see Table 9).

This correlation between *Presence* and learning aligns with findings by Weber et al. [316], suggesting that *Presence* is a significant factor for engagement. However, unlike other studies emphasizing UX in virtual learning [258], UX in this study was only secondarily associated, possibly due to the VLE's high baseline *Usability*.

The abstract nature of the updated VLE raises questions about whether a more realistic setting—such as a seminar room or lecture hall—could further enhance learning or if *Presence* alone effectively supports learning outcomes. Using spaces that closely resemble real-world learning environments could improve plausibility, potentially enhancing the user's immersive experience.

As previously noted, *Usability* was not explicitly measured, given that the original version had received excellent ratings. With further optimizations in this version, it was assumed that *Usability* would not significantly differ or worsen. In this evaluation, UX aspects and immersive tendencies had a secondary influence on *Presence*, without a direct effect on learning. Conversely, UX correlated with *Spatial Presence* and immersive tendencies, suggesting it may support integration into the use case. Workload exhibited an inverse relationship with UX and *Presence*, suggesting that high engagement-related *Subjective Workload* could benefit learning, even if it detracts from *Usability*. These findings highlight the importance of *Presence* in learning within VLEs, while indicating that UX, *Subjective Workload*, and personal traits play supporting roles rather than directly impacting learning outcomes.

Table 10: Summary of qualitative feedback. Reprinted from Kunz and Schott et al. (2025) [Core3], reused in accordance with Elsevier's author rights policy.

Feedback Category	**Summary**
Interactivity and Spatial Imagination	Participants often praised the interactivity and the ability to view the heart in 3D. Challenges in grasping and rotating were mentioned, with the potential cause being the use of hand tracking.
Audio-Text Combination	The combination of audio explanations and text was deemed helpful in catering to various learning styles. It was noted that occasionally, text and audio were not consistently uniform across the application.
Clarity and Learning Benefits	The application assisted participants in better understanding and visualizing complex embryonic developmental processes. Some participants found the application more helpful than traditional teaching materials such as books or videos.
Comfort and MR Experience	Some participants experienced issues with the MR headset, including discomfort and headaches. The MR experience was viewed positively and interestingly by many, especially for spatial visualization.
Suggestions for Improvement	Suggestions were made to enhance handling, highlight text during audio explanations, and integrate notes or annotations into the MR environment. Some participants proposed incorporating the application into the curriculum.
Additional Remarks	The 2nd year students indicated that they would have benefited from the application before exams or assessments. It was also noted that the application is particularly suitable for spatial understanding, making challenging processes more tangible.

5.3.4.4 *Limitations*

A limitation of this study is the use of the N-TLX for cognitive load measurement. Although widely used in computer science, the N-TLX primarily assesses stress and mental burden rather than specific types of cognitive load associated with learning. For educational applications like this, a cognitive load questionnaire focused on learning processes would be more appropriate. Future studies could consider using specific cognitive load questionnaires, such as those by Sweller et al. [292] or Leppink et al. [153]. Notably, the questionnaire by Klepsch et al. [134] differentiates between types of cognitive load and is designed for learning contexts, allowing for more precise cognitive load assessment in VLEs.

5.3.5 *Conclusion*

This study underscores the potential of VLEs to enhance knowledge acquisition and retention for complex anatomical content, such as embryonic heart development. Findings reveal that a single VLE session can improve understanding of embryological transitions and support long-term retention, with *Presence* emerging as a critical factor. While UX, *Subjective Workload*, and immersive tendencies play secondary roles, further research could explore these factors' contributions in more depth. Overall, a multi-dimensional evaluation approach is recommended for VLE studies, focusing on learning outcomes, UX, *Presence*, and immersion, rather than solely on individual characteristics.

Supporting prior findings on user-centered VLEs, these results underscore the importance of designing VLEs that are adaptable to different learner profiles, as suggested by De Freitas [63]. Such adaptability is crucial to enhance learner engagement and ensure that immersive experiences contribute effectively to educational outcomes. This comprehensive framework aligns with calls for rigorous evaluation in MR-enhanced education, particularly in medical fields [7, 335]. By adopting this approach, VLEs can be further refined to improve learning experiences and outcomes across diverse educational contexts.

5.4 CHAPTER SUMMARY

This chapter presented the development and evaluation of an MR application designed to support students in exploring theoretical knowledge on early human heart development through a problem-based learning approach. The application is intended for use as a self-study tool following initial anatomy lectures.

The VLE provides a controlled setting that allows users to concentrate on essential learning content without distractions, encouraging students to individually construct knowledge and making it an ILE.

The learning experience emphasizes embodiment and natural hand interactions, delivering a hands-on approach intended to enhance spatial understanding and facilitate comprehension of physiological development through movement-based exploration of 4D heart models.

The original system, *CardioGenesis4D*, introduced in *Experiment 2*, provided the technical and methodological foundation. This system was developed to accommodate different learning preferences by supporting both active and passive learning style through various features, offering a guided exploration with a didactic design. Evaluation results showed that the VLE achieved high *Usability* ratings and successfully induced a *Sense of Presence*.

As the study design and sample size in *Experiment 2* were not sufficient to demonstrate the impact on learning outcomes, *Experiment 3* was conducted as a follow-up investigation. This study, based on an enhanced version of *CardioGenesis4D*, specifically examined learning effects and the factors influencing them.

The findings from *Experiment 3* confirm that the enhanced *CardioGenesis4D* VLE can effectively improve students' understanding of complex anatomical transitions, showing a direct, positive impact on knowledge acquisition specifically related to heart embryology. Additionally, the VLE supports lasting knowledge retention, with effects on heart embryology knowledge remaining stable over two weeks, indicating its value in promoting long-term memory transfer.

Further, the study revealed that factors such as *Presence*, UX, and *Subjective Workload* influence learning effectiveness in varying ways. Presence was positively associated with the learning effect, suggesting it may enhance engagement and focus within VLEs. UX and *Subjective Workload* play supportive roles, with UX correlating with spatial presence and *Subjective Workload* showing an inverse relationship with both UX and *Presence*, providing valuable indicators for the continued development and refinement of VLEs for educational purposes.

The design of both applications followed a HCD process, involving various stakeholders throughout development and evaluation. This approach aimed to address the needs of two key user groups: students, who require intuitive interaction and effective content comprehension, and educators, who must ensure content accuracy and consider opportunities to integrate MR systems into their teaching.

Given the project's successful implementation, discussions were held in collaboration with educators from our university to explore additional didactic and technical possibilities for this VLE. The following chapter will delve into this topic, evaluating different hardware options for a collaborative learning scenario. It will conclude with a comparison of the various teaching and learning approaches used in these implementations.

COLLABORATIVE LEARNING ENVIRONMENTS FOR EMBRYONIC HEART EDUCATION

ch. 6

SYNOPSIS *Building on the foundation of the previous chapter, the following chapter presents the iterative development of a collaborative MR-based learning system, shaped by the requirements of both experts and students. Through three interlinked experiments, the potential integration of a VLE into concrete anatomical training scenarios is explored. This culminates in a comparison of different didactic approaches within a shared VLE, aimed at examining their effectiveness in enhancing teaching and learning outcomes.*

ABOUT THIS CHAPTER *Portions of this chapter were previously published in Schott et al., "AR-Based Multi-User Learning Environment for Anatomy Seminars" [Core4], "CardioCoLab: Collaborative Learning of Embryonic Heart Anatomy in Mixed Reality" [Core2], and "Stand Alone or Stay Together: An In-situ Experiment of Mixed-Reality Applications in Embryonic Anatomy Education" [Core1] and have been incorporated into this thesis. The article published via IEEE [Core4] is reused here in accordance with IEEE's policy for thesis use. © 2024 IEEE. Reprinted with permission. The article published by the Eurographics Association [Core2] is open access under a Creative Commons Attribution 4.0 International License (CC BY 4.0, https://creativecommons.org/licenses/by/4.0/). The article published via ACM [Core1] is reused in accordance with ACM's policy for open access research articles. Additionally, some of the methods described in Experiments 4 and 5 of this chapter were developed as part of Mr. Jonas Mandel's Bachelor's thesis, titled Exploration medizinischer Visualisierung in Multiuser Augmented Reality, which I supervised.*

MY CONTRIBUTION *I developed the main research idea and was responsible for the conceptualization and methodological approach, building upon and refining the research concept initially formulated in Mr. Mandel's Bachelor's project for Experiments 4 and 5. This included expanding the theoretical foundations and conducting a comprehensive literature review. Additionally, I managed the research project, overseeing the overall methodology, which encompassed requirement analysis through expert interviews, interface design, and visualization. Data collection was performed collaboratively. For Experiment 6, I further advanced the research idea, supervised the technical implementation, and was responsible for the study design. Data collection and analysis were conducted collaboratively. Across all experiments, I took primary responsibility for drafting, reviewing, and editing the manuscripts of the original papers. Furthermore, I supervised and actively contributed to the creation of visuals, producing new photos, tables, illustrations, and additional content specifically for this thesis.*

6.1 INTRODUCTION

The use of immersive MR offers a significant advantage in educational contexts by allowing learners to isolate themselves from external distractions and focus entirely on the learning content. In the previous chapter (Chapter 5), such an application was introduced and evaluated, revealing evidence that VLEs can positively impact user engagement and learning outcomes, particularly in the study of embryonic heart development. This application has the potential to become a valuable addition to the medical curriculum, especially as part of foundational anatomy education. It could serve as a supportive medium that students use alongside regular classes, enabling them to independently revisit classroom material and deepen their understanding of complex morphological changes through MR's interactive capabilities.

CHALLENGES Following the successful evaluation of *CardioGenesis4D* (Chapter 5), which involved a large proportion of medical students in their learning phase and received positive feedback from both students and educators, questions arose about the broader potential of this concept. Specifically, experts expressed interest in how such an interactive MR system could be integrated into regular classroom instruction. This prompted the exploration of new didactic and technical formats to realize such integration.

One of the primary challenges identified was the integration of MR technology into traditional teaching workflows in a resource-efficient manner. Educators emphasized the need for systems that enable multiple users to participate in a shared VLE. At our university, heart embryology is traditionally taught as part of an embryology seminar conducted in small groups of 10 students during the dissection course. These 90-minute seminars rely on PowerPoint-supported frontal teaching to convey cardiac embryology concepts. The goal was to enhance this traditional seminar format by enabling multiple users to interactively and synchronously explore the developmental stages of the embryonic heart. Consequently, it became essential to identify alternative MR systems capable of supporting group learning while including an instructor for effective guidance.

This chapter addresses these technical and pedagogical challenges. It presents an approach that aligns more closely with traditional seminar formats and instructional theory (see Section 2.2.1). At the same time, it seeks to ensure that dynamic communication and knowledge exchange within the group are fostered, leading to the creation of a collaborative learning environment. Additionally, the feasibility of adapting the visualizations and interactions from *CardioGenesis4D* to meet these new requirements was explored, considering various input and output devices.

RESEARCH QUESTIONS Building on the foundation of *CardioGenesis4D*, this chapter aims to expand learning opportunities by emphasizing collaboration and establishing a framework for knowledge sharing within dynamic educational settings. By integrating new technological approaches, the chapter bridges diverse didactic formats and fosters a more interactive learning experience. The research question for *Experiment 4* is:

> **RQ4** | What are the technical and pedagogical requirements for a collaborative MR-based system to effectively support the learning of embryonic heart development?

Once the technical framework is established, the focus will shift to asses the pedagogical suitability of the concept. This led to the following question:

> **RQ5** | How can a collaborative MR-based learning environment for understanding embryonic heart development be effectively integrated into an anatomy seminar setting?

Finally, addressing a gap in the literature, this study explores the direct comparison of different didactic approaches within MR-based learning environments for anatomy education. To this end, two distinct VLEs were developed based on the application: an ILE and a CLE. This gives rise to the third research question:

> **RQ6** | How do individual and collaborative MR-based learning environments differ in supporting educational outcomes for embryonic heart development?

ITERATIVE DEVELOPMENT This chapter highlights the iterative design process guided by the HCD framework and involves three consecutive experiments aimed at identifying suitable didactic formats for teaching and learning about embryonic heart development. In *Experiment 4*, new hardware was introduced to enable group interactions and collaborative viewing. Requirements for interactions and visualization options were evaluated with educators through a participatory workshop. Based on these findings, the system was refined and evaluated in a simulated seminar setting in *Experiment 5*. Insights from these experiments informed the transition to new hardware in *Experiment 6*, which culminated in a comparison of different learning formats—individual versus collaborative (seminar-oriented)—to assess their relative effectiveness.

The Experiments aim to enhance learning opportunities by shifting the focus towards fostering collaboration and establishing a foundation for knowledge sharing in a more dynamic educational setting using MR technology. This progression represents a natural extension by integrating the social aspects of learning and bridging the gap between diverse didactic formats.

A video demonstration of the applications is linked in the Appendix (see Section 9.1).

6.2 EXPERIMENT 4

This section presents the first of three experiments in this chapter. It describes the development of a collaborative MR system based on the application introduced in Chapter 5. The focus lies on evaluating the system's configuration, including visualization and interaction capabilities, while also considering potential integration into anatomy seminar settings.

6.2.1 *Material*

As previously mentioned, the conceptual and technical foundation of this system builds upon the enhanced *CardioGenesis4D* application (Section 5.3). To meet the previously outlined requirements, a new hardware solution was evaluated. For this purpose, the Tilt Five tabletop system[1] was selected.

The Tilt Five system features a HMD with dual HD projectors and a gameboard covered in retroreflective material. This material reflects light from the projectors directly back to the HMD, creating a stereoscopic 3D effect that gives the impression of a holographic display. The gameboard measures 800mm x 1066.7mm and is equipped with tracking markers around its edges. These markers enable the HMD to accurately determine its position relative to the board, allowing each player to experience a unique perspective on the displayed content.

The system requires a PC or Android device for operation and includes its own Software Development Kit (SDK). The HMD is also equipped with an infrared camera for head tracking. Input is supported via controllers, referred to as "wands," which can be used by multiple users (HMDs) simultaneously.

While marketed as an AR system, it could also be classified as Fish Tank VR, as defined by Ware et al. [314]. This classification reflects its unique combination of immersive visual presentation with a tabletop, semi-enclosed display environment.

6.2.1.1 *Hardware Choice Rationale*

The specific decisions for this system can be justified as follows:

- *Cost-efficiency and Accessibility:* The Tilt Five gameboard offers a more cost-effective and portable solution compared to a fully immersive MR setup, as used in the previous experiment.
- *Collaborative Potential:* Unlike fully immersive MR systems that isolate the user, the Tilt Five gameboard enables cooperative learning experiences. Multiple users, including students and in-

1 Tilt Five Inc., https://www.tiltfive.com

structors, can interact within the same physical and virtual environment, fostering communication and collaborative learning.

- *Reduced Cognitive Load:* The semi-immersive nature of the Tilt Five system can reduce the cognitive load associated with MR learning by minimizing sensory overload.
- *Adaptability to Traditional Classroom Settings:* The tabletop form factor of the Tilt Five system is better suited to traditional seminar and classroom setups. This facilitates integration into existing teaching formats and ensures a smoother transition from conventional to MR-based teaching methods.
- *Scalability and Shared Perspectives:* The Tilt Five system supports shared perspectives and simultaneous interaction, making it ideal for groups where multiple users can view and manipulate the same 3D models.
- *Technical Evaluation:* Testing the Tilt Five system provides an opportunity to evaluate its performance, *Usability*, and suitability for anatomy education compared to a fully immersive MR setup. Insights gained from this evaluation could inform future decisions about hardware selection for educational applications.

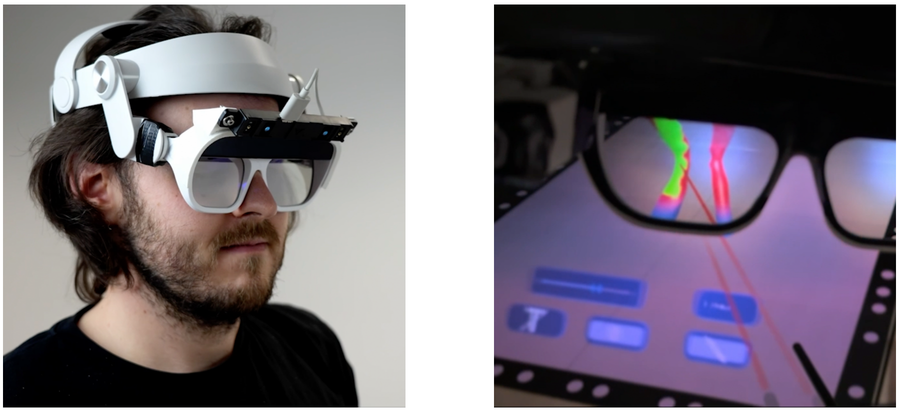
6.2.1.2 *Technical Setup*

The system was initially designed to accommodate four users, as supporting a larger number would require an additional PC, and the setup was intended to remain minimal during the initial implementation phase. The same development phases and corresponding 3D models from the enhanced *CardioGenesis4D* application (Section 5.3) were utilized. Using the "wand" controllers, users could interact with the 3D models by performing operations such as rotation, scaling, and adding individually colored markings via a ray-casting beam (see Figure 28 (b)).

To explore touchless interaction, one headset was modified with an ergonomic head strap and equipped with Ultraleap's Stereo IR 170 Evaluation Kit[2] (see Figure 28 (a)). This modification enabled the testing of interactive deformations of the 3D models, whereas controller input was limited to the playback of predefined animations. As hand interaction was also highly favored by instructors, it was retained as a primary input method.

Software development was conducted using Unity[3] on Microsoft Windows[4]. The application integrated the latest version of the MRTK

2 Ultraleap Ltd., `https://www.ultraleap.com`
3 Unity Technologies, `https://unity.com`
4 Microsoft Corporation, `https://www.microsoft.com`



(a) (b)

Figure 28: (a) User wearing Tilt 5 HMD with additional hand-tracking sensor using custom 3D printed mount; (b) Photograph captured through HMD.

3[5] for UI elements and utilized Ultraleap's Leap Motion Core Unity Plugin packages for precise hand tracking and seamless interaction.

The application was run on a PC with the following specifications: Intel Core i7-8700 3.7 GHz 12-core CPU[6], NVIDIA GeForce RTX 3090 Ti[7], 32 GB RAM, a 256 GB PCIE 3.0 SSD, and a 1 TB HDD.

6.2.1.3 *Player Roles*

The MR system defines two distinct player roles with varying levels of rights and control options: a moderator and up to three students. This role division was designed to facilitate structured seminar sessions while promoting dynamic group discussions. Each user is equipped with controller input to manipulate the 3D content, such as rotating, scaling, and marking. The moderator role includes extended functionalities, such as toggling annotations, enforcing a unified model alignment view for all participants, and locking the beam for specific tasks. Additionally, only the moderator has the ability to deform the model using hand interaction and control animations. This design decision was driven by both practical and pedagogical considerations. Supporting hand-tracking for four simultaneous users would have required additional sensors, increasing the technical complexity of the setup. Moreover, multiple users interacting with the model simultaneously could lead to visual obstructions and overlapping inputs, disrupting the collaborative experience. Furthermore, the proximity required for effective interaction within the system necessitated lim-

5 Microsoft Corporation, https://github.com/microsoft/MixedRealityToolkit-Unity
6 Intel Corporation, https://www.intel.com
7 NVIDIA Corporation, https://www.nvidia.com

iting direct hand interactions to the moderator, ensuring a smoother and more organized learning environment.

6.2.1.4 *Configurations*

Three potential positioning configurations of the game board were developed. In the first configuration, *Flat*, the game board lies on a table. In the second variant, *Poster*, the game board is aligned vertically. The third variant, *Curved*, enhances visualization by incorporating an additional retro-reflective foil wall for increased depth perception. An overview of the configurations in a group scenario can be found in Figure 29, which displays screenshots from the Mixcast Streaming Software[8].

6.2.2 *Evaluation*

With the transition of the application from a single-user (ILE) to a multi-user (CLE) system, designed to function both as a visualization aid and teaching tool, collaboration with domain experts became essential. To validate initial assumptions and evaluate the utility of the system's functionalities, qualitative feedback was gathered through a workshop with experts in anatomy education. This study sought to provide early-stage insights within an iterative development process, focusing on input modalities, role allocation, and gameboard/player positioning.

6.2.2.1 *Participants*

The workshop was moderated by a team comprising a software developer, a designer, and a cardiologist, with the goal of qualitatively assessing the visual and interactive aspects of the proposed VLE concept for seminar integration. Four anatomy experts (three female, one male) from the Institute of Anatomy our university participated in the study. The participants, ranging in professional experience from 3 to 37 years (Mdn = 14.5), included Research Assistants and Professors. They reported substantial teaching experience (Mdn = 17.3) across both theoretical and practical formats, such as lectures, seminars, and hands-on training in human medicine, macro- and neuroanatomy, and histology.

6.2.2.2 *Data collection*

To gain insights into expert behavior and acceptance of the new technology, the TAM questionnaire was employed. The quality of immersion was assessed using the ARIQ questionnaire. To collect comprehensive feedback and suggestions for improvement, experts were en-

8 Blueprint Reality Inc., `https://mixcast.me`

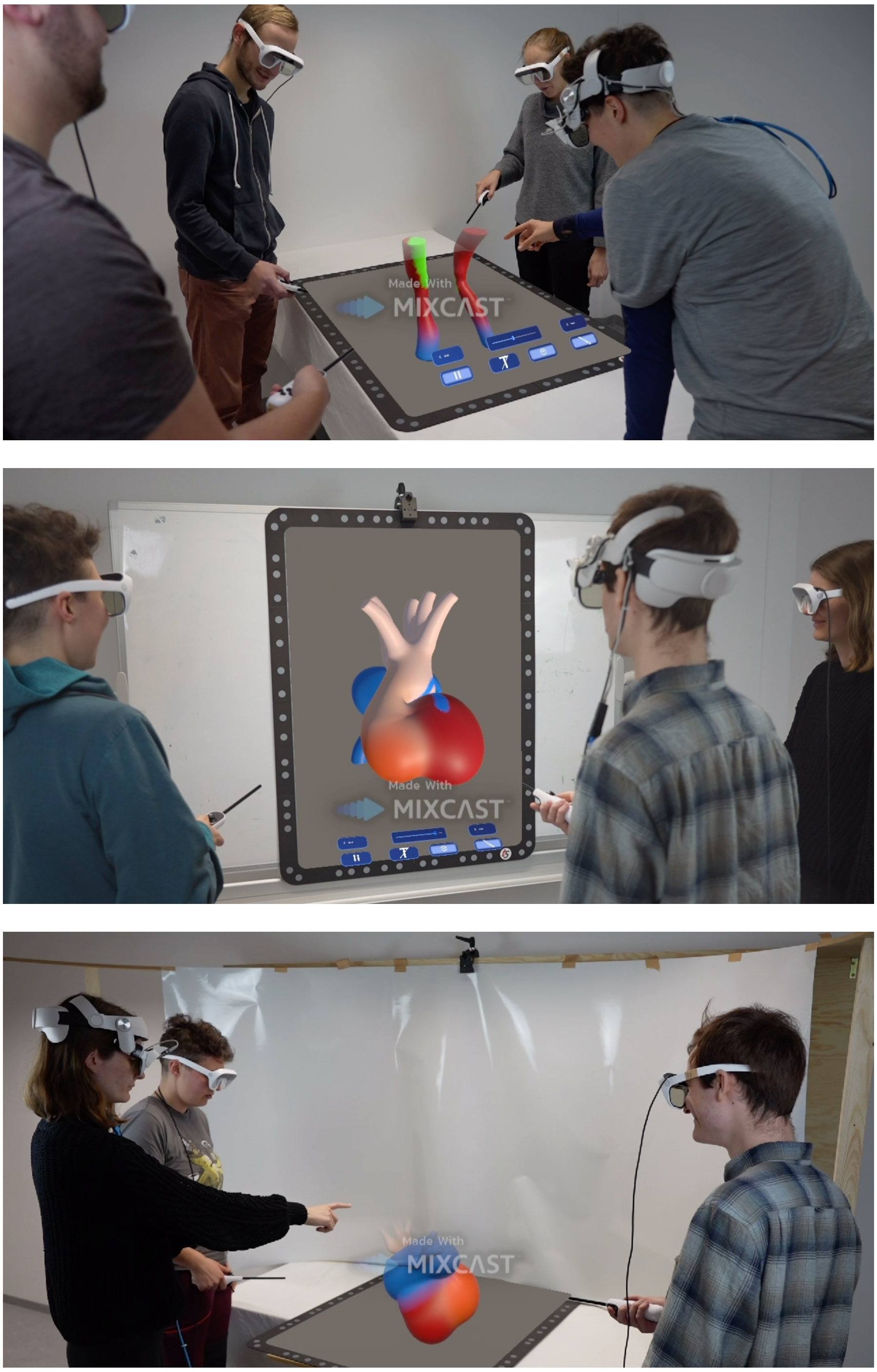


Figure 29: Screenshots from the Streaming Software used showing the evaluated configurations. From top to bottom: *Flat*, *Poster*, and *Curved*.

couraged to freely express their thoughts and discuss them within the group. Through audio recordings analyzed and statements were clustered and documented if at least two participants provided matching feedback.

6.2.2.3 *Procedure*

After welcoming the participants and obtaining their consent for data recording, the MR system was introduced. Each configuration (as outlined in Figure 29) was then presented and discussed sequentially: first the flat setup, followed by the poster configuration, and finally the curved display. Participants engaged with the implemented phases of embryonic heart development, taking turns in the roles of moderator and student. This allowed them to test various positions, input modalities, and all elements of the UI. There were no time constraints at each station. The workshop lasted approximately two hours, and participants received no material compensation for their involvement.

6.2.3 *Results and Discussion*

The subscales of the TAM achieved median ratings of 6.1 (SD = 0.2) for PU and 5.2 (SD = 0.2) for PEU, indicating general acceptance of this technology among the user group and aligning with the qualitative feedback. The ARIQ scored an overall 5.1 (SD = 0.4), placing it in the upper mid-range. This suggests a positive UX, correlating with a conducive learning environment. Among the configurations, experts ranked *Flat* first, followed by *Poster* in second place, and *Curved* last.

The TAP revealed feedback emphasizing the effectiveness of the MR concept for leading small groups through step-by-step explanations, fostering vivid understanding, and facilitating dynamic communication. The designated role distribution and the moderator's focus on visualization and interaction were praised as key factors enhancing the learning process. Experts valued the availability of two input options, particularly precise pointing and controller-based rotation/scaling. Hand tracking was deemed essential for all users, as it enhanced motor movements and understanding of physiological development, aligning with research on individual learning activities [63].

The *Flat* configuration received high praise for its equal visibility and space efficiency in seminar rooms. The *Poster* configuration, while suitable for larger groups, was noted for potential obstructions and discomfort, as users needed to look upward. The *Curved* configuration provided additional depth but was too large for seamless seminar integration and limited effective hand interaction due to the distance.

The ergonomics of the standard headset were criticized, with issues such as precise positioning requirements and discomfort for eyeglasses wearers. However, the modified headset successfully addressed these concerns. Visual cutoffs occurred when the model extended beyond the game board, and button operation difficulties were reported during hand interactions, especially with the *Poster* and *Curved* setups. Enabling distance interaction via controller input could improve the UX. Partial hand-tracking issues were observed, possibly linked to headset alignment with the retro-reflective material. Improved placement near the game board could resolve this but might restrict the interaction range. Seating users instead of requiring them to stand was suggested to improve ergonomics and visualization.

A four-player limit was noted as a drawback, but it was acknowledged that the learning approach is better suited for smaller groups. These insights provide valuable guidance for optimizing future iterations of the system.

6.2.4 *Conclusion and Future Research*

This experiment presents an MR-based CLE aimed at fostering a participatory learning environment, using embryonic heart development as a case study. Through expert evaluations across three configurations, interaction modalities and player roles were assessed for potential integration into anatomy seminar sessions. Experts recommended expanding the concept to include the development of other complex organs, such as the intestine or brain.

The current CLE is limited to four simultaneous users, highlighting the need to explore a server-based solution for increased capacity in the future. This evaluation aims to serve as the foundation for an in-depth analysis of system requirements and the development of an optimized didactic concept.

6.3 EXPERIMENT 5

This section presents the second of three experiments in this chapter, focusing on the technical advancements of the MR system described in the previous section. Building on the findings from *Experiment 4*, which involved an expert workshop to assess didactic strategies, visualization methods, and input possibilities, an iterative design process was employed to refine the system. Feedback from the workshop guided the selection of the most promising visualization techniques and interaction modalities. This experiment evaluates the *updated system* through a user study involving students and moderators, conducted within a simulated session modeled after an anatomy seminar.

6.3.1 *Material*

The *updated system* underwent significant improvements to its features and overall functionality. Based on the evaluation of the *original system*, the flat game board configuration was identified as the most suitable setup due to its superior visibility and interaction opportunities for multiple users. Consequently, this configuration was adopted in the *updated system*.

6.3.1.1 *Visualization*

The *updated system's* core structure remained consistent with the version described in *Experiment 4*. However, performance issues observed during the expert evaluation—stemming from the complexity of visualizing and animating internal structures using MudBun[9]—necessitated adjustments. To resolve these issues, a collection of individual meshes was implemented, loaded sequentially. This approach enabled smooth animations compiled from approximately 200 exported meshes for each 3D model stage.

6.3.1.2 *Controller Input Instead of Hand Interaction*

Unlike the *original system*, controller input was implemented for all users in the *updated system*. Although hand interaction is ideal for understanding morphological changes, having multiple users around the game board often resulted in obstructions and visibility issues. Tracking challenges due to the distance between users and the play area further complicated hand interactions.

A new feature, previously absent in the *original system*, allows users to scale the 3D model using the controller. By pressing the corresponding buttons, the model size can be adjusted. However, when scaled too large, the model can extend beyond the game board's edges, leading to frame cancellation—a phenomenon where virtual objects are cut off by the screen's edges. To mitigate this, alpha blending [141] was applied using custom shaders, fading the model's edges when exceeding the board boundaries (see Figure 30).

6.3.1.3 *Paintings and Cursor*

Due to the transition to individual meshes for the 3D models, the inactive painting (annotation) feature was omitted in the *updated system*. Discrepancies occurred when switching between models, making it impractical. Experts also recommended better differentiation between users to avoid overlapping colors during painting.

As an alternative, a continuous, directed beam was introduced from the controller. When the beam intersects the surface of a mesh, it

9 http://longbunnylabs.com/mudbun

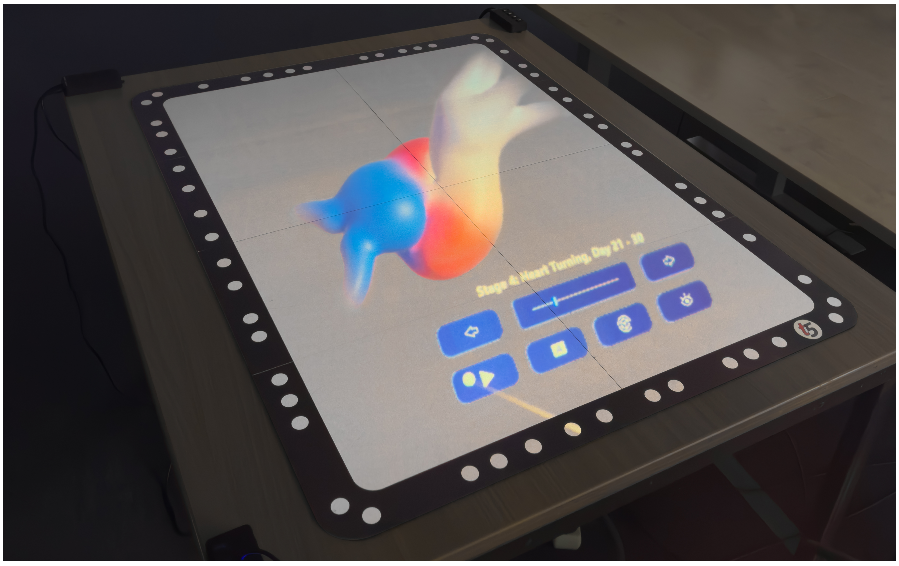


Figure 30: Photograph through HMD showing frame cancellation when the model exceeds the game board's edges. Alpha blending with custom shaders fades the edges to mitigate this effect. Adapted from Schott et al. (2024a) [Core2]. Licensed under CC BY 4.0.

generates a cursor at the intersection point. Activated by the moderator interface, this feature dynamically displays labels with medical descriptions of cardiac structures. These labels are aligned to the user's view and positioned to the left or right of the model based on the structure's location. Cursors are numbered for user identification, facilitating communication. The three basic interactions—rotation, scaling, and cursor—are accessible to all users.

6.3.1.4 *Changes in the UI*

The moderator's operating interface retained its core elements, but consistent icons were added for clarity. Activating annotations now displays a white orientation frame with medical position labels, alongside the name and timeframe of the current phase.

A new feature was introduced specifically for the study: the ability to display example questions and answers related to the system's content via controller buttons. This assists the moderator during sessions by providing contextual information.

In the final stages of development, the heart model was modified to enable internal visualization. The model consists of a fixed mesh representing half of the heart, with animations added as a separate object. During playback, the fixed mesh becomes transparent, offering an unobstructed view of internal processes from all angles (see Figure 31).

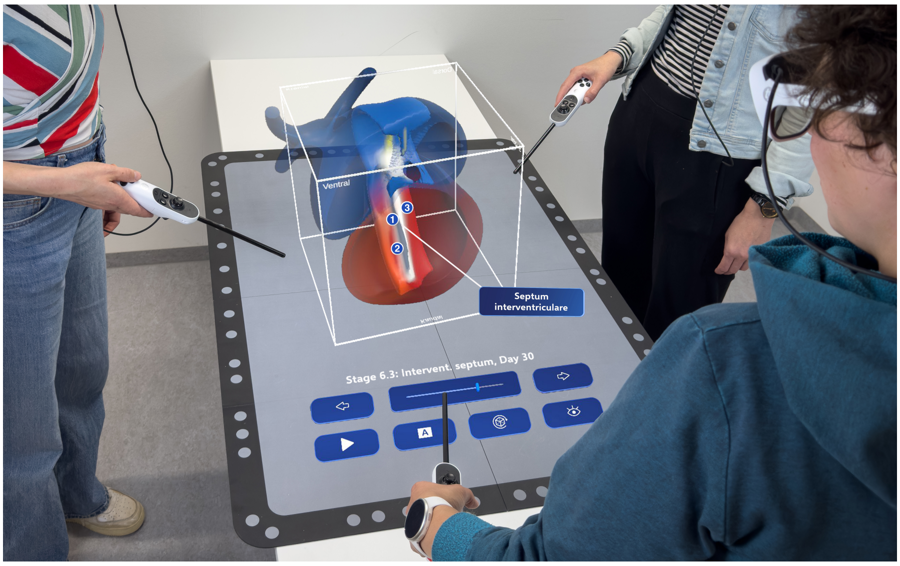


Figure 31: Photomontage of the setup, with the moderator interacting with the UI, displaying all control elements and overlay information such as the orientation grid and labels. Adapted from Schott et al. (2024a) [Core2]. Licensed under CC BY 4.0.

6.3.2 *Evaluation*

A user study was conducted to evaluate *Usability*, gaming experience, immersion, and technology acceptance across different user groups, aiming to gain deeper insights into the integration of such systems into anatomy education. The study sought to assess the perspectives of both moderators and students.

6.3.2.1 *Study Design*

Four sessions were carried out, each involving three participants: one acting as the moderator and two as students. The reduced number of student participants was intended to stabilize the system's performance. Moderators were required to have teaching experience within a broad medical context. A total of four moderators (three male and one female) participated, with a mean age of 27.75 years and an average of 3.38 years of teaching experience. Additionally, eight medical students were recruited, seven of whom were female. The students had an average age of 25.83 years (with one student not reporting their age).

6.3.2.2 *Procedure*

In each session, a seminar was simulated with a moderator guiding two students. The study duration was 60 to 90 minutes for the students, while moderators were asked to arrive 30 minutes earlier for preparation and briefing. During this preparation period, informed consent and demographic information were collected, and the moderators were trained in using the application. Moderators were provided with potential questions developed in collaboration with anatomy education experts, covering the scope of the application. These included one or two questions per developmental stage, designed to stimulate discussion, though their use was not mandatory. The list of questions is available in Section 9.6.

While moderators were encouraged to familiarize themselves with the topic, the questions and answers were available in printed form and could also be accessed virtually during the session. After the students arrived, their informed consent and mandatory study documentation were collected, followed by an introduction by the study instructor. The moderator then independently introduced the application and explained the procedure.

Once the HMD were on, a 3D model of the controller was displayed to illustrate its functions. When all participants were ready, the content phase began, guiding the group through the various stages of heart development. Throughout the session, the moderator facilitated lively communication and interaction within the group.

6.3.2.3 *Measures*

At the end of the seminar, all participants completed several questionnaires. *Usability* was assessed using the SUS. The level of immersion experienced by users in the MR environment was evaluated using the ARIQ. Given that the type of MR learning platform was novel to the participants, the TAM was employed to measure PU and PEU, which are key determinants of users' attitudes and behavioral intentions. Additionally, the *Core*, *Social Presence*, and *Post-game* components of the GEQ were used to evaluate the overall experience of the application. Finally, the experiment concluded with an open discussion and interviews, during which qualitative feedback was collected on the interaction and visualization schemes of the MR system, as well as its feasibility as a potential part of the curriculum.

6.3.3 *Results*

This section summarizes the findings of the user study, encompassing both quantitative and qualitative measures.

6.3.3.1 *Immersion*

Results of the ARIQ questionnaire indicated that a comparable (medium-high) degree of immersion was perceived by both study groups while using the MR application (Mod: $M = 5.11, SD = 0.66$; Stud: $M = 4.80, SD = 0.85$; see Figure 32 (a)). This outcome was expected, as immersion is primarily influenced by the technical factors of the hardware, which remained identical for both groups.

6.3.3.2 *Usability*

Promising *Usability* ratings were recorded for the prototype across both user groups. Moderators provided slightly lower SUS scores ($M = 76.25, SD = 9.24$) compared to students ($M = 83.13, SD = 6.51$). Nevertheless, both values exceeded the generally acknowledged thresholds for good to excellent *Usability* (see Figure 32 (b)).

6.3.3.3 *Technology Acceptance*

The TAM assessments (see Figure 32 (c)) reflected these findings. Lower scores indicated better PU and PEU. For both metrics, the application was rated more favorably by the student group (PU: $M = 24.65$, $SD = 16.43$; PEU: $M = 32.29$, $SD = 15.99$) compared to the moderator group (PU: $M = 34.72$, $SD = 33.37$; PEU: $M = 52.03$, $SD = 33.59$). Although good scores were generally achieved, the only moderate PEU rating for the moderator role indicated room for improvement.

6.3.3.4 *Gaming Experience*

A detailed overview of all GEQ subscales is presented in Figure 33. A higher sense of flow was reported by moderators ($M = 2.20, SD = 0.94$) compared to students ($M = 1.74, SD = 1.02$), suggesting that moderators were more engaged while conducting the seminar. Students reported a minimal degree of challenge ($M = 0.78, SD = 0.25$), whereas moderators perceived the seminar as considerably more challenging ($M = 1.80, SD = 0.67$). Negative affect, i.e., negative emotions towards the application, was rated low across both groups (Mod: $M = 0.50, SD = 0.54$; Stud: $M = 0.59, SD = 0.97$). Positive affect, i.e., positive emotions towards the application, was rated similarly high (Mod: $M = 2.80, SD = 0.71$; Stud: $M = 2.90, SD = 0.96$).

Behavioral involvement, i.e., the influence of users on each other's actions, was perceived as higher by moderators ($M = 2.75, SD = 0.50$) than by students ($M = 1.50, SD = 0.77$). This finding suggests that the teaching functionalities provided to the moderator successfully fulfilled their intended purpose, as moderators perceived their actions as having a greater impact on students.

The *Post-game* GEQ component indicated minimal difficulty in transitioning from MR back to real life (Mod: $M = 0.67, SD = 0.72$; Stud: $M = 0.38, SD = 0.33$). The application was not perceived as tiring (Mod: $M = 0.63, SD = 0.95$; Stud: $M = 0.38, SD = 0.88$). Additionally, positive experiences (Mod: $M = 1.63, SD = 0.79$; Stud: $M = 1.67, SD = 0.81$) outweighed negative ones (Mod: $M = 0.54, SD = 0.64$; Stud: $M = 0.25, SD = 0.41$).

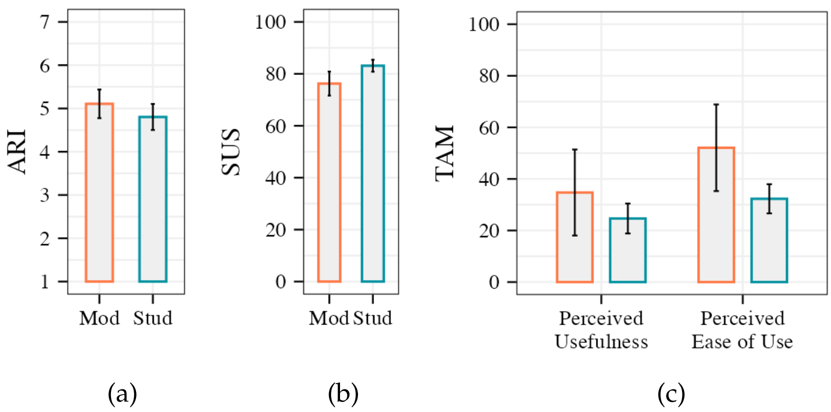


Figure 32: ARI (a), SUS (b) and TAM (c) ratings of the ■ Moderator ■ Student groups. Mean values and standard error bars. Adapted from Schott et al. (2024a) [Core2]. Licensed under CC BY 4.0.

6.3.3.5 *Qualitative Feedback*

During the interview phase, the animated 3D visualization of the embryonic heart was rated as a significant improvement over conventional 2D cross-sections found in textbooks. Particular emphasis was placed on the final stages of the application, where changes inside the heart were visualized using a halved, transparent heart model. This feature was highlighted as especially helpful. However, it was suggested that the transparency effect should remain active even when the animation is not being played. Additionally, it was recommended that more complex information, such as visualizing blood flow, be incorporated into the heart model to enhance its educational value.

The system's interaction capabilities were assessed as intuitive and easy to learn. The variety of interaction options was positively received, with participants noting that the available features were sufficient to engage with the heart visualization effectively without overwhelming them with excessive choices. Occasionally, alignment issues with the heart model were reported, but these were found to be mitigated by adjusting the user's head position and perspective.

The MR hardware was generally rated positively. However, some discomfort was noted, particularly with the HMD being uncomfortable around the ears and warming in the forehead area. The resolution of the HMD was also perceived by some as too low to read text properly.

Regarding the systems's suitability as a VLE, participants found the concept to be highly useful for anatomical training. It was envisioned as a valuable extension of, or even a replacement for, the dissection course. However, limitations related to group size were identified, including restricted space around the area of interest and visual clutter caused by multiple interaction markers on the model.

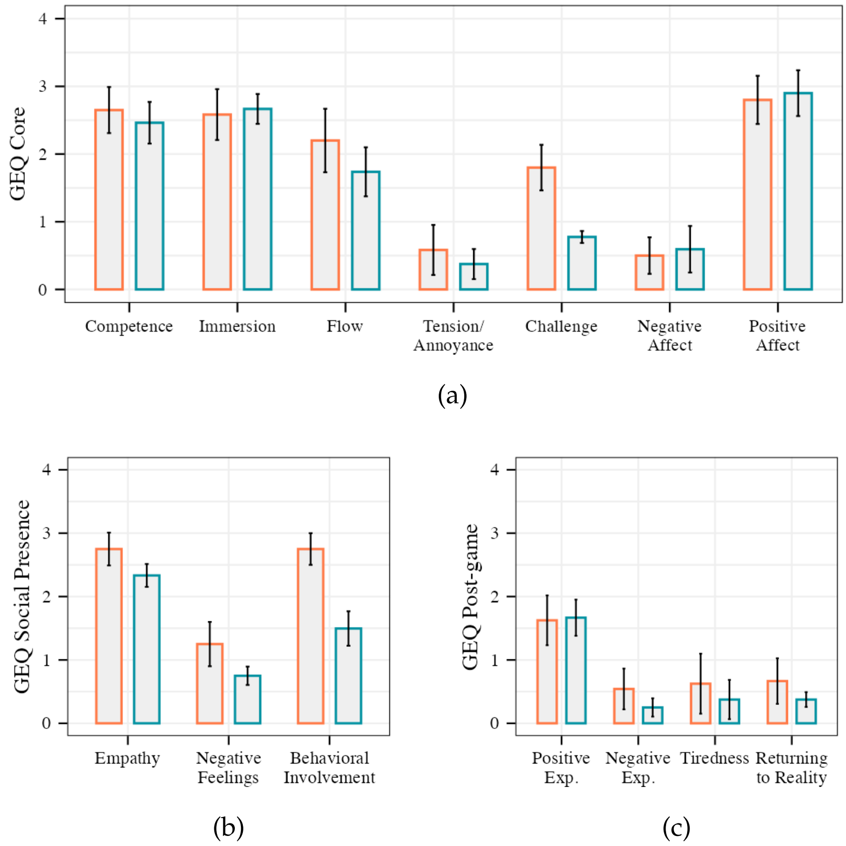


Figure 33: GEQ *Core* (a), *Social Presence* (b), and *Post-game* (c) module ratings of the ■ Moderator and ■ Student groups. Mean values and standard error bars. Adapted from Schott et al. (2024a) [Core2]. Licensed under CC BY 4.0.

6.3.4 *Discussion & Conclusion*

Data from the quantitative questionnaires indicated that the CLE was positively perceived by both moderators and students. High levels of competence, flow, immersion, and positive emotions were reported, accompanied by low levels of tension, challenge, and negative feelings. Favorable scores in the TAM further suggested a high level of acceptance, which is likely to contribute to successful learning outcomes.

Students rated the *Usability*, usefulness, and ease of use of the application more favorably than the moderators. This difference is likely due to the moderators' additional *Subjective Workload* in controlling the interface while conducting the seminar. This increased responsibility may have negatively influenced their overall perception of the MR environment, as reflected in their higher challenge ratings on the GEQ *Core* subscale.

Qualitative feedback also supported the positive perception of the application while highlighting areas for improvement. Participants suggested additional visualization features, such as blood flow, and identified challenges related to larger group sizes. Conventional seminars often involve more students than the current hardware supports, making small-group settings more suitable for the CLE. Expanding the number of simultaneous users would enhance the platform's potential for broader application.

The study was limited by a relatively small sample size, which precluded further statistical analyses. Future work should include a larger participant pool to generate more robust findings and gather additional feedback. This is particularly important for the moderator group, which included only four participants. A more diverse gender distribution should also be considered to mitigate potential sample bias.

During the simulated seminar, moderators were encouraged to foster consistent group communication. The provided example tasks and questions were used to varying degrees, with moderators demonstrating diverse approaches, including direct questioning, posing open-ended group questions, and addressing individuals directly. Their teaching experience allowed them to adapt effectively to different situations and questions. As a result, not every developmental step was presented in the same manner, highlighting the flexibility of the system. The small group size contributed to a lively and engaging atmosphere, ensuring active participation from all attendees. In larger groups, such interactivity and communication might be less effective, as not every participant would have the opportunity to contribute. Additionally, peer-to-peer interactions were observed, with students attempting to explain concepts to one another, further enriching the learning experience.

In conclusion, the CLE demonstrates strong potential as a supplemental tool for anatomy seminars. It offers a positive and supportive UX, with considerable educational and teaching value. However, further refinements and development are needed to fully realize its potential.

6.4 EXPERIMENT 6

This section presents the final of three experiments in this chapter, where insights from the previous experiments were applied to a new MR platform. The primary focus is a comparative evaluation of the ILE introduced in Chapter 5 and the collaborative approaches explored in the preceding experiments.

6.4.1 *Motivation*

Building on the technical advancements and didactic strategies developed in *Experiments 4 and 5*, *Experiment 6* addresses a critical gap in the literature: the direct comparison of individual and collaborative learning environments within MR-based anatomy education. This experiment shifts from refining specific system features to evaluating their educational effectiveness in different formats.

The two approaches reflect distinct pedagogical paradigms, grounded in constructivist learning theory. The ILE supports self-directed learning, allowing students to actively construct knowledge by exploring independently at their own pace. Constructivist theory emphasizes that learners develop understanding through personal experiences and problem-solving, making the ILE particularly suited for fostering individual exploration and deep engagement with the content [63, 109].

In contrast, the CLE incorporates teacher- and learner-centric roles, fostering SI and collaboration within a shared environment. This aligns with the constructivist principle that appropriate guidance and group interaction can enhance learning, especially during the early stages of knowledge acquisition [114]. By simulating a traditional seminar structure, the CLE provides opportunities for dynamic communication, shared problem-solving, and role-based participation, making it an ideal platform for collaborative learning.

By comparing these two approaches in a user study involving $n = 90$ medical students, this experiment investigates the impact of MR on knowledge acquisition, retention, and *Usability*, while examining factors such as UX, *Social Presence*, and *Subjective Workload*. The findings aim to provide deeper insights into how MR can complement conventional learning, enhancing both individual and group-based educational experiences, and offer a broader perspective on the integration of immersive technologies into medical curricula.

6.4.2 *Material and Methods*

The enhanced *CardioGenesis4D*, described in Chapter 5, and the *updated system* from *Experiment 5* were transitioned to a unified hardware platform: the Meta Quest 3[10]. The Quest 3 employs a pass-through-based (VST) MR approach, in which real-world visuals captured through its cameras are blended with interactive virtual content to create a semi-immersive learning environment. This approach was chosen to ensure comparability between the two concepts, as the primary focus was on evaluating the didactic strategies rather than the technical differences of the systems.

The switch to new hardware was driven by several limitations observed with the Tilt Five system. The Tilt Five had a lower resolution (720p), which resulted in blurry representations of both text and models. Tracking was occasionally lost, and the HMD were noted to be uncomfortable, with reports of warmth in the forehead area. Ad-

10 https://www.meta.com

ditionally, the Tilt Five required a wired connection to a powerful PC, limiting portability and ease of use.

The Meta Quest 3 offers several advantages over the Tilt Five. With its VST MR mode and controller tracking, it provides a similar interactive experience while delivering significantly higher resolution, enhancing the clarity of both textual and visual elements. Furthermore, the wireless design of the Quest 3 increases mobility and *Usability*, making it a more practical choice for educational applications in settings such as seminar rooms.

Accessibility remained a key consideration. The VST-based MR approach is less isolating and has the potential to reduce simulator sickness compared to fully immersive virtual environments, thereby increasing acceptance among the targeted user groups. Additionally, as the application is designed to be used in familiar settings, such as a seminar room, this setup allows users to focus more effectively on the learning content without the need to adapt to a fully immersive MR environment. An overview of the implemented ILE and CLE is provided in Figure 34.

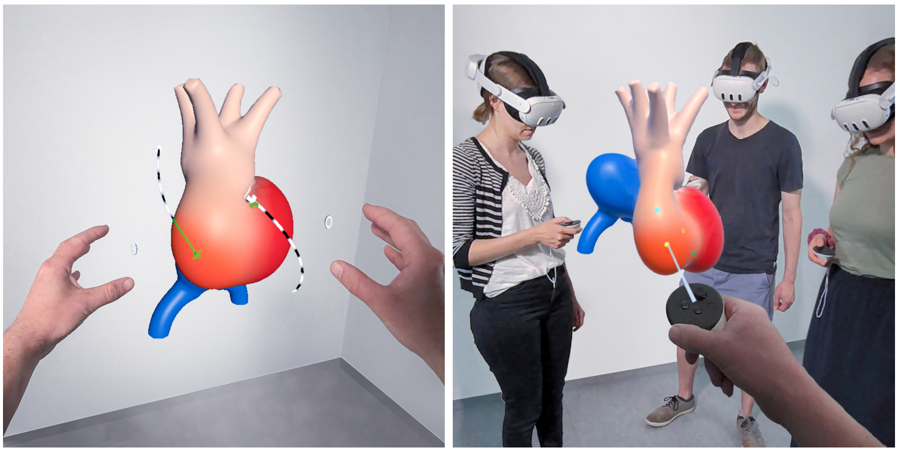


Figure 34: Screenshots from the Quest's pass-through mode of the evaluated systems (slightly edited for improved aesthetics). Left: ILE, where a single user manipulates a 3D heart model using hand interactions. Right: CLE from a student's perspective, showing multiple users interacting with a heart model using controller input. Adapted from Schott et al. (2024b) [Core1]. Reused in accordance with ACM's policy for open access articles.

6.4.2.1 *Individual Learning Environment*

The ILE is a standalone, single-user application designed to facilitate individual exploration. Learners can familiarize themselves with each developmental stage step-by-step by interacting with virtual heart models at their own pace, allowing for detailed examination of specific aspects of embryonic heart development.

The UI and content were largely maintained as described in Chapter 5. The ILE consists of three main interaction components: a timeline providing an overview of all phases, with the current stage indicated by an arrow; an interactive 3D model of the heart with various manipulation and selection options; and minor improvements to the control elements (see Figure 35).

Notable enhancements include the addition of a separate button to activate the audio guide, now accompanied by a widget that provides control over the audio content. Additionally, an orientation frame with anatomical directional labels can now be displayed where labels were previously toggled on or off.

All elements are automatically angled and oriented toward the user for optimal visibility. These elements can be freely repositioned in space through raycast interaction. Furthermore, an ellipse on the floor allows horizontal and vertical adjustment of the entire scene, providing enhanced flexibility and *Usability*.

Software development was carried out using Unity v2022.3.12f1[11] on Microsoft Windows 10 Pro (build 19045). MRTK v2.7.2[12] was employed for hand interaction and UI building blocks. Oculus Integration v57[13] was installed for hand tracking and pass-through functionality, and integrated into MRTK.

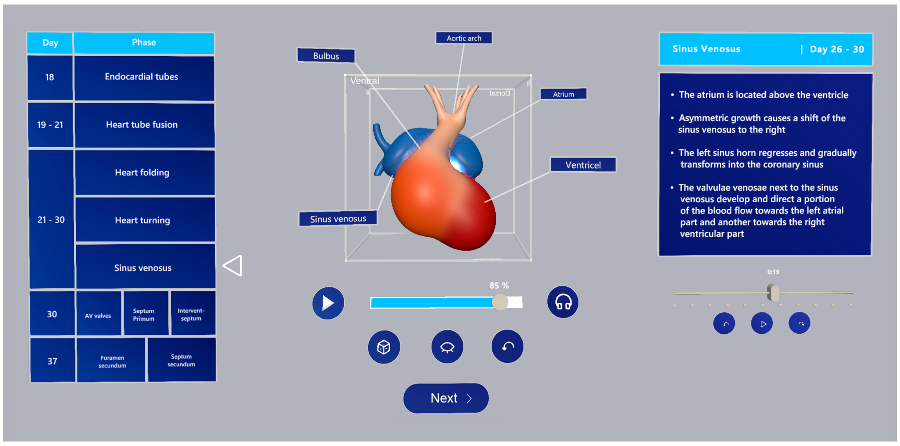


Figure 35: Isometric view of the arrangement of all interactive components in the UI of the *ILE*. Reprinted from Schott et al. (2024b) [Core1]. Reused in accordance with ACM's policy for open access articles.

6.4.2.2 *Collaborative Learning Environment*

The CLE is a multi-user application designed to enable real-time collaboration between students and instructors. Modeled after a traditional seminar, the application assigns specific roles: the lecturer acts

11 https://unity.com/de
12 https://learn.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/mrtk2
13 https://assetstore.unity.com/packages/tools/integration/oculus-integration-deprecated-82022

as a moderator and knowledge guide, while students engage as active participants. This structure fosters peer interaction, idea sharing, questioning, and collaborative problem-solving. By leveraging advanced visualization and interaction capabilities, the CLE enhances dynamic group engagement, creating a more interactive and enriching learning experience.

Classic controller inputs were implemented in the CLE for interactions: rotation (joystick), scaling (A/B buttons), and spatial movement (trigger and joystick). These inputs were chosen to address potential challenges such as obstruction of the 3D model and varying distances between users in the room. A directional beam is continuously emitted from each controller, generating a cursor at the point of intersection with the mesh surface. Each cursor is uniquely colored, replacing the previously used numerical markings, which were occasionally hard to read. This adjustment simplifies communication, as users can now easily reference their colors, enhancing collaborative interactions. All users have access to these interaction options.

Studies contrasting hand-tracking with controller-based interactions, such as those by Adkins et al. [5] and Luong et al. [159], emphasize that the choice of interaction modality has limited impact on collaboration but significant implications for precision and user fatigue. These findings support the use of controllers in tasks requiring accuracy and minimal exertion, such as medical simulations.

The lecturer's UI was slightly optimized, retaining core functionalities while improving layout organization for better *Usability* (see Figure 36). A numerical indicator on the interface now displays the total number of connected users, replacing the less frequently used block annotation feature from *Experiment 5*. This update emphasizes tracking participant engagement and provides a more streamlined and practical seminar experience.

Software development was conducted using Unity v2021.3.35f1[2] on Windows 10 Pro[14]. MRTK 3[15] was used to implement buttons and UI elements. The Meta XR All-in-One SDK v63.0[16] was utilized for controller tracking and to enable pass-through functionality. The CLE setup was implemented using Photon Fusion 2[17], configured as a host-client system. Users transmit data such as animation progress, stage number, cursor position and rotation, and moderator settings (e.g., play/pause, annotations, forced perspective) to the host, which processes these inputs and broadcasts updates to all connected clients.

14 Microsoft Corporation, `https://www.microsoft.com`
15 Microsoft Corporation, `https://learn.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/mrtk3-overview/`
16 Meta Platforms, Inc., `https://assetstore.unity.com/packages/tools/integration/meta-xr-all-in-one-sdk-269657`
17 Exit Games GmbH, `https://doc.photonengine.com/fusion/current/fusion-intro`

This setup ensures consistent visualizations and interactions across all instances of the application.

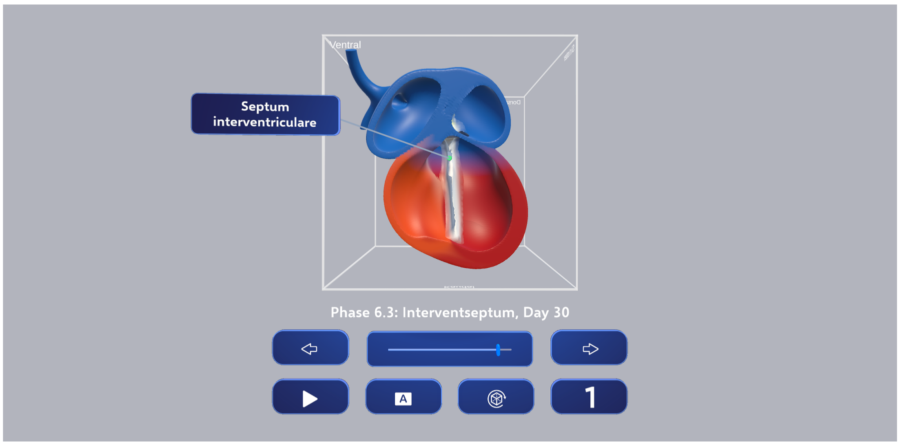


Figure 36: Overview of the UI in the CLE from instructor's perspective. Here, students are presented with a comparable view that is, however, missing the interactive elements, e.g., buttons and the slider. Reprinted from Schott et al. (2024b) [Core1]. Reused in accordance with ACM's policy for open access articles.

6.4.3 *Evaluation*

A user study was conducted to investigate the differences in educational outcomes between the two MR-based didactic concepts (ILE and CLE). The primary objective was to evaluate the educational effectiveness of each application and to analyze the influence of factors such as *Usability*, UX, SI, CP, and *Subjective Workload* on the learning process. Learning success was assessed through a combination of qualitative and quantitative measures, including pre- and post-interaction knowledge assessments and a series of questionnaires utilized during the experiment. To ensure reproducibility and facilitate further research, both Unity projects, along with the acquired data and analysis scripts, have been made publicly available in an online repository[18].

6.4.3.1 *Study Design*

The study was scheduled at the beginning of the summer semester 2024, just before the exam period, to evaluate students at their peak level of knowledge. The entire cohort of medical students participating in the anatomy seminar was evenly divided across two consecutive weeks. This arrangement allowed the study to be conducted during this timeframe, enabling the inclusion of students with varying levels of prior knowledge. While both groups took the actual exam

18 https://github.com/ovgu-var-labs/cardiogenesis

a few days after the study, this data was not analyzed due to data protection concerns.

A between-subjects design was employed, where all participants experienced both MR environments (ILE and CLE), but only assessed their respective first environment. Participants were invited to the study via the internal course management system and were assigned an ID prior to the study. They were also required to complete a pre-study knowledge test.

To ensure consistency and reliability, each session was monitored by experienced general medical or anatomy lecturers who were familiarized with the study procedure and the MR applications. Technical support was available to address any hardware or software issues. An overview of the study design is provided in Figure 37.

Ethical approval was obtained under Nr. 72-24 from our university's ethics committee. Participants provided informed consent, were briefed on the general purpose and specific procedure of the study, and were offered a debriefing at the end of the experiment.

6.4.3.2 *Hypotheses*

Several hypotheses were formulated to investigate the impact of MR applications on medical education, specifically in the context of embryonic heart development. These hypotheses were designed to compare the educational outcomes and experiences provided by the MR environments to those achieved through traditional seminar-based learning.

h1 Learning outcomes in the MR applications are not inferior to those in traditional seminars.

h2 There are significant differences in the effectiveness of learning between ILE and CLE.

h3 There are no significant differences in subjective *Usability* ratings for both applications.

h4 Task load is significantly higher in the single-user application.

h5 UX is significantly higher in single-user applications.

h6 SI/CP are significantly higher in CLE.

A comprehensive set of variables was analyzed to evaluate the applications thoroughly. Two independent variables were examined. The first was the *application type*, which included two levels: ILE and CLE. The second was the *education method*, comparing MR-based learning (Group A) to traditional seminar-based learning (Group B).

The dependent variables included the EKT, as described in Section 5.3.2.3 (see also Section 9.5), to measure knowledge acquisition

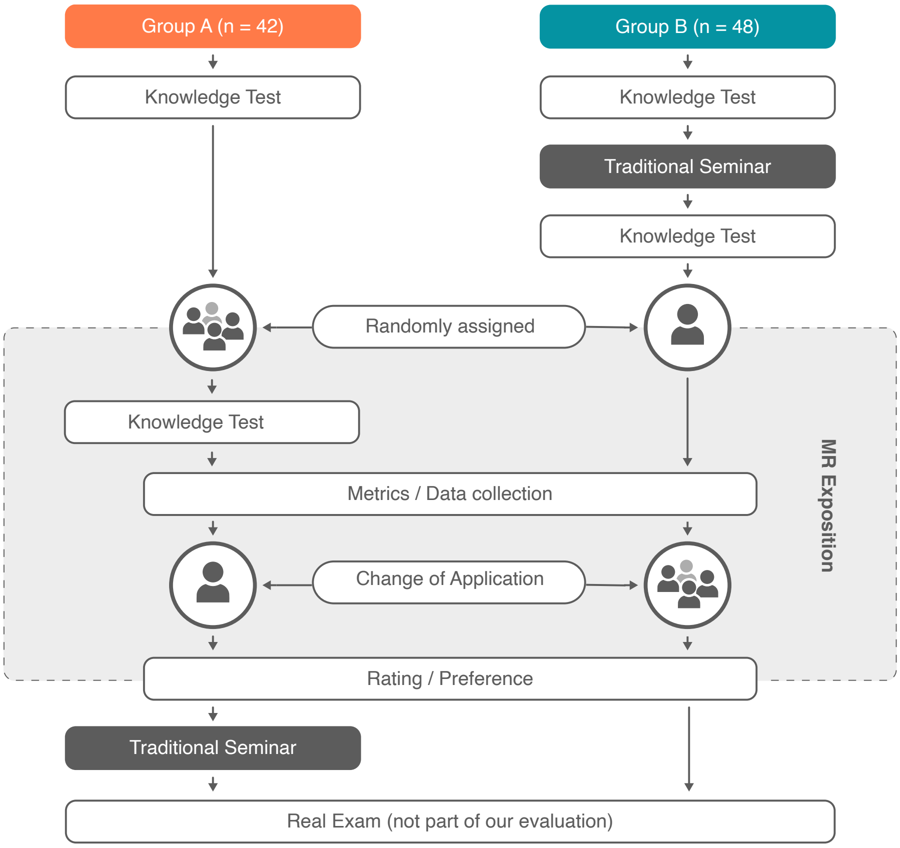


Figure 37: Flowchart illustrating the study design for evaluating MR applications. The study cohort was divided into two groups: Group A ■ (n = 42) and Group B ■ (n = 48), each undergoing a sequence of activities including a knowledge test, a traditional seminar, and MR exposure. On the day of the study, all participants provided informed consent and were surveyed for demographic data. Group A started with the knowledge test followed by the MR session, while Group B began with a traditional seminar before the MR session. The assignment of ILE or CLE was randomized. Each group underwent a different application procedure to ensure balanced exposure. Participants' performance was assessed through additional data collection. Adapted from Schott et al. (2024b) [Core1]. Reused in accordance with ACM's policy for open access articles.

and retention. The SUS was employed to provide a *Usability* score for the applications, while the N-TLX was used to assess *Subjective Workload* and stress levels.

To investigate CP and SI, the study examined how collaboration with real individuals in the CLE, such as a lecturer, compared to the influence of a computer voice (audio guide) in the ILE. This was assessed using CP and SI questions, as described in Section 2.4.2.2. To evaluate a broad range of UX dimensions, the UEQ was applied, offering insights into various UX qualities and enabling straightforward comparisons across different conditions.

In addition to standardized measurement instruments, participants provided ratings and recommendations for the applications based on their individual preferences. These were collected using three 7-point Likert-type scale questions with endpoints -3 (ILE) to 3 (CLE):

- *General Preference* (GP): "Which application did you like better?"
- *Ease-of-Use* (EU): "Which application was easier to use?"
- *Learning Preference* (LP): "Which application would best support your learning?"

Additionally, the educators who supported this study and regularly conduct similar anatomy teaching sessions in real-life settings were given the opportunity to provide feedback. A questionnaire was sent to them two weeks after the conclusion of the study to gather their insights and recommendations (see Section 9.7).

6.4.3.3 *Participants*

Out of approximately 200 first-semester medical students enrolled at our university, 90 participants were recruited for the user study conducted within the Anatomy seminar. Recruitment focused on feedback about the course management system provided by the lecturers. Participants were compensated with €10 for their involvement. The gender distribution (59 female, 31 male) represented approximately 66% female students, aligning with the average first-time enrollment rates in germany. Additional demographic data and the distribution of participants across the individual study groups are presented in Table 11.

The demographic assessment was conducted based on the template provided in Section 9.2. Similarly to the approach outlined in Chapter 5, self-assessments of General Embryology Knowledge (GEK) and Heart Embryology Knowledge (HEK) were collected using the same scale, as described in Section 5.3.2.4.

6.4.3.4 *Study Apparatus*

The study was conducted in two seminar rooms, each measuring 20 to 30 m^2, located at the Institute of Anatomy at the university. These rooms accommodated three students and one instructor or study leader each. Floor markings were placed to designate participant positions in the ILE setup, ensuring adequate safety distances. In the CLE room, a central marking was designated for placing the virtual heart model. Participants had at least a 2 m x 2 m area for free movement.

Each room was equipped with four Meta Quest 3 HMDs, with six headsets allocated for participants, one for the instructor, and one as a backup. To ensure prolonged battery life, the headsets were equipped

Table 11: Demographic participant data. Sample size and gender distribution are quantities; all other values are medians. Reprinted from Schott et al. (2024b) [Core1]. Reused in accordance with ACM's policy for open access articles.

	Group A		Group B		Total	
Variable	**ILE**	**CLE**	**ILE**	**CLE**	**ILE**	**CLE**
Sub-sample size	17	25	26	22	43	47
Number of women	15	16	16	12	31	28
Age	21	20	21	21	21	21
Technical affinity	3	4	3	3	3	4
MR experience	1	2	1	1	1	1
Gaming experience	2	2	3	2	2	2
GEK self-assessment	2	3	3	3	3	3
HEK self-assessment	2	2	3	3	2	2

with comfort head straps and additional batteries. For sanitary purposes, silicone face protectors were attached to each HMD.

Two QR codes were affixed to the walls of each room. QR Code 1 linked to an online survey for evaluating the respective application, while QR Code 2 was used for post-experiment feedback and ranking after participants had tested both applications.

Due to the size of the participant sample and the limited timeframe of four days, the study was supported by six instructors. All instructors had teaching experience and were briefed on the study procedures and trained in handling the hardware and software. Additionally, technical supervisors were available to address any hardware or software issues. This setup allowed for simultaneous assessment of up to six study participants, ensuring smooth and efficient operation of the study.

6.4.3.5 *Procedure*

Upon arrival, participants entered the designated room for ILE testing, where they were briefed on the study procedures, provided informed consent, and completed compensation forms. Completion of the embryology knowledge test sent via email prior to the study was mandatory; participants who had not completed it were required to take the test on-site.

Participants were then divided into two groups of three: one group remained in the ILE room, while the other was escorted to an adjacent room for CLE testing. In cases of participant cancellations, adjustments were made to ensure that each CLE session included three participants. The study instructor in the ILE room was responsible for supervising both the process and the technology.

In each testing room, participants scanned a QR code using a personal or provided mobile device to access an online survey. They entered their assigned ID and completed demographic information. Depending on the experimental sequence, which might have included a conventional seminar prior to MR exposure, participants could also retake the embryology knowledge test.

After completing the survey, a prompt indicated that the application was ready to start. Before the session began, participants received a briefing on the procedure, adjusted their headsets (e.g., headband, eye distance), and were informed of precautions against cybersickness and the option to withdraw from the study at any time.

ILE PROTOCOL The HMDs were donned by the participants, who lined up in previously marked positions on the floor. This ensured even spacing and sufficient freedom of movement. The instructor had pre-initiated the application and provided instructions on hand interactions. Participants proceeded to a tutorial scene, which introduced the basic operating elements and interaction options. Text fields within the tutorial were utilized to verify text readability and make any necessary adjustments to the headset.

Subsequently, the learning application was activated via a button, allowing participants to freely explore the content. No time limit was imposed, and participants had the option to repeat the session if desired. Upon signaling completion, participants removed their headsets and completed the online survey. They were prompted to identify the application they had just tested and then provided their ratings on the respective questionnaires in the following order: *Social Presence*, CP, SUS, N-TLX, and UEQ.

Once the survey was completed, the test group switched rooms. For participants testing their second application, a second QR code was scanned, directing them to the final rating survey (Rating Recommendation).

CLE PROTOCOL While the CLE was evaluated, participants followed similar instructions and accessed the online survey using the provided QR code. Unlike the ILE setup, no dedicated tutorial scene was included; instead, the system's usage was explained by the instructor. The instructors had been trained in advance on the technical handling of the headsets and the application.

Participants donned their HMDs simultaneously and sat on chairs arranged in a circle. The lecturer in the room also wore an HMD. Once all participants initiated the application, roles were automatically assigned: the lecturer served as the host, and the students acted as clients. After the lecturer interface confirmed the correct number of users, the MR scene was positioned at a marker in the center of the room for all participants. The group then proceeded through the application together.

The session concluded with participants completing the online survey. Participants testing their first application switched rooms, while those testing their second application completed the final assessment (Rating Recommendation) and were dismissed. Each application session, excluding time spent on form completion, lasted approximately 20 to 30 minutes. Overall, the study duration ranged between 1.5 and 2 hours and was conducted over four days.

6.4.3.6 *Data Analysis*

First, the raw study data were preprocessed. SUS and raw N-TLX scores were calculated from the individual items. For the UEQ, only the overall score was determined by averaging across items, rather than analyzing all subscales. Knowledge test data were evaluated as the differences between test scores collected after and before the learning sessions, to assess whether comparable levels of knowledge gain were achieved between groups (GEK-Diff and HEK-Diff).

To evaluate the hypotheses, several statistical tests were conducted. First, the data for each variable were checked for normality using Shapiro-Wilk tests and for homogeneity of variances using Levene's test. Across variables, these assumptions were violated. Therefore, independent sample Yuen's tests, robust t-tests based on trimmed means, were performed to assess differences between groups. Effect sizes were calculated following the δ_t estimate proposed by Algina et al. [11]. Additionally, Bayes factors were calculated to evaluate the support for the null hypothesis of no differences between groups when no significant effects were observed [107]. The interpretation of the Bayes factor results followed the guidelines of Lee and Wagenmakers [152].

6.4.4 *Results*

This section presents the results of the conducted *Experiment 6*. The descriptive results for the ILE are provided in Table 12, while those for the CLE are detailed in Table 13. Additionally, the forest plot in Figure 38 visualizes the mean differences between the individual and collaborative learning environments. An overview of the statistical analyses of the dependent variables is presented in Table 14.

6.4.4.1 *Learning Outcomes*

No statistically significant differences in knowledge gain were observed between the two education methods, MR-based learning and traditional learning (GEK-Diff: $t = 0.36, p = 0.72$, HEK-Diff: $t = 1.66, p = 0.10$). Bayes factor analyses provided moderate evidence supporting the null hypothesis that there are no differences between the groups regarding GEK ($BF_{10} = 0.25$) and anecdotal evidence for the null hypothesis concerning HEK ($BF_{10} = 0.37$). Figure 39 visualizes the pre- and post-learning session test results for both questionnaires.

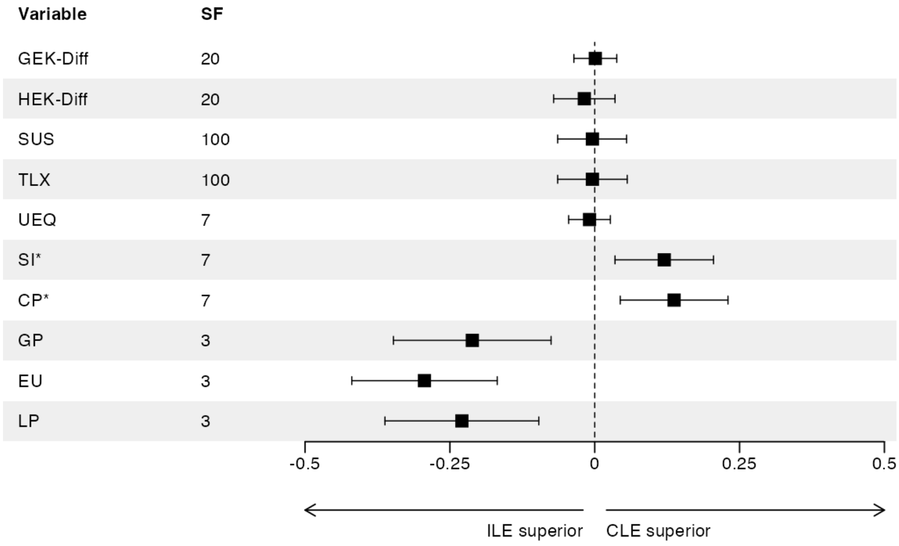


Figure 38: Differences of means and 95% confidence intervals of investigated measures. The final three items represent the mean value and the 95% confidence interval of the three final preference questions. Values were divided by their respective scale maximum (see SF = scaling factor), to achieve values in the range $[-1, 1]$. Asterisks (*) signalize statistical significance. SI = Social Interaction, CP = Co-Presence, GP = General Preference, EU = Ease-of-Use, LP = Learning Preference. Adapted from Schott et al. (2024b) [Core1]. Reused in accordance with ACM's policy for open access articles.

While no increase in GEK scores was observed on average, comparable heart embryology knowledge gain was noted between the MR-based and traditional learning groups. These findings affirm H1, as the MR applications were not inferior to the traditional course regarding learning outcomes, likely due to the same learning content being conveyed.

The differences between the MR-based learning applications, ILE and CLE, were further analyzed in detail. Regarding learning outcomes, only the data from study group A was considered. Again, no statistically significant differences were found between the groups (GEK-Diff: $t = 0.07, p = 0.95$, HEK-Diff: $t = 0.69, p = 0.50$). Bayes factor calculations provided moderate evidence supporting the null hypothesis for GEK-Diff ($BF_{10} = 0.31$) and anecdotal evidence for the null hypothesis for HEK-Diff ($BF_{10} = 0.36$). The findings are illustrated in Figure 40. Pre- and post-exposure questionnaires showed no differences in GEK scores, while similar HEK score increases were observed in both the ILE and CLE sub-groups. These results lead to the rejection of H2. It was hypothesized that varying levels of interactivity and cooperation would result in different learning outcomes; however, no differences in learning effectiveness were detected between the two application types.

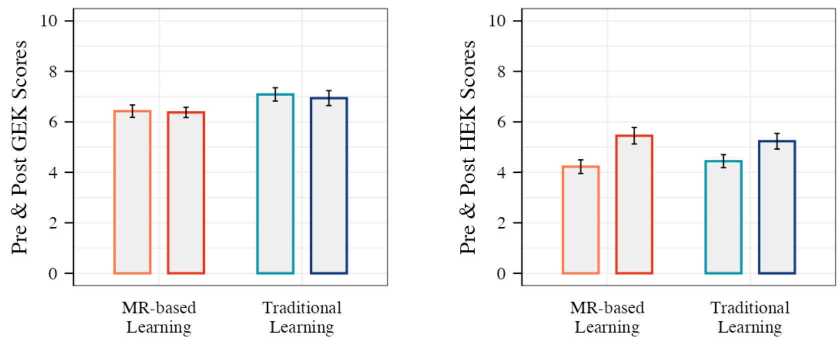


Figure 39: Results of the scores in the Embryology Knowledge Tests (GEK & HEK) in relation to the MR-based learning group (A) and the traditional learning group (B). ■ & ■ represent 'Before Class', while ■ & ■ represent 'After Class'. Adapted from Schott et al. (2024b) [Core1]. Reused in accordance with ACM's policy for open access articles.

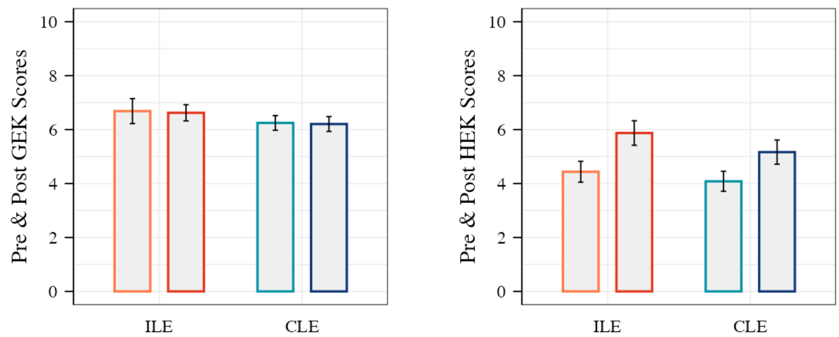


Figure 40: Results of the scores in the Embryology Knowledge Tests (GEK & HEK) in relation to the individual and collaborative learning environments (Only Group A). ■ & ■ represent 'Before Class', while ■ & ■ represent 'After Class'. Adapted from Schott et al. (2024b) [Core1]. Reused in accordance with ACM's policy for open access articles.

6.4.4.2 *Usability, Task Load & UX*

Consistent findings across the *Usability*, *Subjective Workload*, and UX measures indicated that the ILE and CLE conditions provided students with comparable experiences. Yuen's t-tests showed no statistical significance for any variable (SUS: $t = 0.40, p = 0.70$, N-TLX: $t = 0.27, p = 0.79$, UEQ: $t = 0.13, p = 0.90$). Moderate evidence was found supporting the respective null hypotheses that no differences in means exist between the ILE and CLE applications (SUS: $BF_{10} = 0.22$, N-TLX: $BF_{10} = 0.22$, UEQ: $BF_{10} = 0.24$). Consequently, H3 can be affirmed, as no subjective *Usability* differences were identified between the applications. However, H4 and H5 were rejected. It was speculated that the increased interactivity in the ILE would result in higher *Subjective Workload* demands but an enhanced UX. The results, however, indicated that participants rated the CLE equally.

Table 12: Summary of descriptive results for *ILE*. All entries are in the format: mean value [standard deviation]. Adapted from Schott et al. (2024b) [Core1]. Reused in accordance with ACM's policy for open access articles.

Variable	Range	Group A	Group B	Total
GEK-Diff	[-10:10]	-0.06 [1.18]	-0.47 [1.18]	-0.27 [1.18]
HEK-Diff	[-10:10]	1.44 [1.09]	0.94 [1.98]	1.18 [1.61]
SUS	[0:100]	81.18 [16.59]	82.21 [14.60]	81.80 [15.23]
TLX	[0:100]	36.23 [14.05]	27.15 [15.01]	30.74 [15.15]
UEQ	[-3:3]	1.97 [0.58]	2.14 [0.49]	2.07 [0.53]
SI	[-3:3]	4.13 [1.39]	4.49 [1.82]	4.35 [1.65]
CP	[-3:3]	5.06 [1.87]	5.04 [1.81]	5.05 [1.81]

Table 13: Summary of descriptive results for *CLE*. All entries are in the format: mean value [standard deviation]. Adapted from Schott et al. (2024b) [Core1]. Reused in accordance with ACM's policy for open access articles.

Variable	Range	Group A	Group B	Total
GEK-Diff	[-10:10]	-0.04 [1.00]	0.18 [1.91]	0.05 [1.43]
HEK-Diff	[-10:10]	1.08 [2.19]	0.65 [1.80]	0.90 [2.02]
SUS	[0:100]	83.50 [12.89]	78.98 [12.76]	81.38 [12.89]
TLX	[0:100]	28.43 [12.82]	32.58 [13.72]	30.37 [13.27]
UEQ	[-3:3]	2.22 [0.43]	1.77 [0.83]	2.01 [0.68]
SI	[-3:3]	5.35 [1.24]	5.00 [0.91]	5.19 [1.10]
CP	[-3:3]	6.28 [0.88]	5.70 [1.41]	6.01 [1.18]

An inspection of the descriptive data in Table 13 and Table 12 suggested the presence of potential interaction effects on these variables. Students in group B, who had already completed a traditional learning seminar, rated the ILE higher than the CLE. Conversely, students in group A, who had not yet attended the seminar, appeared to favor the CLE application. A post-hoc robust two-way ANOVA confirmed a significant interaction effect on N-TLX related to this observation ($F = 5.38, p = 0.03$).

6.4.4.3 *Co-Presence*

Significant differences between the ILE and CLE applications were identified regarding the two CP measures (SI: $t = 2.41, p = 0.02$, CP: $t = 2.56, p = 0.02$). The respective scores for the CLE were higher than those for the ILE, although the observed effects were small (SI: $\delta_t = 0.39$, CP: $\delta_t = 0.41$). Therefore, H6 can be affirmed. This outcome is likely attributable to the interaction with real human beings in the CLE.

6.4.4.4 *User Preferences*

After completing both MR-based learning sessions and experiencing the ILE and CLE, user preference ratings were collected on a scale

Table 14: Summary of statistical analyses on dependent variables. Significant effects were found for social interaction (SI) and co-presence (CP). The last column interprets evidence against H_0 for these variables. EM = Education Method (MR-based learning and traditional learning), AT = Application Type (ILE/CLE).

Variable	Factor	df	t	p	δ_t	Effect	BF_{10}	Evidence for/against H_0
GEK-Diff	EM	43.35	0.358	0.722	0.06	-	0.251	moderate
GEK-Diff	AT	17.56	0.065	0.949	0.06	-	0.314	moderate
HEK-Diff	EM	39.12	1.664	0.104	0.25	small	0.371	anecdotal
HEK-Diff	AT	20.35	0.690	0.498	0.21	small	0.361	anecdotal
SUS	AT	53.69	0.396	0.694	0.07	-	0.223	moderate
TLX	AT	53.87	0.272	0.786	0.05	-	0.222	moderate
UEQ	AT	53.71	0.127	0.899	0.03	-	0.244	moderate
SI	AT	38.59	2.406	0.021	0.39	small	7.212	moderate
CP	AT	33.90	2.555	0.015	0.41	small	10.382	strong

from -3 (ILE) to 3 (CLE). Across items, the students slightly favored the ILE application over the CLE application (GP: $M = -0.63, SD = 1.98$, EU: $M = -0.88, SD = 1.83$, LP: $M = -0.69, SD = 1.93$) (recap Section 6.4.3.2). However, this tendency was marginal.

6.4.4.5 *Expert Feedback*

The multi-user MR sessions were positively received, particularly for their ability to facilitate interactive and engaging learning experiences. One of the most appreciated features was the option to switch between free rotation and forced perspectives, which enabled targeted questioning and enhanced student engagement. The sessions were considered intuitive once all participants were logged in, and the multi-user format was seen as suitable for introductory lessons, though single-user setups were deemed more practical for individual use.

However, technical challenges were noted, including the need to restart the application after each session, which disrupted workflow. Students often lacked sufficient time to fully explore the animations, and some encountered difficulties understanding certain technical aspects of the HMD.

Compared to traditional teaching methods, the multi-user MR system was seen as a significant improvement in fostering participation and engagement, with its 3D representations providing clearer insights into complex structures. Nonetheless, concerns were raised about maintaining student attention during longer sessions. The quality of student interaction and questioning was also observed to depend heavily on their prior knowledge of the subject.

The system received an overall positive rating, with an average score of 4.2/5 for its ability to improve understanding of embryonic heart development. Key features contributing to its effectiveness included the ability to toggle labels, switch perspectives, and enable collaborative exploration in small groups. However, technical issues such as frequent reconnections, Wi-Fi limitations, and battery constraints hindered the UX. Additionally, imprecise labels and animations were identified as areas needing improvement.

The instructor interface was rated 4.2/5, with feedback praising its intuitive design. Most features were actively utilized by students, although some faced challenges, particularly during independent exploration. While the multi-user setup was engaging, students expressed a preference for single-user environments that allowed them to explore the material at their own pace.

Participants recommended several enhancements to the system, including faster session transitions to reduce disruptions, clearer labeling for each developmental phase, and additional animations for critical processes like septation of the outflow tract. Incorporating congenital malformations as optional learning topics was suggested to encourage deeper exploration and engagement. Expanding the application to include additional embryological stages, related organ systems, and blood flow visualization was also proposed.

While the system demonstrated promise as an innovative teaching tool, challenges in integrating it into the curriculum were noted. These included ensuring accessibility for all students, managing costs, and allocating sufficient time within existing lesson plans. Nevertheless, the system's potential for increasing motivation and enhancing the learning experience was widely acknowledged, particularly when combined with traditional teaching methods.

6.4.5 *Discussion & Conclusion*

The learning effect observed in all groups and both applications indicates that both ILE and CLE were effective supplements to traditional teaching methods in the transfer of knowledge. However, the initial hypothesis (H1), which focused on whether MR-based learning would be inferior to traditional learning in terms of knowledge gain, was examined. The results suggest that MR-based learning was not inferior to traditional methods in terms of overall knowledge acquisition (HEK). No increase in GEK scores was observed in any group, which warrants further investigation. Potential reasons for this could include the complexity of the learning materials, the limited session duration, or the specific content covered in the GEK questionnaire. It is also possible that the knowledge test was not suitable for multiple uses within the study, despite the randomization of the question order and items.

The relatively small increase in HEK scores across all groups could be attributed to various factors, such as the complexity of the topic, the limited session duration, or the type of knowledge assessed by the HEK questionnaire. These factors could be explored in more detail in future research. Based on the absence of significant differences in GEK scores and the comparable increase in HEK scores, it can be affirmed that MR-based learning applications were not demonstrably worse than traditional learning in terms of knowledge acquisition. Fluctuations in CLE completion times were observed due to strict protocol adherence and varying levels of instructor familiarity with the technology. Occasional errors and connection issues resulted, leading to differences in the amount of time participants spent in the application, which may have affected their learning outcomes

COMPARISON TO CONVENTIONAL MEDIA The results are consistent with the literature, which suggests that immersive virtual environments can increase engagement and generative processing but may also heighten distractions and extraneous processing [170]. While these environments have the potential to enhance learning, they have not been consistently shown to be more effective than conventional media. Therefore, it is suggested that immersive technologies be combined with proven instructional strategies to improve learning outcomes. Furthermore, the comparability of learning effects between individual and collaborative settings indicates that no definitive evidence favors one over the other.

In a study conducted by Veer et al. [305], MR was perceived more positively by students in terms of enjoyment and perceived usefulness. However, traditional textbook resources were found to result in higher immediate post-test scores. Both methods were demonstrated to be effective for knowledge retention over a two-week period. Consequently, it is proposed that a combination of MR and conventional teaching methods could optimize learning outcomes in medical education.

NOVELTY EFFECT It was observed that the UEQ scores were similar in both groups. The interactivity was found to differ greatly between the two applications: little to no control over the MR environment was offered to students in the CLE (except for moving and scaling the heart model), while the ILE provided a moldable 3D model, an audio guide, and full control over the steps and phases. Initially, it was assumed that the novelty of the MR experience would overshadow the UX in both groups, as the test subjects had little prior MR experience.

However, it is suspected that familiarity with the seminar setting in group B (ILE) and potentially the novelty effect for both groups (since collaborative learning and individual learning with an audio guide were likely new experiences) could explain the similar UEQ scores. This finding aligns with those of Bork et al. [34], where no significant difference in anatomical knowledge was identified between a seminar group using MR and a traditional learning group. Interestingly, similar to the findings in this study, the technology and its collaborative aspects were appreciated by students in their experiment, though challenges associated with solely using MR for learning were also reported [34].

KNOWLEDGE ACQUISITION Although these MR environments are considered capable of providing engaging learning experiences, mixed results regarding their impact on immediate knowledge acquisition compared to traditional methods have been reported in studies [34, 291]. This finding contrasts with existing literature, which suggests that higher interactivity is more engaging and therefore more effective for learning [68, 92]. The *Subjective Workload* experienced during initial exposure to the topic was analyzed, and the results were found to support this hypothesis. However, other studies have argued that higher *Subjective Workload* may indicate that students are overwhelmed [55, 205, 223].

COGNITIVE LOAD The selected questionnaire for measuring *Subjective Workload* (N-TLX), in hindsight, was not the most suitable tool for differentiating between the *Subjective Workload* associated with the application and that arising from engagement with the learning task. As discussed in Section 5.3.4.4 as a limitation, employing alternative tools, such as a cognitive load questionnaire specifically focused on learning processes, would be more appropriate.

Consequently, an additive *Subjective Workload* effect is suspected, as the CLE group could passively engage with the topic, whereas the ILE group was required to manage high interactivity and a new topic simultaneously. To address this in future studies, it is suggested that a familiarization period with the technology be included before the actual learning sessions, ensuring that participants' responses are based on the educational content rather than the novelty of the medium. The overwhelming effect (novelty bias) discussed by other authors was not observed, as the learning effect remained consistent across both groups [55, 162, 176].

SOCIAL PRESENCE The audio guide integrated into the ILE was used to welcome students, introduce them to the MR environment and interactions, explain all developmental steps in detail, and provide feedback upon successful task completion. However, it did not replicate the advantages of having a real instructor present to address individual situations and questions from the group, as reflected in the results regarding SI and CP. Additionally, the questionnaire results may have been influenced by the presence of up to two other students in the room during the evaluation of the ILE. Studies indicate that MR is more effective for collaborative and contextual learning, as it allows interaction with both real and virtual elements, enhancing spatial understanding and contextual learning. In contrast, fully immersive experiences have been shown to increase engagement and emotional response [12]. Further research is recommended to definitively determine the effectiveness of combined approaches and to identify the optimal teaching methods for various learning objectives and student groups.

6.5 CHAPTER SUMMARY & ONGOING WORK

This chapter demonstrated the iterative HCD process involved in the development of an MR-based CLE. Initially, a new medium, along with potential interaction and visualization formats, was explored to adapt the concepts developed in Chapter 5. This exploration was conducted through the participatory involvement of experts in anatomical education (*Experiment 4*). Feedback from this phase was incorporated, and the application was further developed both technically and conceptually. It was then evaluated in a simulated environment modeled after an actual anatomy seminar (*Experiment 5*). Finally, the CLE and the ILE from Chapter 5 were transferred to a unified technological platform and compared in a study involving both medical students and experts.

The MR environments evaluated were shown to enhance learning experiences and outcomes in traditional teaching settings. Thus, MR can serve as a valuable supplement to conventional teaching methods. Participants in our study generally favored the ILE due to its interactive and self-directed learning features, although user preferences were only slightly higher than those for the multi-user approach. The study has shown the importance of striking a balance between interactivity and *Usability*. While the single-user application provided a higher level of user control and interaction, the multi-user application facilitated greater SI and CP, which are crucial for collaborative learning. This balance is essential for the development of effective MR learning tools that cater to different learning needs and preferences.

Expert feedback highlighted the system's ability to foster engagement and participation, particularly in multi-user settings. The collaborative MR approach was praised for its interactive features, such as the ability to switch perspectives and toggle labels, which enhanced understanding of complex structures like the embryonic heart. However, technical challenges, including session interruptions and connectivity issues, were noted as areas needing improvement. Experts also suggested expanding the content to include additional animations and congenital malformations to deepen engagement. While logistical challenges such as cost and integration into existing lesson plans remain, the system was widely acknowledged as a promising supplement to traditional teaching methods.

Combined teaching methods incorporating traditional seminars and MR environments could potentially optimize learning outcomes in medical education. The insights gained from this iterative process provide a strong foundation for further development and integration of MR-based tools in curricula.

The success of this application provided an opportunity for further development. Plans are underway for students to utilize the ILE as a permanent supplementary learning tool for self-study. Additionally, work is in progress to develop a unified platform for the CLE and ILE, which is intended to be made available to the Institute of Anatomy at our university in upcoming semesters. For example, clinical information could be integrated, allowing the system to display concurrent cardiac defects along a timeline, thereby demonstrating clinical relevance for cardiology trainees. While the current VLE focuses on heart development, it is planned to be adapted to other complex processes, such as the formation of the inguinal canal or the pharyngeal arches.

This work lays the groundwork for understanding the effects of MR-based learning, yet critical design features remain to be explored. For instance, the role of realism and the cognitive workload in fully immersive VLEs are still open questions. Additionally, strategies to enhance SIs in traditional seminar settings within MR, and their connection to CLEs across various virtual and physical spaces, require further investigation.

The next chapter extends the CLE and ILE concepts, expanding them across multiple MR platforms. It shifts focus to a different organ system—the liver—and targets a more advanced stage of medical education, specifically addressing students engaged in anatomical training for liver surgery. In contrast to the abstract 3D models of the embryonic heart, this system leverages curated clinical data to create 3D reconstructions, offering a more tangible and clinically relevant learning experience.

CROSS-MODALITY PLATFORM FOR LIVER ANATOMY EDUCATION

ch. 7

SYNOPSIS *This chapter presents the development of a collaborative and cross-modal learning platform designed as a tool for liver anatomy education. With support from surgeons, curated clinical use cases were implemented into diverse MR approaches, and various teaching and learning scenarios were explored.*

MY CONTRIBUTION *I was responsible for the conceptual refinement and oversaw the implementation of the methodological approach. This included developing the theoretical foundations and conducting an extensive literature review. Additionally, I contributed to the requirements analysis, which involved expert interviews and technical testing for the implementation. I designed the study, while data collection was carried out collaboratively. I took primary responsibility for drafting, reviewing, and editing the manuscript of the original paper. Furthermore, I supervised and actively contributed to the interface design and the creation of visual elements, including new photographs, tables, illustrations, and additional content specifically produced for this thesis.*

7.1 INTRODUCTION

The use of collaborative MR systems as a suitable medium to support knowledge transfer in the early phases of anatomy education, with a generally positive influence on user engagement and learning effectiveness, has already been demonstrated in the course of this thesis.

In the previous chapter, the development from an ILE to a CLE was explored, highlighting its practical benefits for foundational medical education. However, to gain a comprehensive understanding of the potential of MR technology, it is essential to broaden the scope and examine other areas of the educational spectrum. While the focus so far has been on abstract animated visualizations to illustrate developmental processes, the use of real clinical cases has not yet been addressed. Such cases can provide an effective means of conveying applied knowledge about medical conditions and their treatments, thereby better preparing students for clinical practice.

Furthermore, the systems presented so far have been designed with a clear distinction between student-driven self-exploration and teacher-guided learning. This raises the question of how students and instructors can efficiently share a virtual space to foster collaborative learning while maintaining educational effectiveness. Hence the following research question:

RQ7 | What design principles can enhance collaborative MR environments for advanced medical training, specifically in liver anatomy education, by integrating real clinical cases and accommodating varying levels of immersion?

To address this question, this chapter introduces an application aimed at preparing students in advanced clinical semesters for liver surgery training. The system combines multiple modalities, enabling participation in a shared virtual space at varying levels of immersion. This approach not only bridges technological gaps but also supports barrier-free access to MR systems, fostering inclusivity within educational contexts.

Unlike previous concepts designed exclusively as either ILEs or CLEs, this chapter introduces a Cross-Modality Virtual Learning Environment (CMVLE) that facilitates both independent self-exploration by students and teacher-guided learning within a shared virtual space. By supporting diverse modalities, this approach represents a more open-ended methodology than those discussed earlier in this thesis.

Liver surgery represents a highly complex subfield of surgery that exemplifies the challenges of teaching intricate clinical practices. These surgeries often involve managing the liver's dual vascular systems, requiring careful consideration of resection impacts on the portal vein and hepatic artery. Additional complexities include determining the appropriate surgical access to a pathology and deciding the precise amount of tissue to remove. Traditional teaching modalities—including didactic lectures, laboratory practice, and textbook learning, as discussed in Chapter 3—often fall short in comprehensively conveying such multifaceted surgical decision-making processes. By addressing these challenges, MR offers significant potential for teaching complex spatial relationships, such as the intertwined vascular structures critical in liver surgery [47, 132, 267, 282, 338].

This chapter introduces a MR-based CMVLE specifically designed for liver anatomy education. The requirements for this application were developed participatively with clinical partners to integrate HCD principles, ensuring it meets the specific needs of both clinicians and learners. The prototype allows students to explore curated clinical cases, featuring 3D surface models, 2D image data, volumetric reconstructions, and detailed medical information, all within a shared virtual space. These cases simulate real-world scenarios and foster problem-based learning through collaborative interaction with the data. Focused on frequently occurring malignant tumors, the cases include tools for dynamic sorting and deeper exploration of relevant medical context. By supporting diverse modalities—from HMD-based solutions to desktop PCs—the system aims to offer an inclusive and accessible platform for advanced medical education.

A video demonstration of the application is linked in the Appendix (see Section 9.1).

7.2 EXPERIMENT 7

This section presents the final experiment of this thesis. It details the development and evaluation of a CMVLE, exploring its potential use in enhancing students' preparedness for liver surgery training.

7.2.1 *Material*

Nineteen data sets of patients who underwent surgery for liver tumors were selected. The selection encompasses the most common malignant liver tumors, including hepatocellular carcinoma, cholangiocarcinoma, and colorectal liver metastases. A broad range of resection types is represented in the case collection, ranging from small atypical resections or single-segment resections to major surgeries such as extended hemihepatectomies. The data sets were derived from 3D images generated through CT scans. The required segmentation was performed by a liver surgeon at the University Medical Center of the Johannes Gutenberg-University Mainz (Germany), experienced in reconstruction using Synapse 3D[1]. The resulting 3D STL files were converted into the OBJ file format and subsequently imported into the Unity game engine (*Unity Technologies*).

Digital Imaging and Communications in Medicine (DICOM), a standard for the storage, management, and communication of medical image information, is used in this work to refer to all 2D and 3D image data, including associated clinical patient information.

7.2.2 *Development of the CMVLE*

A participatory design process was employed for requirements elicitation. The prototype was iteratively developed in collaboration with experienced liver surgery lecturers from the Department of General, Visceral, and Transplant Surgery at the University Medical Center of the Johannes Gutenberg-University Mainz, Germany. A cross-modality MR-based learning platform is proposed as an exploration and learning environment for liver surgery education. Various interaction possibilities were developed, which are detailed in Section 7.2.2.5. Three core functionalities were integrated for data exploration: an interactive *Liver Shelf* (see Figure 41), an *Information Board* (see Figure 41), and a DICOM workstation, comprising a *DICOM Board* and a *DICOM Cube* (see Figure 41 and Figure 43). The environment itself deviates from the lecture hall-inspired settings of previous sys-

1 FUJIFILM Europe GmbH, https://www.fujifilm.com/uk/en/healthcare/synapse

Table 15: Sorting types and corresponding parameters. Reprinted from Schott et al. (2021) [Core7], copyright 2021 IEEE. Used with permission.

Sorting type	**Sorting parameters**
Resection types	Extended hemihepatectomy left, Hemihepatectomy (left/right), Left lateral resection, Atypical (simple/complex), Mesohepatectomy, in situ split
Vascular Reconstruction	None, Cava, Hepatic Vene, Portal Vene
Intervention	Primary, Recurrence
Tumor type	Hepatocellular Carcinoma, Cholangiocellular Carcinoma, Metastasen Mamma-CA, Colrectal Liver Metastases, Mucinous cystic neoplasia, Metastases Gastrointestinal Stromal Tumor, Focal Nodular Hyperplasia, Echinococcus multilocularis, Gall Bladder CA
Vessel Variation	No, Yes
Resectability	Resectable, limit value

tems, instead adopting a low VF design to deliberately focus attention on the clinical cases.

7.2.2.1 *Learning Objectives*

The CMVLE serves as an entry point to liver surgery education by presenting theoretical content on liver resection planning through a problem-based learning approach using clinical cases [109]. The focus is on fostering symbolic knowledge [213]. The complex 3D structures of the liver are presented with a high degree of clarity to facilitate the memorization of (anatomical) learning objects. The *fully immersive MR mode* is intentionally designed as a safe, closed, and controlled learning environment. Ambient noise and fast animations were deliberately excluded to minimize distractions.

The prototype is intended as an experimental environment in which users can independently explore medical data. Collaborative learning is enabled by allowing students to acquire intricate knowledge through self-exploration in groups. Active participation in MR and passive observation during, for instance, teacher-led sessions, are both supported. The environment complements existing materials and methods for liver surgery education by vividly presenting real patient cases, enabling interactive discussions, and providing students with easy access to data.

7.2.2.2 *Liver Shelf*

The overview visualization was inspired by the metaphor of library shelves, offering a structured and familiar spatial layout for anatomical data. Preim et al. [212] describe this kind of shelf-based representation as a promising metaphor for organizing complexity in medical

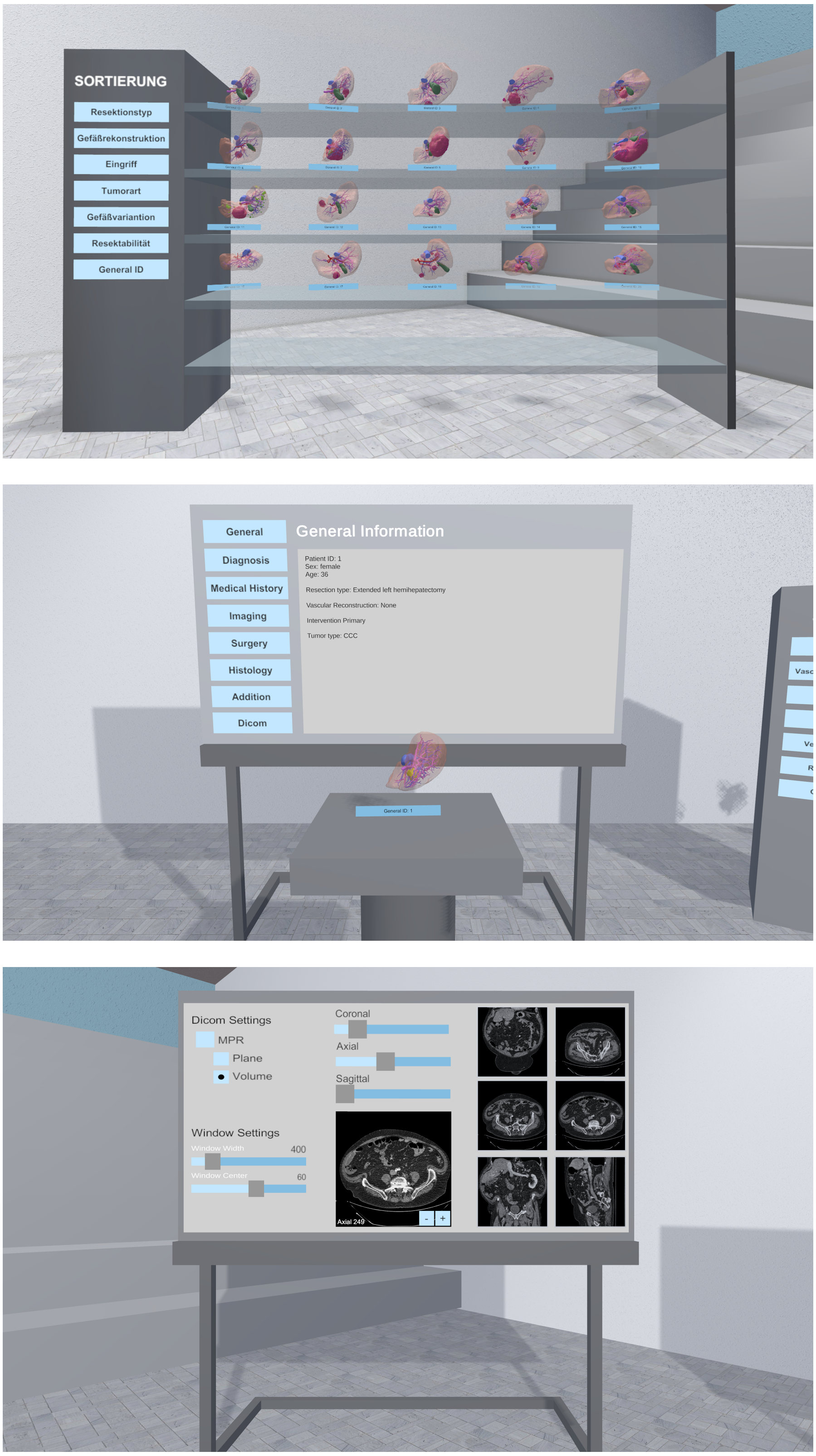


Figure 41: Top: Interactive *Liver Shelf* with 19 medical 3D data sets. Middle: *Information Board* with patient and case information. Bottom: 2D *DICOM board* with CT data from different slice images.

learning environments. Various 3D liver models are arranged in multiple compartments stacked vertically. These models include the liver surface, blood vessels, gall bladder, and different types of tumors or cysts. Using the controller-based *Virtual Hand* technique, the 3D models can be grabbed, translated, rotated, and scaled through bi-manual interaction. The liver surface is displayed transparently to reveal the internal structures (see Figure 42).

The coloring of the structures follows conventions commonly used in medical textbooks (e.g., tumors in yellow, gall bladder in green, vena cava in blue, arteries in red, and hepatic veins in blue). Figure 41 illustrates the *Liver Shelf* in a VR representation, with corresponding 3D models. The shelf can be extended with additional data sets, making it possible to create multiple shelves or even an entire library.

Additional functionalities, not achievable in a traditional library, have also been integrated. For instance, the data sets can be automatically sorted based on various criteria (see Table 15).

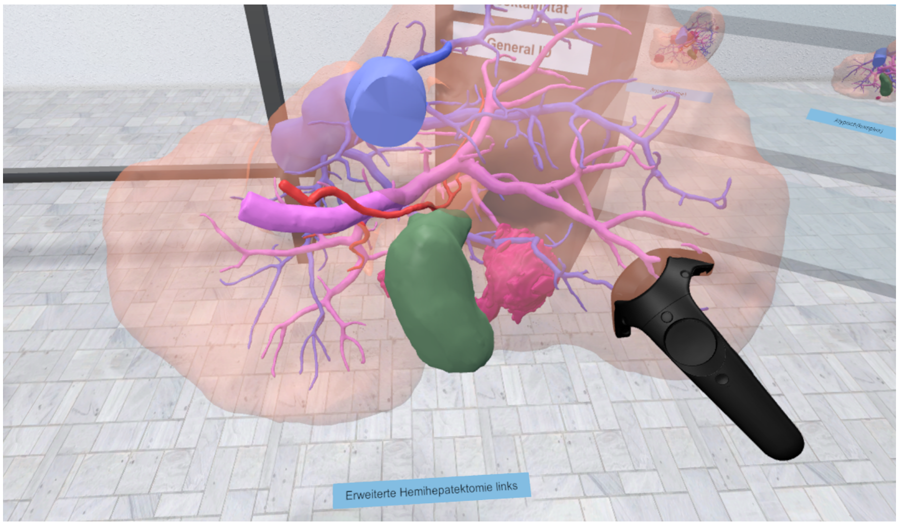


Figure 42: User interacting with a virtual organ using the controller.

7.2.2.3 *Information Board*

Detailed and anonymous information is displayed on the *Information Board*. This includes age, sex, diagnosis, medical history, imaging, surgical history, histology, and various 2D image data, all derived from treatment notes and medical reports. To explore the details of a specific liver data set, users can teleport or walk to the *Information Board*. Activation requires placing a selected 3D data set on the platform in front of the board (see Figure 41). Once activated, the meta-information is organized into different categories, which can be accessed by selecting buttons using ray-based interaction. This interaction is initiated by activating a ray originating from the user's controller, with selections confirmed by pressing a button on the controller.

The ray can also function as a pointer to highlight specific text passages. To access information from another data set, the 3D surface model on the platform can be swapped.

7.2.2.4 *DICOM Board*

DICOM data sets can be selected in the prototype via the *Information Board* and displayed on the *DICOM Board* (see Figure 41). The *DICOM Board* includes a 2D image viewer and the option to activate a multiplanar reformation (Multiplanar Reconstruction (MPR)) view using the *DICOM Cube* (see Figure 43). Basic tools have been implemented to allow users to interact with the data sets through traditional slicing.

These tools enable adjustments to the range of gray values displayed and allow changes to the slice direction across the three anatomical planes: sagittal (left and right), coronal (back and front), and axial (head and tail). Interaction is facilitated using a slider and the ray-based interaction method described earlier.

In addition to traditional slicing, the MPR functionality enables interaction with data sets via the *DICOM Cube*. Users can freely position a plane within the 3D data set to view the resulting slice (see Figure 43). This allows for the selection of planes where the target anatomical structures can be assessed effectively. The *DICOM Cube* itself can also be moved freely. For additional orientation, the data set's hull is outlined with lines, and small cubes in the corners indicate the orientation of the data set, labeled with the anatomical alignments (sagittal, coronal, axial)..

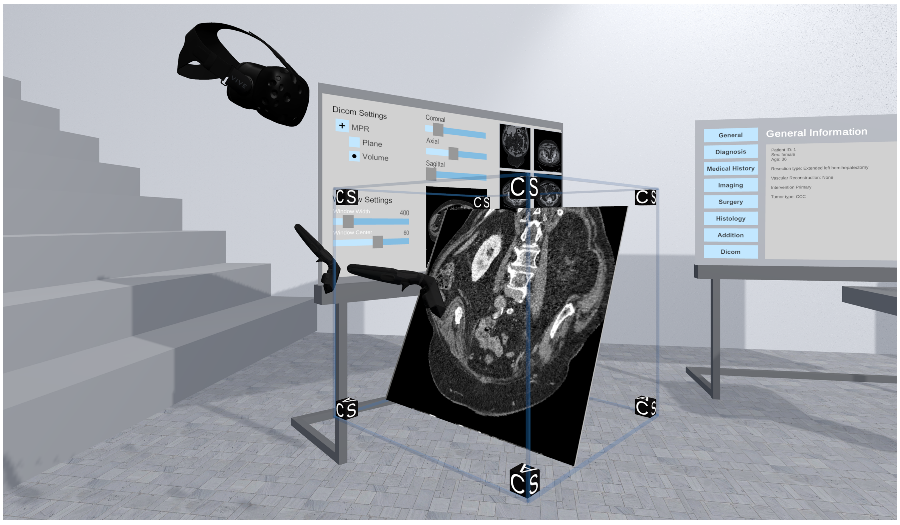


Figure 43: 3D *DICOM Cube* with an interactive plane in front of the 2D *DICOM Board* displaying CT data from different slice images. Reprinted from Schott et al. (2021) [Core7], copyright 2021 IEEE. Used with permission.

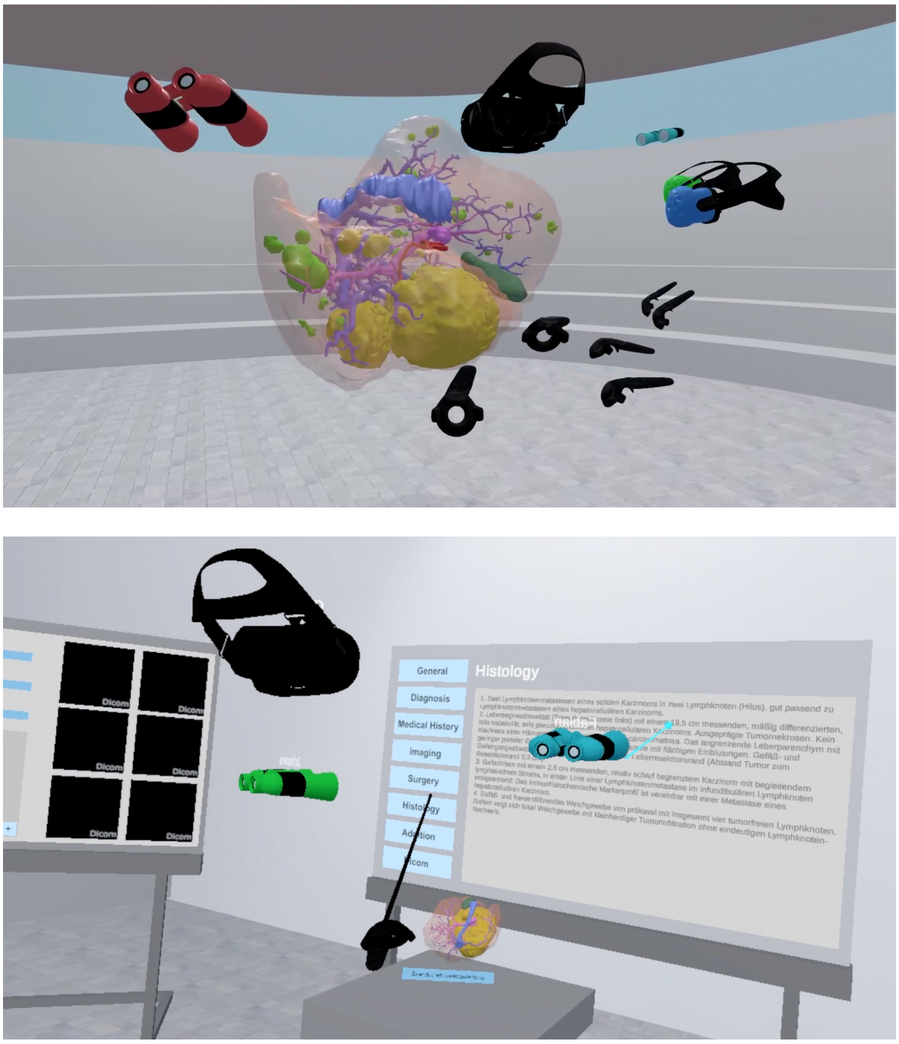


Figure 44: Multi-user scene view with three users in *fully immersive MR mode* (represented by Vive headsets + controller) and multiple *Spectators* (represented by binoculars).

7.2.2.5 *Modalities*

The CMVLE offers multiple modes of use, including a fully immersive MR experience, a *VST MR mode*, and a *Spectator Mode*. The *Spectator Mode* enables users to join the VR scene using a desktop PC.

Additionally, the prototype supports multi-user functionality, allowing several users to interact within the same environment simultaneously. All modes share a consistent interaction design and environment structure, with minor differences that are outlined below.

FULLY-IMMERSIVE MODE A simplified lecture hall environment was created in addition to the interaction elements described earlier. The minimalist design aims to minimize distractions and reduce cognitive load. While real-world movement is possible, it is constrained by hardware limitations (e.g., cable length and tracking space). To address this, users can teleport through the environment using their controllers. Range-limited teleportation was deliberately implemented to enhance the sense of security, preventing users from walking through walls or altering essential objects in the VLE. A blend between the two locations is designed to further reduce cognitive load and prevent user disorientation [320].

Unlike the *VST MR mode*, the *fully immersive MR mode* allows for environments of any size. However, a direct representation of other users is missing in this scenario. To address this, other MR participants are represented by basic avatars, including an HMD and two controllers. Each user is automatically assigned a unique headset color upon joining to facilitate identification. Additionally, users can display a name of their choice above their headsets (see Figure 44).

Head and hand movements are transmitted in real-time, enabling actions such as nodding, shaking heads, or waving to be recognized by other participants. A ray originating from the user's controller serves as a pointer and matches the color of their HMD. Figure 44 illustrates multiple users in the CMVLE, collaboratively examining a scaled-up 3D liver model in *Spectator* and *fully immersive MR mode*.

SPECTATOR MODE If a *fully immersive MR mode* is selected but no HMD is connected, the prototype runs as a desktop application (see Figure 46). This enables users to participate without requiring MR hardware. In this mode, desktop users are represented as binoculars (see Figure 44), which are also color-coded and can display a custom name. Unlike HMD users, desktop participants are passive and operate in a so-called *Spectator Mode*, without the ability to actively manipulate objects.

Interaction in the *Spectator Mode* is facilitated through mouse and keyboard input. Movement is controlled using the arrow keys or alternatively the WASD keys, while the space bar activates a pointer ray. The mouse is used to change the orientation, and the user's vertical position can be adjusted with the mouse wheel. Additionally, the *Spectator Mode* allows users to view the perspective of a fully immersive MR user, enabling them to see exactly what the MR user is experiencing. This feature is particularly beneficial for lecturers, who can supervise a group of students and provide immediate feedback.

It should be noted that participation in the *Spectator Mode* is not supported in the *VST MR mode*.

VIDEO-SEE-THROUGH MODE In preliminary consultations with medical experts, it was noted that complete isolation from the real world could cause discomfort for lecturers. Concerns included potential collisions within the physical space and unease stemming from the lack of awareness of students' physical presence. To address these issues, a VST MR system was developed. Unlike the *fully immersive MR* *mode*, the VST environment removes virtual elements such as the lecture hall and reduces the area where objects are positioned. This adjustment is necessary because free teleportation is not feasible in VST MR, and object placement is constrained by real-world factors such as physical space, headset cable length, and tracking area.

Figure 46 illustrates a user's perspective in the VST MR environment, showing a 3D liver model in front of the *DICOM Board*. In contrast to the *fully immersive MR mode*, real-world structures such as walls and tables are visible. Interaction with the stations (via controller) and the user interface is identical to that in the *fully immersive MR mode*.

7.2.2.6 *System Architecture & Technical Details*

An overview of the system architecture, combined with the VR/AR modes described in Section 7.2.2.5, is shown in Figure 45.

The application was implemented using Unity 2019.1.2f1[2]. To facilitate data exchange—including the position and rotation of objects or users—the prototype employed a network connection powered by the Photon Unity Networking 2 (PUN2) package[3]. Several Vive HMDs[4] were used to run the fully immersive MR scenario; however, the use of this specific headset was not mandatory. The headsets were operated via Steam VR[5], and user interaction was facilitated through the Vive controllers.

For the implementation of VST MRs, the same HMDs were used in combination with the ZED Mini camera[6], enabling the realization of VSTs MRs within a VR headset. The Microsoft HoloLens[7]—despite its widespread use in research—did not meet the requirements regarding detail level and model complexity. To synchronize the coordinate system for the VST MR setups, the "lighthouse settings" and "chaperone files" generated during the Steam VR room setup were utilized.

In a fully immersive MR environment, participants can move freely through space using teleportation, and sharing a real-world registration among users is not mandatory. However, in the VST MRs setups, all participants must operate within the same real-world coordinate system. While separate registrations are typically performed for each system, it is essential in this case for all systems to share the same room registration.

To facilitate this, a small tool was developed to automatically transfer and replace the required registration files from one fully immersive MR setup to all others via the network. These files primarily include the "lighthouse settings" and "chaperone files" generated dur-

2 Unity Technologies, `https://unity.com`
3 Exit Games GmbH, `https://www.photonengine.com/pun`
4 HTC Corporation, `https://www.vive.com`
5 Valve Corporation, `https://store.steampowered.com/steamvr`
6 Stereolabs Inc., `https://www.stereolabs.com/zed-mini/`
7 Microsoft Corporation, `https://www.microsoft.com/en-us/hololens`

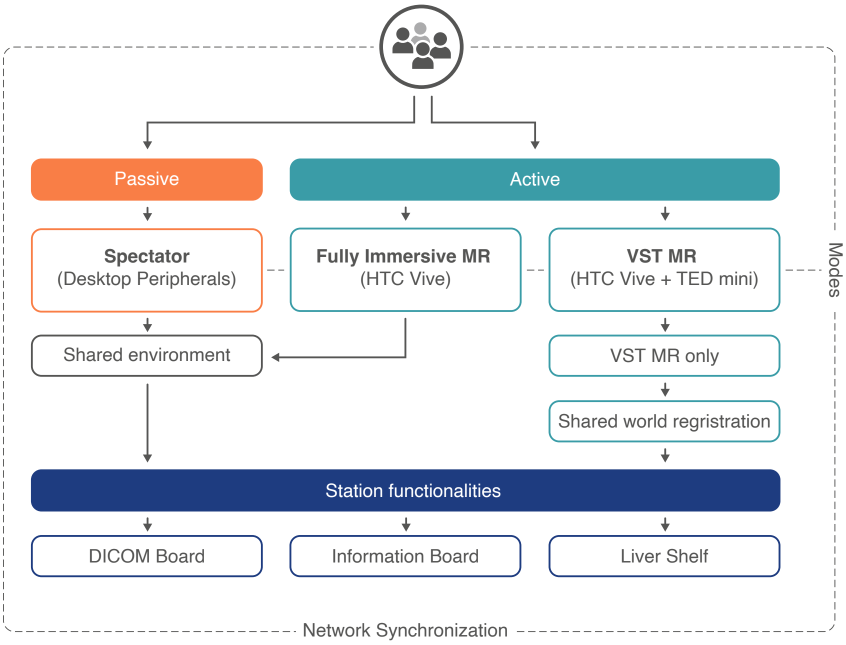


Figure 45: Overview of the system architecture. Adapted from Schott et al. (2021) [Core7], copyright 2021 IEEE. Used with permission.

ing the Steam VR room setup, which are stored in the Steam VR configuration folder.

For the *Spectator Mode*, only standard desktop peripherals (e.g., keyboard, mouse, monitor) are required.

7.2.3 *Evaluation*

An explorative study format was selected, focusing on the collection of qualitative feedback. The insights gathered from potential users were intended to refine and better define usage scenarios for the different modes of the prototype. Particular attention was given to individual aspects related to the quality, quantity, and presentation of the medical data, as well as the *Usability* of the interactions.

The study also emphasized evaluating the suitability of the application for multi-user scenarios, with a specific focus on the *fully immersive MR mode* and the corresponding desktop application.

7.2.3.1 *Study Design*

The evaluation procedure was based on the TAP method, where participants were asked to continuously verbalize their thoughts while interacting with the prototype. Participants were free to move throughout the environment and were encouraged to interact with it in an exploratory manner. Only the fully immersive MR prototype was additionally evaluated using the SUS and IPQ scales. A concluding in-

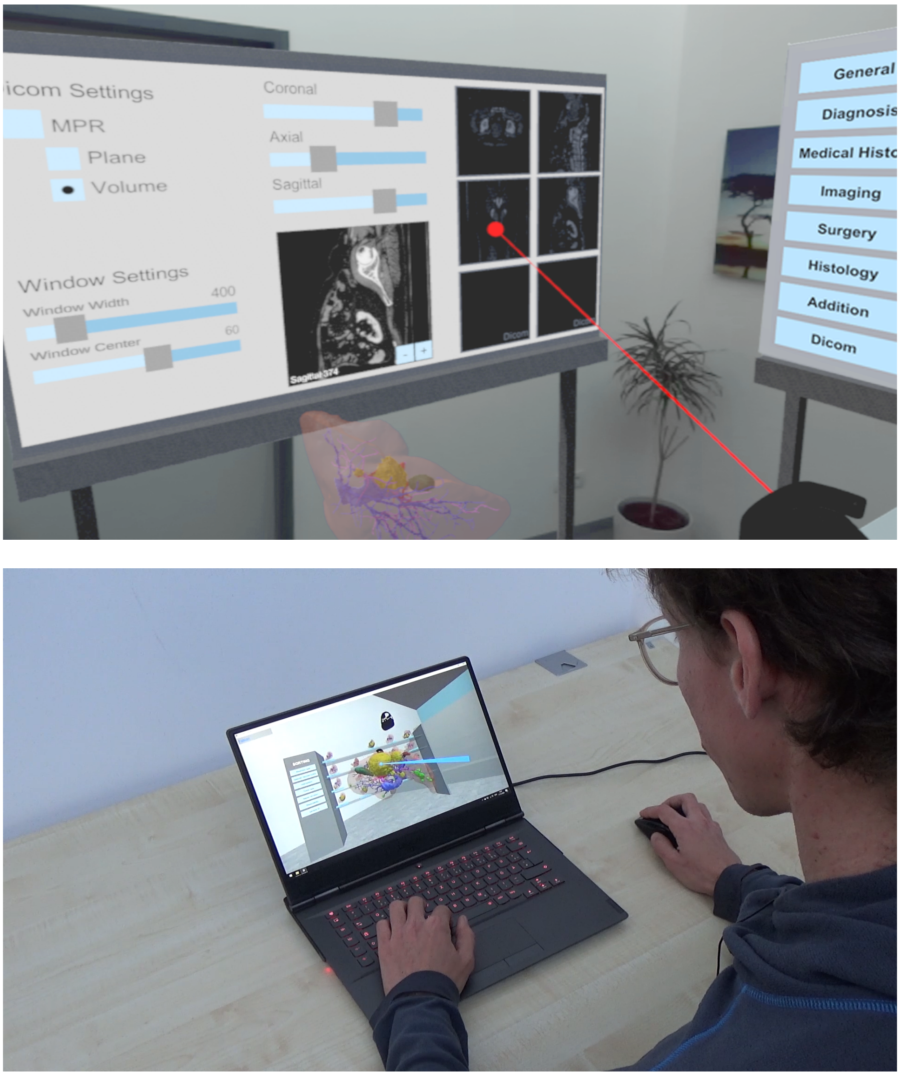


Figure 46: Top: Perspective of a user in the VST MR mode examining a 3D model of a liver and the *DICOM Board*. Reprinted from Schott et al. (2021) [Core7], copyright 2021 IEEE. Used with permission. Bottom: User in front of a PC in *Spectator Mode*.

terview was conducted using a semi-structured questionnaire, with audio recordings made for later analysis.

7.2.3.2 *Participants*

The different modes were demonstrated to a test group of ten participants, aged between 23 and 34 (M = 27.4), all of whom had a medical background. To gain independent insights into the needs of students and experts, the test group was divided into two subgroups:

The first subgroup consisted of five non-paid experienced surgeons (three male, two female) from the University Medical Center of the Johannes Gutenberg-University Mainz, Germany. All but one participant in this subgroup had teaching experience, including tutorials for courses in ophthalmology, surgical suturing, ultrasound, and anatomy. Two participants had more than five years of professional experience as medical doctors. All participants in this subgroup had prior experience with fully immersive MR, while two indicated that they had no experience with VST MR. No participants in this study were involved in the development of the system.

The second subgroup comprised five medical students (three male, one female, one not specified) from the same university, who were compensated with 20 Euros for their participation. One participant had experience as a teaching assistant, and four participants had completed at least their fifth academic semester. All participants in this subgroup had prior experience with MR applications, though two reported having no experience with VST MR.

7.2.3.3 *Setup*

Both test groups completed the experiment in an MR lab within a 3×3 meter tracking space. The study for instructors (medical experts) was conducted via video call. At the participant's location, a technical assistant was present in the same room as the test subject to manage the technical equipment and ensure the participant's well-being. Instructions were provided exclusively by the investigator through a loudspeaker. During the experiment with the medical students, the investigator was physically present in the same room as the test subject.

In both scenarios, the investigator was supported by an assistant located in a separate room, who joined the session as a fully immersive MR user. The investigator conducted the interview and operated in *Spectator Mode* during the MR application.

To test the system's load capacity, a technical test was conducted involving three participants in *fully immersive MR mode* (wearing HMDs) and 15 participants in *Spectator Mode*. No latency issues were observed during the 20-minute test. Due to the limited tracking space and compliance with COVID-19 hygiene regulations, the *VST MR mode* was not tested with multiple users.

7.2.3.4 *Procedure*

The study lasted approximately one hour and included a step-by-step demonstration of the different modes: first, the *fully immersive MR mode*; second, the *Spectator Mode*; and third, the *VST MR mode*. After explaining the process to the participants and obtaining written informed consent along with demographic data, the experiment began in the virtual environment, starting with a demonstration of user-to-user interactions, such as waving hands and shaking heads.

Once participants indicated they were ready, the three stations were visited sequentially, beginning with the *Liver Shelf*. Participants were regularly encouraged to verbalize their thoughts, observations, expectations, and the reasoning behind their actions.

During the first station, the focus was on interaction with the 3D representation. Participants were asked to explore the interaction techniques as independently as possible. If they encountered difficulties, the study assistant demonstrated the techniques. Different 3D models were passed between participants, and structures were examined collaboratively. The sorting functions listed in Table 15 were then presented, and participants were required to change the sorting of the *Liver Shelf* at least once.

Next, participants selected a 3D representation and placed it on the platform of the *Information Board* (second station). There, they explored the medical background information associated with the chosen data set. At the third station, the *DICOM Board* was introduced. After a brief explanation, participants engaged in a free interaction period where they sliced through a dataset and created photocopies. Each participant was required to select at least one random slice, create a copy, and hand it over to the assistant. They were then instructed to test the *DICOM Cube* by using the slice plane at the final station. At the conclusion of the session, participants were asked if they felt unwell during their time in the *fully immersive MR mode*.

After completing the main routine, participants exited the *fully immersive MR mode* and re-entered as *Spectators*. Once they became familiar with the controls using the keyboard, they were given the opportunity to explore the virtual environment. Inside, the study assistant, still in *fully immersive MR mode*, interacted with the environment.

During the setup for the *VST MR mode*, participants completed questionnaires on *Usability* and *Presence* for the *fully immersive MR mode*. Due to hygiene and travel restrictions, it was not possible to demonstrate the *VST MR mode* in a multi-user setup. Instead, this mode was presented via a live video stream from the perspective of the study assistant. The assistant performed interactions within the *VST MR mode* while the participants observed and provided comments.

7.2.3.5 *Data analysis*

Upon completion of the study, the recorded individual statements were compiled into a table. In the first step, the statements were labeled to associate them with specific stations and participants. Subsequently, the statements were categorized into themes, including input methods and devices, visual processing, and contextual awareness. Overlaps were then identified, and clusters were formed containing at least two similar statements. Finally, a summarizing statement was created for each cluster.

7.2.4 *Results*

A total of 435 individual statements were recorded. Forty-nine of these statements were summarized and assigned to ten categories, as presented in Table 16. The categories include feedback on the virtual organ model or the *DICOM Cube* ("3D Representation"), the placement of these models ("Spatial Arrangement"), and direct interaction with models and the cube ("3D Interaction"). Interactions explicitly related to the graphical (UI) were classified under "2D Interfaces."

Additional categories addressed statements about multi-user application scenarios ("Multi-user"), the input device used ("Input Device"), locomotion within the virtual environment ("Locomotion"), and the perception of the virtual world ("Virtual Environment"). One participant reported feeling slightly uncomfortable in the virtual environment during the initial phase of the experiment. All other participants reported no discomfort. It was not possible to demonstrate the *VST MR mode* to two participants.

The *fully immersive MR mode* achieved a *Usability* score of 79 (maximum = 100; SD = 7.8) on the SUS. According to the results of the IPQ, spatial presence received the highest rating (M = 4.8; SD = 0.5), while experienced realism was rated the lowest (M = 3.08; SD = 0.87). No significant differences were identified between experts and students regarding *Usability* or *Presence* aspects. A visualization of these results can be found in Figure 47.

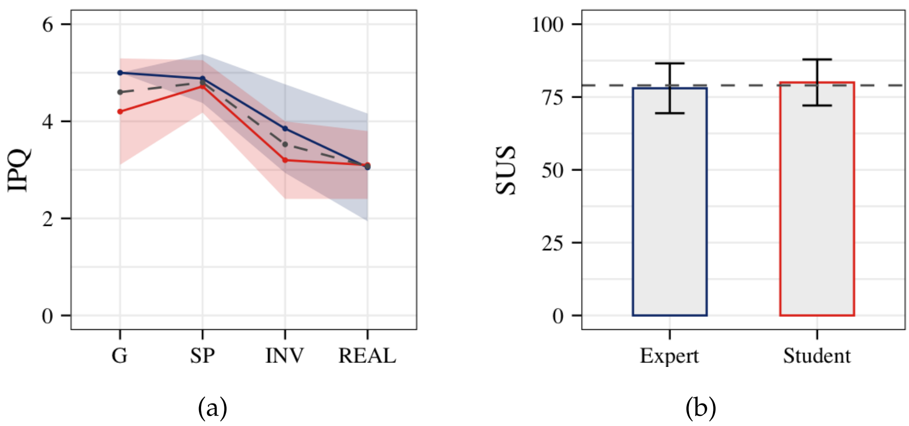


Figure 47: IPQ score (a) and SUS score (b) with respect to participants' background. ■ represents Expert background, while ■ represents Student background. Shaded error bands in (a) represent the standard deviation. ■ dashed lines indicate Mean Values for each Total Score. Adapted from Schott et al. (2021) [Core7], copyright 2021 IEEE. Used with permission.

Table 16: Summary of the collected statements of the respective stations under allocation of different categories. The identifier (ID) represents the respective test person and serves for contextualization. ID = 1-5 experts; ID = 6-10 students. Reprinted from Schott et al. (2021) [Core7], copyright 2021 IEEE. Used with permission.

Category	Statements	ID
Liver Shelf		
3D Representation	Color scheme of the structures is appropriate	1,2,7,8,9,10
	More detailed exploration of individual pathologies desired	2,9,10
	Pathologies are clearly visible and adequately presented	1,2,6,7,9
	When enlarging the model more information should appear	7,8
	When rotating the model the context to the position in the body is missing	8,10
Spatial Arrangement	3D models were placed too low	1,5,7
	Arrangement of 3D representation is suitable	5,8
	Hierarchy of 3D models is not obvious	7,9,10
3D Interaction	Direct 3D interaction (scaling, translation, rotation) with Organs feels natural	1-7,9,10
	Exploration methods are easy to understand	2,4,5,9
	Mix of direct and ray interaction leads to confusion	6,7,9
	Object removal is expected with ray instead of gripping it directly	3,5,7,10
	Possibility to hide specific structures, change transparency and brightness	1,2,3
	Ray should hit internal structures	5,6,7,9
	Unused models should automatically sort themselves into shelves	1,3,5,8
2D Interface	Extension of the sorting function by adding more sub parameters	1,3
	Labeling is poorly readable	4,5
	Sorting option of patient ID is not helpful	7,8
	Sort function is useful	1,5,6
Information Board		
3D Interaction	Individual structures should be selectable and point to information board details	4,8,9
2D Interface	Scope and presentation of the information appropriate	3,6,8
	Text is too small and contrast is too weak	1,4,5
Multi-user	Presence of several information boards for parallel interaction and exploration	7,8
DICOM Board		
3D Representation	DICOM Cube needs further orientation hints	3,4,6,8
	Registration between 3D model and DICOM Cube	2,4
3D Interaction	Board functionalities should be available on preview image and photocopies	3,5,8
	DICOM Cube is a helpful addition because it promotes spatial understanding	2,4,6,7,8
	DICOM Cube and plane should be scalable, because visible areas are wasted	2,3,4
	Interactive photocopies are a useful addition to the static view	1,2,4,5,9
	Ray should better hit the plane directly instead of ending at the DICOM Cube	3,7
2D Interface	Data set should start centered for better orientation	1,3,4,8
	Preview image should be scalable	4,5
	Ray interaction leads to confusion while using Sliders	7,8
	Step by step slicing by using +/- symbol was interpreted as zoom function	2,5,8
Input Device	Ray interaction (slider movement) via controller is too inaccurate to slice data	1,4
Miscellaneous	Insufficient resolution of CT data in DICOM Cube	1,5,6,8,9
	Terms such as DICOM unknown	6,10
	Uncertainties during initialization of the DICOM board	1,2,3,5,6,9
General		
Multi-user	Adopting the fully immersive MR user view is helpful for better understanding	5,9,10
	No more than 5 people should be in VST MR mode at the same time	2,3,5,9
	Spectator does not disturb the immersion	6,7
	Spectator mode is suitable for passive participation in larger groups	1,2,3,6
	Teacher is in spectator mode and can passively support students in fully immersive MR mode	8,10
	fully immersive MR mode is especially suitable for small learning groups	1,3,8,10
Virtual Environment	VST MR mode looks more familiar, because participants and environment are in view	2,6,8
	Spatial conditions limit movement in VST MR mode	6,7
	The distraction from the environment is greater in VST MR mode	7,9,10
Locomotion	Preference for walking instead of teleporting in fully immersive MR mode	6,10
Miscellaneous	Implementation of VST MR mode seems unstable	4,8

7.2.5 *Discussion*

Overall, the prototype was well received by participants, with the *fully immersive MR mode* achieving a *Usability* score of 79, as measured by the SUS. This positive result can be attributed to the fact that most controls were explained in advance. However, many interactions were explored independently by users, with some being described as intuitive (e.g., enlarging 3D objects). The ability to directly grasp virtual organ models was positively noted. Nonetheless, the combination of direct interaction and indirect interaction (using a ray) occasionally caused confusion and limited user actions.

Participants highlighted the benefit of additional information for students. The preparation of the data and the presentation of the 3D model were particularly praised for providing realistic insights into liver anatomy (see Table 16, 3D Representation, 3D Interaction). The inclusion of separate functions to show or hide individual structures and enable highlighting was suggested as a means of improving contextual understanding.

Some terms and hierarchical structures were occasionally misunderstood, which might be due to the early stage of medical education of some participants (see Table 16, 2D Interface). Additionally, the shelf-based presentation had certain disadvantages; objects located in the lower compartments received less attention and were more challenging to reach. As a result, users often needed to bend down when manually returning a model, as only direct grabbing was possible (see Table 16, Spatial Arrangement).

7.2.5.1 *Information Board*

The *Information Board*, containing all case data and prepared similarly to a classic medical report, was positively received by participants. However, the interaction between the 3D model and the displayed text was identified as an area for improvement. It was suggested that selecting structures on the 3D model should automatically highlight corresponding passages in the text. Conversely, marking text passages could also trigger the highlighting of specific structures (see Table 16, Information Board).

7.2.5.2 *DICOM Board*

The interface of the *DICOM Board*, designed to resemble familiar desktop interactions from the real world, was positively received, as it allowed participants to navigate it quickly. However, the initial relationship between patient data (e.g., the 3D liver model) and the DICOM CT display was not immediately clear to users. Difficulties arose in activating the *DICOM Board*, as participants were required to adjust the CT slices using sliders first, which was not intuitive.

The creation of photocopies and the ability to interact freely with them were especially appreciated. This functionality allowed views from different sectional planes to be generated and discussed in parallel. However, the image resolution had been reduced to improve performance, which was negatively perceived by users, as the lower resolution made it difficult to discern structures clearly.

Almost all participants highlighted the benefits of presenting the data set as a three-dimensional cube. This representation, particularly in the *fully immersive MR mode*, was noted as effective for creating an interactive relationship to body positioning. However, some student participants experienced initial difficulties in orienting themselves with respect to the human body and suggested that a positional hint would be helpful. Additionally, there was a desire to display the liver directly within the cube to enable hybrid rendering (see Table 16, DICOM Board).

7.2.5.3 *Interaction*

With respect to the 3D interaction with the organ models and the *DICOM Cube*, it was suggested that the ray interaction be improved to enable both direct manipulation and precise pointing on surfaces or structures. While the adaptation of the 2D interface from real-world applications was noted as intuitive, some interface elements were reported as difficult to reach (e.g., small buttons) or inconsistent in their controls, being either too fine or too coarse (e.g., sliders). These limitations were attributed to the hardware choice (Vive Controller) and its implementation.

Issues related to poor readability and text size being too small were also reported. These challenges were associated with improper headset placement and the absence of lens correction adjustments.

7.2.5.4 *AR Mode*

VST MR technologies are known to provide a sense of presence, enabling cooperative and situational learning, which can be beneficial for educational purposes [330]. Unlike fully immersive MR, VST MR allows users to see each other, increasing the sense of social interaction. However, the VST MR application in this study could not yet be tested with multiple participants. Participants did, however, note potential spatial limitations and the increased space requirements associated with the *VST MR mode*.

It was suggested that the teacher might experience an enhanced sense of being seen and heard in VST MR, but participants raised concerns that the students' focus could be disrupted due to increased distractions. Additionally, the lower fidelity in the representation of structures was criticized. Although the interactions for system operation are identical to those in the fully immersive MR environment, the *Usability* was deemed less comparable due to differences in the perceived sense of presence.

7.2.5.5 *VR Mode*

The IPQ measurement shows that the *fully immersive MR mode* induces a sense of presence. Particularly high scores for the subscales "Spatial Presence" and "General Presence" indicate that participants in the sample felt present in the virtual environment and were able to act independently and freely. This suggests that the environment is well-suited as an explorative learning environment.

As expected, the subscale "Experienced Realism" received low scores due to the abstract representation of the virtual environment. The first item of this subscale, "How real did the virtual world seem to you?" recorded an Avg value of 2.2 (SD = 1.14). In contrast, the second item, "How much did your experience in the virtual environment seem consistent with your real-world experience?" received the highest rating within this subscale (M = 3.9; SD = 1.52), emphasizing the intuitive interactions and real-time communication.

On the subscale "Involvement," the Avg score of 4.8 (SD = 0.63) reflects a strong level of engagement. Specifically, the last item, "I was completely captivated by the virtual world," highlights the suitability of the environment for learning scenarios that demand concentration. However, the third item in this subscale, "I still paid attention to the real environment," received a low value, corresponding to the low degree of realism in the virtual environment.

7.2.5.6 *Learning Environment*

A high level of *Presence* enhances motivation, which is essential for achieving learning objectives. The fully immersive MR and VST MR environment promotes active learning by allowing users to explore anatomical structures through natural interactions, thereby enabling embodied cognition and reducing cognitive load [213]. The CMVLE approach facilitates collaborative learning groups, fostering communication and social interaction [109].

The evaluation revealed several potential learning scenarios. When all students participate in the fully immersive MR environment, the concept is particularly suited for small learning groups. In this scenario, students can benefit from interactivity, knowledge exchange, and collaborative discussions. However, this requires a more dynamic structure, allowing each participant or group to access individual information via dedicated boards. The placement system presented in the current concept is more appropriate for a traditional teacher-student setting, where the teacher serves as the primary source of information. The inclusion of the *Spectator Mode* was originally intended to enable participation for a larger audience without requiring specialized hardware.

Given the increased adoption of virtual teaching methods during the COVID-19 pandemic, this type of virtual environment could positively influence students' learning behavior. However, the evaluation highlighted that the current interaction possibilities are too limited, which may negatively impact motivation. For teaching concepts in this domain, it is recommended that the teacher adopts a passive role, providing support, instructions, and guidance to the students.

Additionally, a single observer does not appear to significantly impact the *Presence* of fully immersive MR users. However, this aspect should be further evaluated with larger groups, as multiple observers could potentially cause distractions. To address this, a fade-out and mute function would be a valuable addition. Furthermore, the perception of participants among themselves requires improvement. The inclusion of avatars with facial expressions and gestures could enhance communication and social interaction among users.

7.2.6 *Conclusion*

A CMVLE was presented to support students in liver surgery education using clinical cases. Various teaching scenarios were demonstrated, enabling collaborative and cooperative learning in diverse group constellations. The system was designed with multiple modes, allowing users to select configurations suitable for specific applications, including distance learning. Due to hygiene regulations, the VST MR system was only presented as a video demonstration. Future evaluations under more study-friendly conditions are expected to provide comprehensive insights. The integration of user feedback and subsequent improvements has been identified as a key next step. These enhancements should be incorporated into real training sessions and validated in those contexts. Additionally, the prototype remains expandable, both in terms of the number of participants and the volume and variety of medical data, including the potential integration of concepts such as the Bento Box [126].

This setup also holds potential for adaptation to training in additional surgical disciplines. The approach presented here offers a promising outlook for complementing complex theoretical and practical teaching content in surgical education using fully immersive MR and VST MR technologies.

This technical foundation, including some of the 3D models and interaction principles, served as the basis for further developments by Chheang et al. (see Chapter 3), particularly in [Further2] and [Further7], with contributions from the original paper discussed in this chapter. These advancements highlight the versatility and scalability of CLEs, building upon the foundational concepts established in this thesis.

7.3 CHAPTER SUMMARY

In this chapter, the final *Experiment 7* introduced a CMVLE that facilitates both independent self-exploration by students and teacher-guided learning. The system was designed to prepare students in advanced clinical semesters for liver surgery training. The chapter provided an overview of the medical background and challenges in liver anatomy. Requirements were gathered in collaboration with experts, allowing curated clinical cases to be integrated into an interactive MR environment. Furthermore, the developed MR system combines multiple modalities, enabling participation in a shared virtual space with varying levels of immersion. Potential teaching and learning scenarios using various modalities were explored, and approaches for integrating MR into anatomical education were proposed. The study demonstrated the potential of the CMVLE to support collaborative and explorative learning in liver surgery education. Positive feedback highlighted the system's *Usability* and *Presence* qualia, while areas for improvement included interaction consistency and the representation of structures within the VEs. This experiment represents the methodological conclusion of the investigations conducted in this thesis. The following chapter summarizes the overall findings, discusses the advantages and disadvantages of MR systems in medical education, addresses the limitations of this work, and outlines future potential.

ch. 8

SYNOPSIS *This chapter summarizes the findings of this thesis by addressing the research questions posed at the beginning and throughout the individual experiments. Furthermore, it discusses limitations and highlights directions for future research.*

8.1 THESIS SUMMARY

This thesis begins by posing the overarching question: *How can immersive experiences be designed to enrich medical education?* Its primary objective is to explore this question through a series of experiments employing various MR systems and learning approaches. The investigation starts with an experiment aimed at understanding the role of design components in creating engaging and effective MRs environment for medical education and task simulation.

8.1.1 *Effects of Visual and Interaction Fidelity on Medical Task Simulations*

RQ1 The first research question, *What are the critical visual and interaction fidelity factors that contribute to creating engaging and effective medical task simulations in MR?*, was addressed through *Experiment 1* (Chapter 4). This study identified key factors in VF and IMs that enhance *Usability* and engagement in MRs simulations. High levels of VF, such as realistic rendering, appropriate lighting, and detailed textures, were shown to improve measures like *General Presence*, *Involvement*, and UX. Interestingly, lower VF environments, while less immersive, remained effective for specific tasks, especially when development constraints or performance limitations were considered.

These findings align with previous research that highlights the positive influence of UX on *Presence* [37, 44], which in turn significantly affects a learner's ability to engage with their environment [63]. Enhanced *Presence* has been linked to increased motivation, problem-solving capabilities, and knowledge construction [109]. Thus, the experiment not only validates prior findings but also extends them by addressing medical education-specific use cases.

The simulation environment developed for this experiment was versatile, functioning both as a testing platform for prototypical evaluations and a training environment for various interventional scenarios. Tangible IMs were also compared to traditional controller-based inputs. The results showed no significant differences in task performance between the two IMs, suggesting that interaction fidelity depends more on individual user preferences than on the fidelity levels themselves, at least for the tasks studied. However, for tasks requiring higher precision, tangible IMs paired with realistic visual representations may offer advantages. This aligns with findings that mismatched fidelity levels—such as realistic visuals combined with subpar interaction paradigms—can lead to an *Uncanny Valley* effect, causing discomfort or distraction [186]. The study emphasizes the importance of maintaining congruence between the user's expectations (coherent sensory elements) and the VE to ensure an effective and engaging experience.

Thus, effective simulations must balance VF and IMs, aligning them with the specific task and context to maintain user expectations and engagement. These results underscore the interplay of visual and interaction fidelity in determining the effectiveness of medical simulations. While high fidelity enhances *Presence* and engagement, *Usability* and task performance depend on aligning fidelity with educational goals and user expectations. The insights from Chapter 4 provide a practical framework for designing VEs, offering a balance between realism, development effort, and user needs.

8.1.2 *Individual Learning Environments*

RQ2 The second research question, *How can suitable visualizations and interactions in MR be designed to effectively represent embryonic heart development?*, was addressed through *Experiment 2*, which focused on a specific application in anatomical education: the developmental processes of the embryonic heart. Foundational visualization and interaction principles were developed for this concrete use case, yielding significant insights.

Expert feedback emphasized the value of dynamic visualizations for understanding structural deformations, particularly when integrated into a fully immersive MR-based environment. The presented concept effectively fostered 3D comprehension of embryonic heart development, offering clear insights into complex morphological processes through streamlined visualizations and guided interactions tailored to the study sample. These visualizations successfully directed attention to key anatomical changes while minimizing visual distractions. The high levels of *Spatial Presence* and *General Presence* further underscore the efficacy of these models in immersing users and supporting anatomical understanding.

In terms of interaction, active deformation was shown to encourage engagement and motivation but posed initial challenges for some users. This highlights the importance of designing simplified interactivity and providing guidance, particularly for users with limited technical experience or lower spatial reasoning abilities.

In summary, effective visualizations in MR must balance simplicity with sufficient contextual detail, while interactions should accommodate varying levels of user expertise and spatial reasoning. Combining active engagement with guided interactions and ensuring adaptability for diverse educational contexts can significantly enhance the effectiveness of MR-based learning environments in representing embryonic heart development.

RQ3 Building upon the technical foundation established in the previous experiment, the third research question, *Are there measurable learning effects when using MR to understand embryonic heart develop-*

ment, and which factors influence these outcomes?, sought to explore the educational impact of MR on medical students.

Through improvements to the application developed in *Experiment 2*, based on user and expert feedback, a large-scale evaluation was conducted with medical students during their examination period in *Experiment 3*. The findings revealed that a single VLE session significantly enhanced knowledge of embryonic heart development, as evidenced by increased *Heart Embryology* scores in the tailored knowledge test. These gains were retained after two weeks, with only minimal decline in scores observed among the post-VLE group. This stability underscores the VLE's potential for supporting long-term memory retention. These results align with prior research indicating that immersive MR environments can enhance both immediate learning and sustained knowledge retention [335].

Presence, particularly as measured by *Spatial Presence* and *General Presence*, demonstrated a strong correlation with learning outcomes, echoing the findings of *RQ1*. While *Subjective Workload* and UX played secondary roles, higher engagement-related workload was associated with improved learning, emphasizing the importance of active engagement and immersion in educational settings.

The VLE proved effective in supporting both knowledge acquisition and retention, suggesting its utility as a supplementary learning tool for medical students.

In summary, this study demonstrated that MR-based VLEs can effectively enhance knowledge acquisition and retention for complex anatomical topics such as embryonic heart development. *Presence* emerged as a pivotal factor in facilitating learning, while UX, *Subjective Workload*, and immersive tendencies served as complementary influences. These findings reinforce the potential of VLEs as valuable educational tools, particularly when integrated into broader medical curricula.

8.1.3 *Collaborative Learning Environments*

RQ4 While previously discussed VLEs facilitate individual learning, integrating them into seminar-style settings or combining them with traditional teaching methods could further enhance their utility. This insight formed the basis for the development of the fourth research question, *What are the technical and pedagogical requirements for a collaborative MR-based system to effectively support the learning of embryonic heart development?*

To address this question, the existing concept developed in *Experiment 2 and 3* was adapted to new hardware, incorporating novel interaction and visualization techniques specifically tailored for a seminar-based use case. The system was evaluated in collaboration with experts to assess its functionality and pedagogical efficacy.

The evaluation revealed general acceptance among experts, who praised the system's design and functionality. Among the tested configurations, the *Flat* setup was preferred for its equal visibility and space efficiency in seminar settings. The MR-based approach was commended for its ability to guide small groups through step-by-step explanations, fostering vivid understanding and dynamic communication. The designated role distribution within the CLE was identified as a critical factor in enhancing the learning process. Moderators focusing on visualization and interaction significantly contributed to student engagement and comprehension. However, the learning approach was noted to be more effective for smaller groups, emphasizing the importance of tailoring the CLE to group size and interaction dynamics.

This experiment underscores the importance of balancing technical and pedagogical requirements to create an effective CLE for teaching complex anatomical topics like embryonic heart development. Key findings highlight the need for intuitive interaction modalities, ergonomic considerations, and tailored learning approaches to foster engagement and accommodate diverse learner needs. These insights serve as a foundation for optimizing the CLE and expanding its application to other organ systems and educational contexts.

RQ5 The next step was to determine how this concept could be effectively and practically applied, leading to the fifth research question of this thesis: *How can a collaborative MR-based learning environment for understanding embryonic heart development be effectively integrated into an anatomy seminar setting?*

To address this question, a simulated anatomy seminar was conducted, utilizing the refined prototype from *Experiment 4* in a user study. Moderators—recruited medical professionals with teaching experience—used the application to guide two students through the topic of embryonic heart development. Both students and moderators reported favorable experiences with the CLE, as reflected in high immersion scores and strong *Usability* ratings. The general acceptance of the approach was further supported by positive results from the *Technology Acceptance Model*.

This more traditional seminar-style setup demonstrated that students benefited from a reduced *Subjective Workload* compared to moderators, allowing them to concentrate more effectively on the learning content. The system proved particularly well-suited for small-group settings, encouraging active participation and peer-to-peer interactions. However, limitations were identified for larger groups, where interaction quality could diminish due to restricted space and visual clutter caused by multiple interaction markers.

RQ6 The CLEs developed thus far demonstrate strong potential as supplemental tools for anatomy seminars, effectively integrating interactive and immersive features to enhance anatomical education while fostering dynamic group interactions and engagement. However, further refinements—such as scaling for larger groups, increasing sample size in evaluations, enhancing visualization capabilities, and addressing hardware-related issues—are necessary to fully realize their potential. Similarly, the ILEs explored in *Experiments 2* and *3* proved efficient by promoting individual knowledge reconstruction and serving as complementary approaches for this use case. Consequently, the sixth research question arose: *How do individual and collaborative MR-based learning environments differ in supporting educational outcomes for embryonic heart development?*

To address this question, the study compared the individual learning environment (ILE) and the collaborative learning environment (CLE) to evaluate their effectiveness in supporting educational outcomes for embryonic heart development. Both approaches were implemented on the same (new) hardware, a VST HMD, to ensure comparability and leverage the benefits of immersive and semi-immersive MR systems.

The findings revealed no significant differences in knowledge gain between ILE and CLE groups, with evidence indicating comparable knowledge acquisition across the two environments. Both approaches were shown to effectively support the increase of anatomical knowledge related to embryonic heart development, confirming their value as educational tools.

Both ILE and CLE received high ratings for *Usability*, and no substantial differences in *Subjective Workload* were observed. However, students in the CLE groups demonstrated slightly higher engagement through increased social presence, attributed to interaction with real individuals in the collaborative environment. Participants expressed a marginal preference for ILE due to its self-directed and interactive features, while CLE was praised for fostering collaboration and enhancing social presence.

In conclusion, the study found no significant differences in educational outcomes between ILE and CLE, with both environments effectively supporting knowledge acquisition for embryonic heart development. While ILE promoted individualized learning with greater user control, CLE enhanced social presence and collaborative learning dynamics. The choice between ILE and CLE may depend on specific learning objectives and user preferences.

8.1.4 *Cross-Modality Learning Environments*

RQ7 The exploration of the transition from ILE to CLE underscored the practical advantages of collaborative MR environments in enhancing foundational medical education. Building on these findings, the final experiment aimed to broaden the scope by exploring liver anatomy education within the context of surgical training. While previous experiments primarily utilized abstract, animated visualizations to illustrate developmental processes, the integration of real clinical cases had not yet been addressed. Clinical cases provide an effective means of conveying applied knowledge about medical conditions and treatments, preparing students more comprehensively for clinical practice. Additionally, prior systems were designed distinctly for either student-driven self-exploration or teacher-guided learning. This raised the question of how students and instructors can efficiently share a virtual space to foster collaborative learning while maintaining educational effectiveness. These considerations led to the final research question: *What design principles can enhance collaborative MR environments for advanced medical training, specifically in liver anatomy education, by integrating real clinical cases and accommodating varying levels of immersion?*

To address this question, *Experiment 7* introduced a CMVLE tailored to prepare advanced medical students for liver surgery training. The system integrates multiple modalities, enabling participants to share a virtual space while accommodating varying levels of immersion. This approach bridges technological gaps, promotes barrier-free access to MR systems, and fosters inclusivity in educational contexts.

Co-developed with surgeons and evaluated by experts and students, the application revealed valuable insights into technical challenges, interaction mechanics, and collaborative dynamics. The fully immersive MR mode demonstrated good *Usability* and received high *Spatial Presence* ratings, reflecting strong engagement and effective interaction. Participants praised intuitive features like direct manipulation of 3D organ models, which facilitated anatomical exploration and group discussions. However, *Experienced Realism* received lower ratings, attributed to the abstract representation of the virtual environment compared to the detailed models.

The interplay between 2D and 3D medical data functioned effectively, providing valuable support for navigation and understanding. However, the integration between 3D models and corresponding 2D datasets was found to be less intuitive and offers room for further improvement.

The CMVLE showcased strong potential as an innovative tool for liver anatomy education. Key design principles for optimizing collaborative MR environments include ensuring intuitive and seamless integration between 3D models and medical imaging data, enhancing contextual understanding through advanced features such as dynamic visualization and hybrid rendering, implementing high-performance network structures to ensure a smooth UX for multiple users, and selecting hardware tailored to the specific application needs. These principles establish a robust foundation for advancing MR technologies in medical education and training.

8.2 LIMITATIONS & FUTURE DIRECTIONS

This thesis has several limitations, which are summarized in the following sections.

8.2.1 *Experimental Design and Fidelity*

The first limitation relates to the exploration of VF and its impact on learning outcomes and user engagement. While *Experiment 1* identified the importance of factors such as lighting, shadows, and textures, it did not isolate which specific components were most influential. Follow-up studies are needed to examine the individual effects of these VF elements on overall UX, task performance, and learning outcomes.

Additionally, the environments used across experiments varied significantly, from laboratory settings in *Experiment 2* (Chapter 5) to abstract designs in *Experiment 3*, real-world-inspired classrooms in *Experiment 6* (Chapter 6), and hybrid spaces in *Experiment 7* (Chapter 7). These differences likely influenced UX and *Presence* but were not systematically analyzed. Future research should investigate the interplay between environmental fidelity and educational effectiveness to identify optimal settings for sustained engagement and learning.

8.2.2 *Representation of Users and Interaction Modalities*

Limited attention was given to virtual user representation. Earlier experiments relied on AI-driven voice presence or UI-based interactions, while *Experiment 6* integrated real human interaction. Future studies could explore the use of photorealistic avatars or AI-generated virtual agents with expressive features, such as facial expressions and gestures, to enhance *Social Presence* and engagement, as suggested by prior research [Further2, 54].

Interaction modalities were also underexplored. While *Experiment 1* in Chapter 4 compared tangible IMs, later experiments did not evaluate how users might benefit from selecting their preferred interaction methods. For instance, hand-based interactions in Chapter 5 promoted 3D morphological understanding, but their comparative effectiveness against other modalities remains unclear. Future work should examine how interaction preferences influence *Usability* and learning across various tasks.

8.2.3 *Evaluation Metrics and Cognitive Load*

The evaluation methods used in this thesis presented challenges. While widely validated metrics such as UX, *Presence*, and *Usability* were employed, their variability across experiments limited comparability. Future studies should standardize evaluation frameworks to ensure consistency and generalizability.

The N-TLX was used to measure *Subjective Workload*, but it proved less suitable for assessing cognitive load specific to learning tasks. Domain-specific cognitive load questionnaires, such as those developed for educational VLEs, could provide more precise insights. Additionally, cybersickness, which could influence learning outcomes and UX, were not systematically addressed in this thesis.

Custom knowledge tests, developed in collaboration with experts and modeled after real examination questions, were used to measure learning outcomes. However, these tests were not validated through broader longitudinal studies. Employing standardized tests could improve the reliability and applicability of findings. Future research should also explore long-term retention through repeated VLE sessions, as suggested by Makransky et al. [162].

8.2.4 *Development Limitations and Hardware Variability*

The systems developed for this thesis were research prototypes, optimized for specific experimental needs rather than for end-user deployment. This led to certain technical constraints, such as transmission errors in CLEs and interaction inconsistencies, which may have affected the results.

Various HMDs were utilized to explore their respective advantages and limitations. However, this variability prevents firm conclusions regarding the ideal hardware for balancing ergonomics, VF, and *Usability*. Future studies should evaluate hardware-specific factors, such as comfort during prolonged use and the impact of display quality on learning outcomes.

8.2.5 *Generalizability*

Although the iterative HCD approach ensured that the systems were tailored to the needs of the target audience, the requirements and evaluations were primarily based on participants from a single institution. Educational needs, teaching methods, and available resources vary significantly across institutions. Future research should adapt and validate these systems in diverse educational contexts to ensure broader applicability.

Emerging technologies, such as AI-driven 3D model generation, offer promising avenues for scaling MRs systems cost-effectively. Leveraging these technologies could streamline development processes and reduce costs, making MRs systems more accessible to educational institutions.

Gamification elements could further enhance engagement and motivation, while uniform design guidelines and consistent feedback mechanisms could improve *Usability* and user satisfaction. The flexible, platform-independent approaches discussed in Chapter 7 offer a promising foundation for expanding MRs systems to broader educational applications.

Finally, while technology acceptance was high, the practical implementation of MR-based systems requires balancing development effort, resource allocation, and demonstrated educational effectiveness. Longitudinal studies evaluating cost-effectiveness and learning impact will be critical for the sustainable integration of MR technologies into medical education.

8.3 GENERAL CONTRIBUTION

How can immersive experiences be designed to enrich medical education? This guiding question shaped the foundation of this thesis. The answers lie in the research questions it raised. In line with the integration of MR and HCI considerations, this thesis employed a HCD process to identify and address the specific needs of students and teachers. Through this approach, the development of tailored learning applications was aligned with HCD principles, enabling systematic evaluation of various MR technologies for their effectiveness in achieving educational objectives while enhancing both *Usability* and learning outcomes. The findings of this work contribute to a deeper understanding of the challenges associated with integrating MR into medical education. They provide valuable insights into the design principles, technical implementation, and evaluation methodologies required for developing immersive technologies tailored to medical training contexts. Overall, this thesis advances the field of immersive technologies in medical education, offering promising avenues for improving learning outcomes and fostering greater user engagement.

APPENDIX

A

SYNOPSIS *The appendix provides essential supplementary materials for the conducted experiments, including demographic questionnaires, custom knowledge tests, a spatial reasoning test, moderator guidelines, and more.*

9.1 VIDEOS

This section provides an overview of the MR applications featured in this thesis, accompanied by QR codes that link directly to demonstration videos. Each QR code corresponds to a specific experiment and chapter, offering a visual representation of the described MR systems and their functionalities.

Experiment 1:
Chapter 4

Experiment 2:
Chapter 5

Experiment 3:
Chapter 5

Experiment 4:
Chapter 6

Experiment 5:
Chapter 6

Experiment 6:
Chapter 6

Experiment 7:
Chapter 7

9.2 DEMOGRAPHIC DATA SHEET

This questionnaire was used to survey demographic data and was employed in either its original form or with slight modifications in all experiments conducted as part of this work. For example, the activity status was adapted based on the specific target group being investigated, such as medical students, experts/teachers, or general users (e.g. with a technical background). Additionally, the last question about the use of virtual reality was replaced by questions pertaining to the technology used (e.g., augmented reality, mixed reality).

Alter: ____ **ID:** ____
Age

Geschlecht: [ ] weiblich [ ] männlich [ ] divers
Gender female male diverse

[ ] **Medizinstudent*In im Fachsemester:** ____
Medical student in the semester of study

Haben Sie Lehrerfahrung?
Do you have teaching experience?

[ ] Ja [ ] Nein
Yes No

Wenn ja, bitte kurze Beschreibung: ______________________________
If yes, please provide a brief description

Ist Ihr Sehvermögen beeinträchtigt?

[ ] Nein
No

[ ] Ja, aber korrigiert (Brille/Kontaktlinse, keine Einschränkungen im Alltag)
Yes, but corrected (no restrictions in everyday life)

[ ] Sonstiges: ______________ (Farbsinnstörung, Fehlsichtigkeit, …)
Others (Color vision deficiency, e.g. red-green vision deficiency)

Wie hoch schätzen Sie ihre technische Affinität ein?
How high do you rate your technical affinity?

Niedrig / Low — 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ — Hoch / High

Wie viel Erfahrung haben Sie im Umgang mit Computerspielen?
How much experience do you have with computer games?

Keine Erfahrung / No experience — 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ — Viel Erfahrung / A lot of experience

Wie viel Erfahrung haben Sie im Umgang mit Virtual Reality?
How much experience do you have with virtual reality?

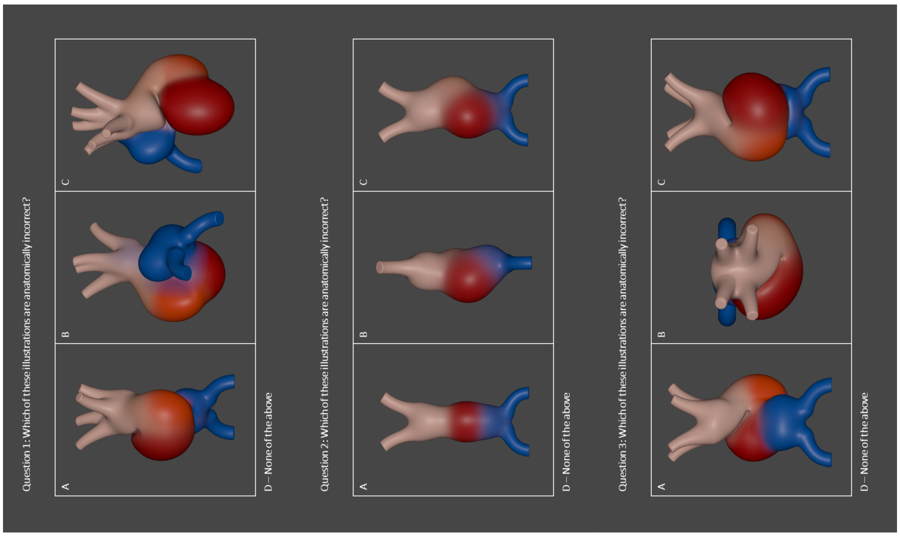
Keine Erfahrung / No experience — 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ — Viel Erfahrung / A lot of experience

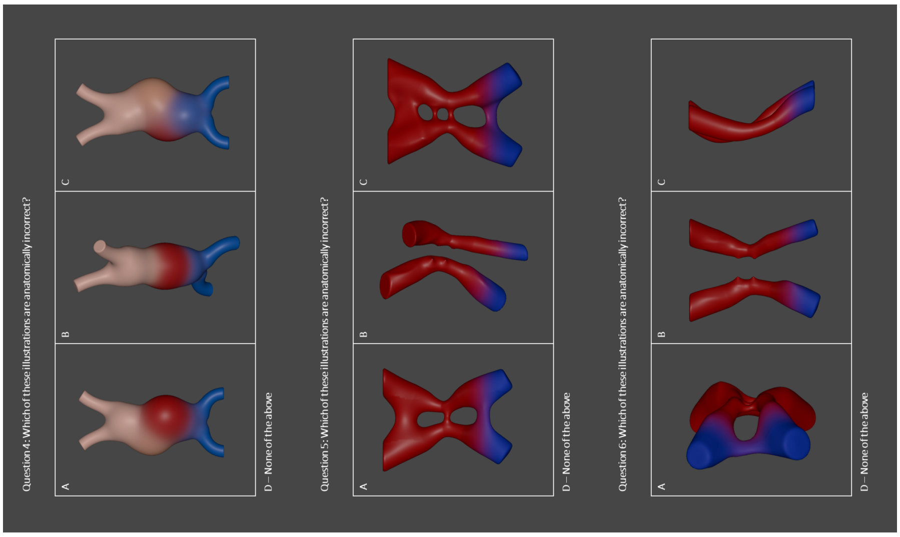
9.3 ANATOMY KNOWLEDGE TEST

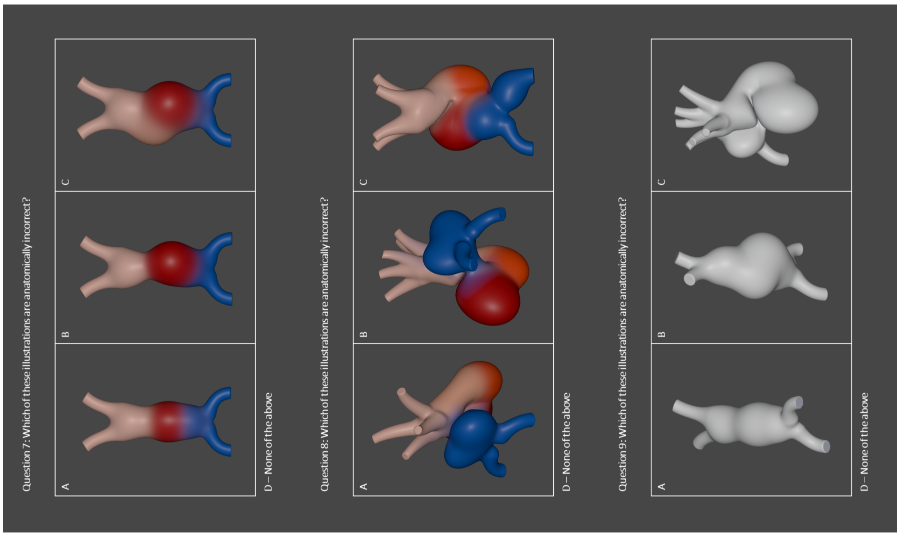
This section presents the anatomy knowledge test developed for use as described in *Experiment* 2 of Chapter 5. The test comprises 30 questions/tasks divided into three categories. To ensure variability and prevent repeated exposure, the order of questions/tasks and image placement on pages were randomized, resulting in two test versions for each participant. The images are screenshots from a Unity application developed specifically for the purposes of this study. In the experiment, the questions were formulated in German.

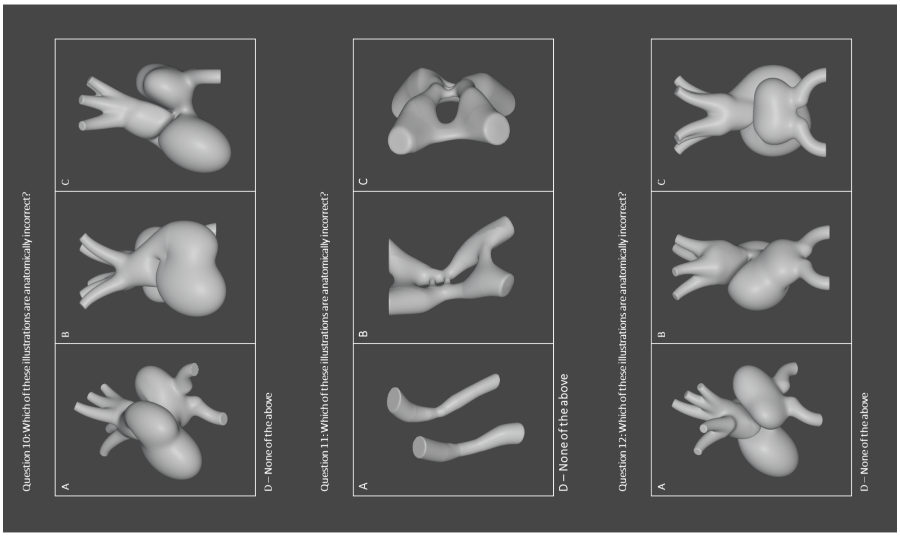
9.3.1 *Category 1: Shape*

This category consists of 15 questions where participants must identify illustrations that deviate from anatomical structures learned during training. Discrepancies are introduced through mirroring, omission, or distortion of sections.



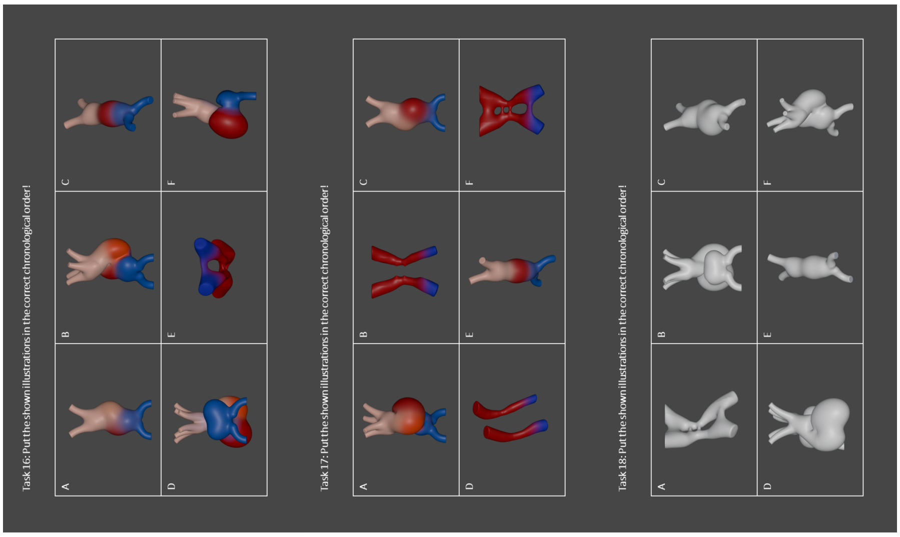


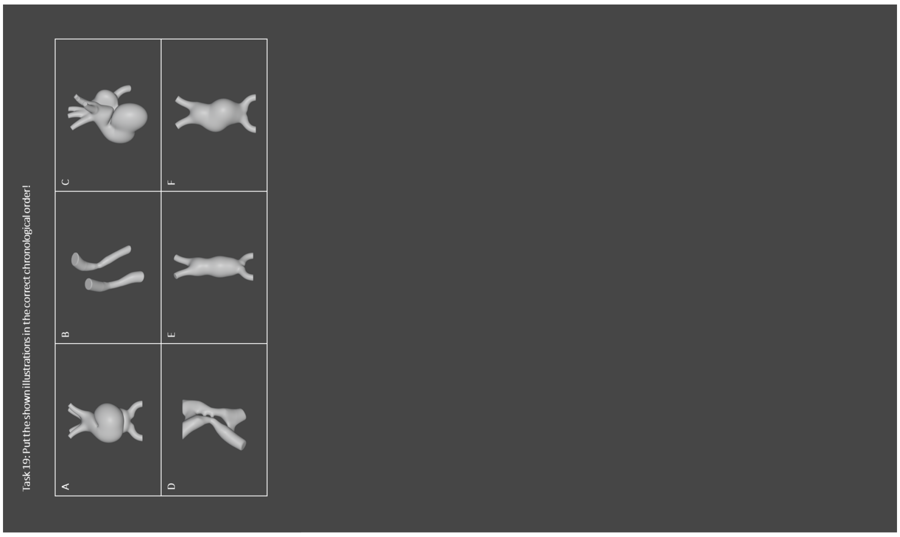




9.3.2 *Category 2: Time*

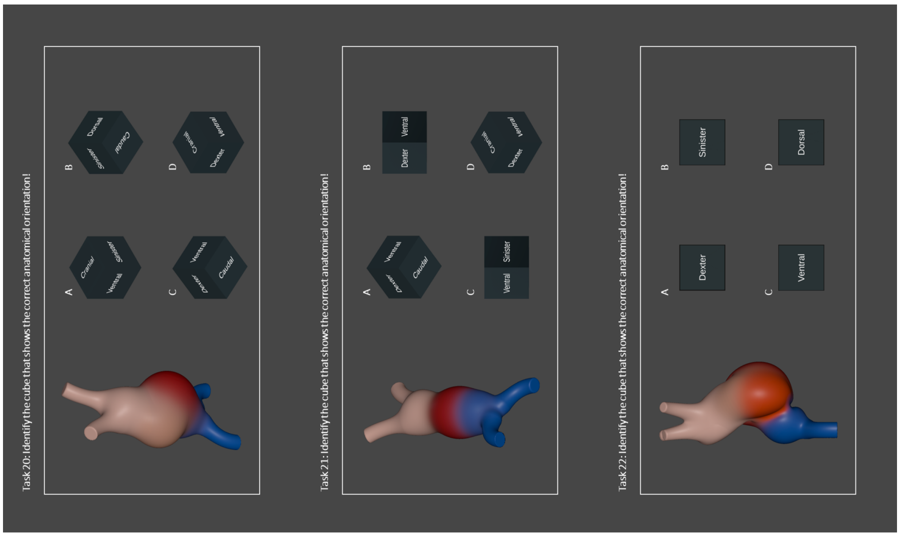
In this category, participants encounter 4 questions requiring them to arrange 6 illustrations in the correct chronological order.

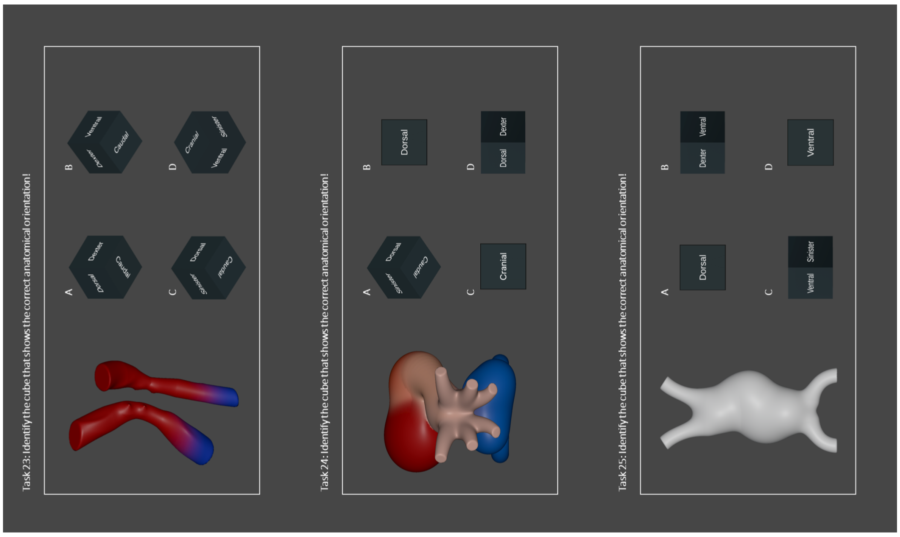

Task 16: Put the shown illustrations in the correct chronological order!
A
B
C
D
E
F
Task 17: Put the shown illustrations in the correct chronological order!
A
B
C
D
E
F
Task 18: Put the shown illustrations in the correct chronological order!
A
B
C
D
E
F

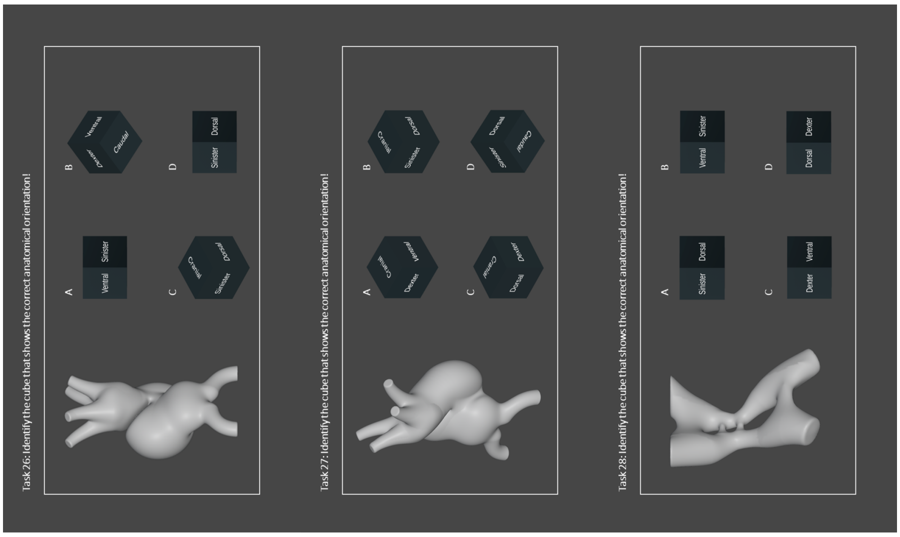

Task 19: Put the shown illustrations in the correct chronological order!
A
B
C
D
E
F

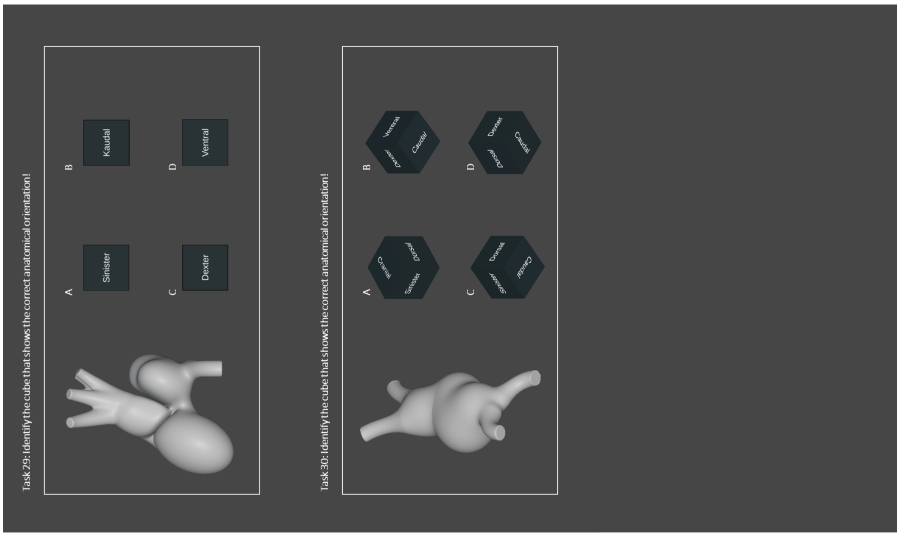
9.3.3 *Category 3: Location*

Comprising 11 questions, this category tasks participants with identifying a cube resembling the navigation cube from the VR application. They must select the variant that matches the illustration presented.



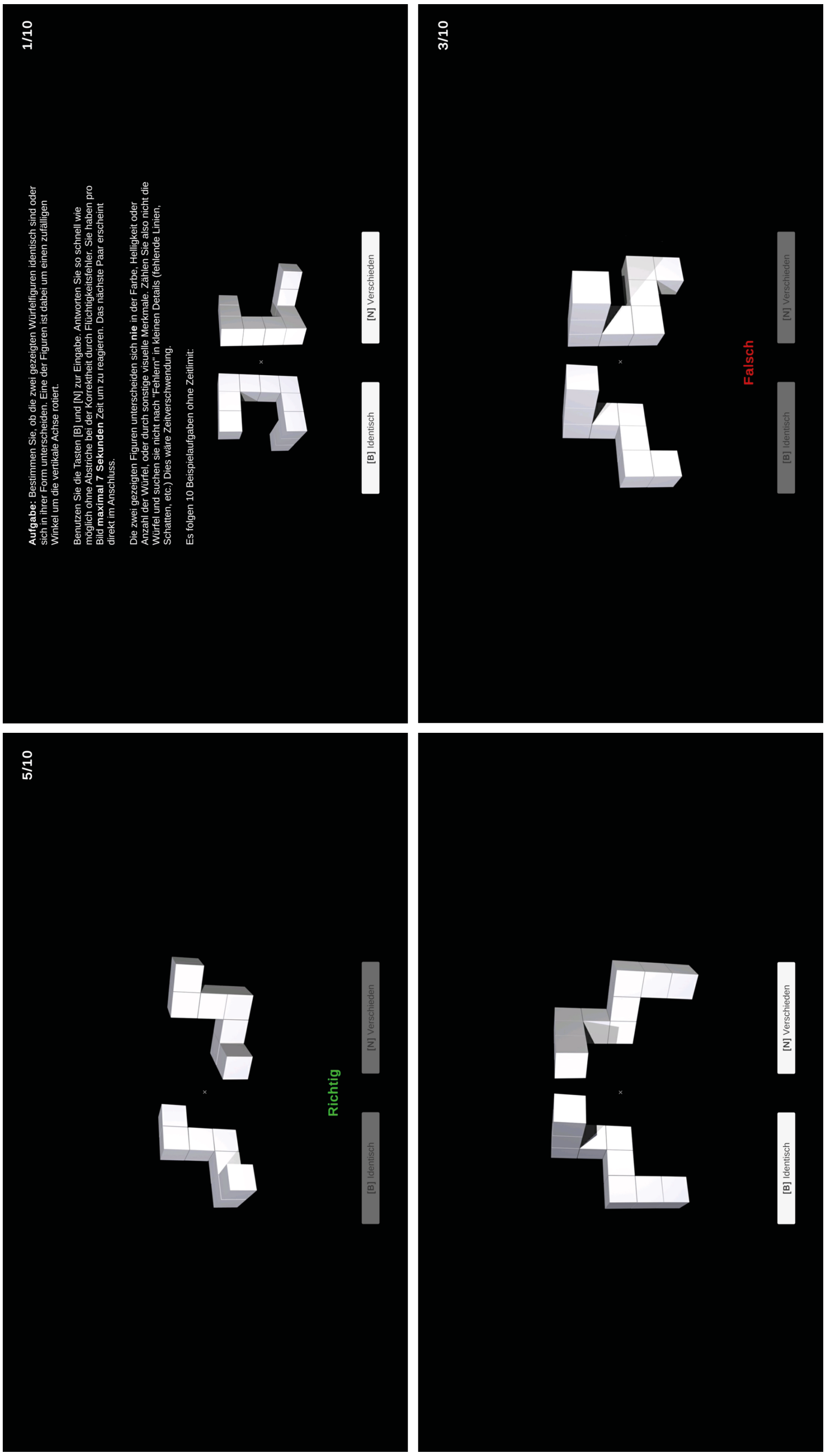






9.4 MENTAL ROTATION TEST

In this section, excerpts from the MRT used in *Experiment 2* of Chapter 5, following the variant proposed by Ganis and Kievit [89], are presented. The images are screenshots of the original Unity application, which was developed in German specifically for the purposes of this study. After the end of the practice phase (10 tasks), a button had to be pressed, which was followed by a countdown and the start of the measured test (96 tasks) that no longer provided feedback on whether the answers were correct or incorrect.



1/10
Aufgabe: Bestimmen Sie, ob die zwei gezeigten Würfelfiguren identisch sind oder sich in ihrer Form unterscheiden. Eine der Figuren ist dabei um einen zufälligen Winkel um die vertikale Achse rotiert.
Benutzen Sie die Tasten [B] und [N] zur Eingabe. Antworten Sie so schnell wie möglich ohne Abstriche bei der Korrektheit durch Flüchtigkeitsfehler. Sie haben pro Bild maximal 7 Sekunden Zeit um zu reagieren. Das nächste Paar erscheint direkt im Anschluss.
Die zwei gezeigten Figuren unterscheiden sich nie in der Farbe, Helligkeit oder Anzahl der Würfel, oder durch sonstige visuelle Merkmale. Zählen Sie also nicht die Würfel und suchen sie nicht nach "Fehlern" in kleinen Details (fehlende Linien, Schatten, etc.) Dies wäre Zeitverschwendung.
Es folgen 10 Beispielaufgaben ohne Zeitlimit:
[B] Identisch
[N] Verschieden
3/10
Falsch
[B] Identisch
[N] Verschieden
5/10
Richtig
[B] Identisch
[N] Verschieden
[B] Identisch
[N] Verschieden

9.5 ADVANCED EMBRYOLOGY KNOWLEGE TEST

Questionnaire for Assessing Knowledge in General Embryology (1 - 10) and Cardiac Embryology (11 - 20), as Employed in Section 5.3 of Chapter 5 and in Section 6.4 of Chapter 6. The questionnaire is provided in German, consistent with the study's methodology.

Allgemeine Embryologie

1) Bei ausbleibender Befruchtung erreicht ein Corpus luteum den Höhepunkt seiner Entwicklung ...
   a) 3 Tage nach der Ovulation
   b) 6 Tage nach der Ovulation
   c) 9 Tage nach der Ovulation
   d) 14 Tage nach der Ovulation

2) Die Decidua basalis ist ...
   a) Fetaler Anteil an der Plazenta
   b) Mütterlicher Anteil an der Plazenta
   c) Einziger gemeinsamer Anteil an der Plazenta
   d) Plazenta-assoziiertes extraplazentares Mesoderm

3) Welche Bestandteile gehören nicht zu einem Somiten oder stehen nicht im engen funktionellen Zusammenhang?
   a) Sklerotom
   b) Myom
   c) Dermatom
   d) Segmentaler Spinalnerv

4) Kinder diabetischer Mütter sind häufig ...
   a) überdurchschnittlich intelligent
   b) sehr viel häufiger Mädchen als Jungen
   c) schwerer und größer als normale Kinder
   d) gegen einen möglichen eigenen Diabetes durch Immunkontakt im Mutterleib geschützt

5) Welche Aussage zum fetalen Kreislauf bzw. fetalen Blut am Ende der Gravidität trifft normalerweise zu?
   a) Das Foramen ovale schließt sich vor der Geburt.
   b) Der mittlere Blutdruck in der V. umbilicalis beträgt etwa 100 mmHg.
   c) Die fetale Herzfrequenz beträgt mehr als 100 min–1.
   d) Die Hämoglobinkonzentration im fetalen Blut beträgt etwa 120 g/L.
   e) Die O2-Sättigung des Blutes ist in der fetalen Aorta etwa doppelt so hoch wie in der V. umbilicalis.

6) Der Körper des Embryos wird (im Alter von 7 Wochen) von verschiedenen Gebilden umgeben. Welche Reihenfolge trifft – vom Embryo aus gesehen – zu?
   a) Amnion – Chorion – Decidua
   b) Amnion – Decidua – Chorion
   c) Chorion – Amnion – Decidua
   d) Chorion – Decidua – Amnion
   e) Decidua – Chorion – Amnion

7) Welche der genannten Scheitel-Fersen-Längen entspricht am ehesten der eines Feten?
   a) 4 mm
   b) 8 mm
   c) 10 mm
   d) 30 mm
   e) 250 mm

8) In welcher Reihenfolge laufen die Stadien der Prophase I der ersten Reifeteilung ab?

a) Leptotän-Zygotän-Pachytän-Diplotän-Diakinese
b) Pachytän-Leptotän-Zygotän-Diplotän-Diakinese
c) Zygotän-Leptotän-Pachytän-Diplotän-Diakinese
d) Leptotän-Zygotän-Diplotän-Pachytän-Diakinese
e) Zygotän-Pachytän-Diplotän-Leptotän-Diakinese

9) Während der Embryonalentwicklung geben bestimmte Zellverbände ihren epithelialen Charakter auf und ihre Zellen migrieren zu anderen Stellen im Organismus. Dies geht am wahrscheinlichsten einher mit einer Verminderung von ...

a) Aktin
b) Colony-stimulating-Factor (CSF)
c) Desmin
d) E-Cadherin
e) Vimentin

10) Welche Aussage zum Ductus arteriosus (DA) trifft im Allgemeinen zu?

a) Er leitet in der Fetalzeit Blut aus der Aorta in den Truncus pulmonalis.
b) Er entwickelt sich aus der 3. Pharyngealbogenarterie links.
c) Er entwickelt sich aus der 4. Pharyngealbogenarterie links.
d) An seinem postnatalen Verschluss wirkt eine Intimaproliferation mit.
e) Er entsteht im 5. Schwangerschaftsmonat.

Herz Embryologie

11) Wo entsteht die Cardiogene Platte?

a) Viscerales Mesoderm
b) Im Bereich der letzten Somiten
c) Im Ektoderm
d) Im passenden Sklerotom
e) In der Decidua

12) Um welchen Tag entsteht die Herzanlage?

a) 8. Tag
b) 12. Tag
c) 14. Tag
d) 18. Tag
e) 24. Tag

13) Von Cranial nach Caudal kann der Herzschlauch eingeteilt werden (4.-5. Somiten):

a) Bulbus-Ventrikel-Atrium-Sinus venosus
b) Ventrikel-Atrium-Sinus venosus-Bulbus
c) Atrium-Bulbus-Ventrikel- -Sinus venosus
d) Sinus venosus- Atrium -Ventrikel-Bulbus
e) Sinus venosus- Atrium-Bulbus-Ventrikel

14) Die Faltung des Herzschlauches entsteht durch:

a) Symmetrisches Längenwachstum in einer unbegrenzten Höhle
b) Asymmetrisches Längenwachstum in einer unbegrenzten Höhle
c) Symmetrisches Längenwachstum und aktive Bewegung der Muskeln
d) Asymmtrisches Längenwachstum und aktive Bewegung der Muskeln
e) Asymmetrisches Längenwachstum in einer begrenzten Höhle

15) Die Valvulae venosae ...

a) Verschließen den Zufluss zum Herzen in der Diastole
b) Lenken den Blutstrom zum linken Ventrikel und linken Vorhof
c) Lenken den Blutstrom zum rechten Ventrikel und linken Vorhof
d) Sitzen im linken Vorhofsanteil
e) Bilden sich aus myoepikard Mantel

16) Die Anlagen der Segelklappen bilden sich aus:

a) Endokardkissen im AV-Kanal
b) Myokard an den Rändern der Ventrikel
c) Gallertigen Bindegewebe um den Herzschlauch
d) Endokardkissen in den Vorhöfen
e) Myokardialen Anteilen der Ausflusstraktes

17) Das linke Sinushorn entwickelt sich im Verlauf zu:

a) dem Sinus coronarius
b) dem Ramus interventrikulares anterior
c) der Vena cava superior
d) der Vena cava inferior
e) der Pulmonal vene

18) Das Septum primum verschließt:

a) das Foramen primum
b) Foramen interventrikulare
c) Foramen secundum
d) AV-Kanal
e) Foramen ovale

19) Welche Aussage über das Foramen sekundum trifft zu:

a) liegt im Septum primum
b) formt später das Foramen interventrikulare
c) bildet sich um den 20. Tag
d) liegt im Septum sekundum
e) liegt im Septum interventriculare

20) Welche Aussage über das Septum secundum trifft nicht zu:

a) Entwickelt sich um den 25. Tag
b) Wächst Richtung AV-Kanal
c) Bildet mit das Foramen ovale
d) Befindet sich rechts des Foramen primum
e) Hilft den Blutstrom im Herzen zu lenken

9.6 MODERATORS TRAINING QUESTIONS

This questionnaire served as a guide for the moderator in the experiment presented in Chapter 6. The questions reflected the content conveyed by the application, allowing the moderator the flexibility to choose which questions to ask. These were intended as guidelines and were not mandatory nor required to be followed verbatim. Responses could also be more elaborate, with the overall purpose being to provide orientation and stimulate discussion.

Phase 1: Paarige Endokardschläuche
Frage: "Wie entwickeln sich die paarigen Endokardschläuche aus der kardiogenen Platte?"
Antwort: "In der kardiogenen Platte entwickeln sich paarige Endokardschläuche. Diese Schläuche sind anfangs getrennt, aber beieinander liegend. Im Verlauf der ersten 10 Tage post conceptionem (nach der Befruchtung) wachsen und verlängern sie sich und beginnen zu verschmelzen."

Phase 2: Fusion zum Herzschlauch
Frage: "Wie unterteilt sich der fusionierte Endokardschlauch in die verschiedenen Abschnitte wie kranialen Bulbus, Ventrikel, Atrium und Sinus venosus?"
Antwort: "Der fusionierte Endokardschlauch gliedert sich in spezifische Abschnitte, wobei der kraniale Bulbus am vorderen Ende liegt, gefolgt vom Ventrikel, dem Atrium und schließlich dem Sinus venosus am hinteren Ende."

Frage: "Welche spätere Rolle spielt der kraniale Bulbus? Beschreiben Sie die Entwicklung des Atriums und des Sinus venosus."
Antwort: "Der kraniale Bulbus entwickelt sich im Verlauf zu den großen Arterien des Herzens. Das Atrium und der Sinus venosus entwickeln sich aus dem kaudalen Teil des Herzschlauchs."

Phase 3: Einknicken
Frage: "Was führt zur Bildung der Herzschlaufe? Was passiert mit dem Einfluss und Ausflusstrakten?"
Antwort: "Das Längenwachstum in der Begrenzung des Herzbeutels sorgt dafür, dass sich der Herzschlauch nach rechts und vorne biegt, wodurch die typische D-Schlaufe entsteht. Der Einfluss- und Ausflusstrakt bleiben dabei zentral positioniert."

Phase 4: Schleifenbildung
Frage: "Wie formen die Krümmungen des Herzschlauchs die drei gegeneinander geknickten Abschnitte?"
Antwort: "Die drei Krümmungen des Herzschlauchs formen die unterschiedlichen Abschnitte: den kaudalen Abschnitt (Sinus venosus), den mittleren Abschnitt (Ventrikel) und den kranialen Abschnitt (Ausflusstrakt)."

Phase 5: Sinus venosus
Frage: "Wie beeinflusst das asymmetrische Wachstum die Position des Sinus venosus und die Entwicklung des Sinus coronarius?"
Antwort: "Durch das asymmetrische Wachstum verlagert sich der Sinus venosus zunehmend nach rechts. Das linke Sinushorn bildet sich zurück und wird zum Sinus coronarius."

Frage: "Welche Rolle spielen die valvulae venosae in Bezug auf den Blutfluss durch den Sinus venosus?"
Antwort: "Die valvulae venosae sind entscheidend für die Lenkung des Blutstroms in die verschiedenen Teile des Herzens. Sie lenken einen Teil des Blutes zum linken Vorhof und einen anderen Teil zum rechten Ventrikel."

Phase 6.1: AV-Kanal

Frage: "Wie wird der AV-Kanal durch die einwachsenden Endokardkissen unterteilt und wie beeinflusst dies den Blutstrom?"
Antwort: "Die Endokardkissen wachsen ventral und dorsal in den AV-Kanal ein und unterteilen ihn, wodurch die Blutströme in den linken und rechten Ventrikel gelenkt werden und eine Unterteilung der Flüsse beginnt."

Frage: "Wie entstehen die Segelklappen?"
Antwort: "Die Endokardkissen bilden auch den Ursprung der Segelklappen. Diese Klappen spielen eine wichtige Rolle als Ventil zwischen den Vorhöfen und den Kammern, die den Rückstrom des Blutes aus dem Ventrikel ins Atrium in der Systole verhindern."

Phase 6.2: Septum primum

Frage: "Wie führt das Septum primum zur Septierung des einheitlichen Vorhofes und zur Bildung des Foramen secundum?"
Antwort: "Das Septum primum wächst in den einheitlichen Vorhof hinein und versucht, das Foramen primum zu schließen. Dabei bildet sich im oberen Abschnitt das Foramen secundum."

Frage: "Wozu dient das Septum secundum?"
Antwort: "Das Septum secundum bildet sich rechts des Septum primum und deckt das Foramen secundum ab, lässt jedoch caudal das Foramen ovale frei. Dies ermöglicht einen rechts-links Shunt durch das Atriale Septum bis zur Druckumkehr nach der Geburt."

Phase 6.3: Kammerseptum

Frage: "Wie entwickelt sich das interventrikulare Septum und wie beeinflusst dies den Blutkreislauf?"
Antwort: "Das interventrikulare Septum bildet sich durch eine Mischung aus Einschnürung und Einwachsen des Septums aus Myozyten. Es lässt das Foramen interventrikulare offen, was die Teilung des Blutkreislaufs mit der Teilung des AV-Kanals bewirkt."

Phase 7: Foramen secundum & Septum secundum

Frage: "Wie entwickelt sich das Foramen secundum weiter und was geschieht mit dem Septum primum und secundum nach der Geburt?"
Antwort: "Das Foramen secundum entwickelt sich weiter, während das Septum primum und secundum zur Bildung des interatrialen Septums beitragen. Nach der Geburt drückt das Septum primum auf das Foramen ovale und verschließt es, wobei die Septen miteinander verschmelzen, was entscheidend für den Übergang des Blutkreislaufs nach der Geburt ist."

Frage: "Welche Rolle spielt das Septum secundum in der Herzentwicklung, insbesondere in Bezug auf das Foramen ovale?"
Antwort: "Das Septum secundum spielt eine entscheidende Rolle bei der Abdeckung des Foramen secundum und der Bildung des Foramen ovale. Es trägt dazu bei, dass das Foramen ovale nach der Geburt verschlossen wird, was für die normale Funktion des Herzens nach der Geburt wichtig ist."

9.7 EXPERT INTERVIEW QUESTIONS

Questionnaires distributed to educators after the study to collect feedback on the use of the MR system described in *Experiment 6* in Chapter 6.

Vielen Dank für eure Unterstützung beim Testen der Anwendung! Folgend haben wir Fragen aufgeführt, die uns dabei helfen sollen, die Anwendung zu verbessern.

1. Wie würden Sie Ihre allgemeine Erfahrung mit der Leitung der Multi-User-AR-Sitzungen beschreiben? Was fanden sie positiv/negativ und wenn ja wieso?

2. Gab es Unterschiede in der Interaktion und im Engagement der Studierenden im Vergleich zu traditionellen Lehrmethoden?

3. Es war festzustellen, dass die Studierenden durch die Multi-User-AR-Anwendung ein tieferes Verständnis der embryonalen Herzentwicklung erlangt haben. (Farblich markieren/unterstreichen)

   Stimme überhaupt nicht zu 1 2 3 4 5 Stimme voll und ganz zu

4. Welche Aspekte der Anwendung glauben Sie, haben am meisten zum Lernen/Lehren beigetragen?

5. Gab es technische Schwierigkeiten oder Herausforderungen bei der Nutzung der AR-Plattform?

6. Wie empfanden sie die allgemeine Bedienung der AR-Anwendung (Rotation, Bewegen, Schaltflächen, Schieberegler)? Wie nützlich waren die Funktionen wie das Ein- und Ausschalten von Beschriftungen und das Forcieren der Sicht auf das 3D-Modell?

7. Das spezielle „Dozenten-Interface“ war intuitiv zu bedienen.

   Stimme überhaupt nicht zu 1 2 3 4 5 Stimme voll und ganz zu

8. Inwiefern haben die Studierenden von den Bedienmöglichkeiten Gebrauch gemacht? Hatten die Studierenden Schwierigkeiten mit der Bedienung und wenn ja, welche genau?

9. Welches Feedback haben Sie von den Studierenden über die Multi-User-AR-Anwendung erhalten?

10. Welche zusätzlichen Funktionen könnten die Lehre innerhalb dieser Anwendung verbessern?

11. Welche Herausforderungen oder Bedenken haben Sie bezüglich der Integration solcher Technologien in den Lehrplan?

12. Wie könnten diese Anwendungen weiterentwickelt werden, um das Engagement und die Motivation zu steigern?

[1] Eiman M Abdel Meguid, Jane C Holland, Iain D Keenan, and Priti Mishall. "Exploring Visualisation for Embryology Education: A Twenty-First-Century Perspective." In: *Biomedical Visualisation: Volume 11*. Ed. by Paul M Rea. Cham: Springer, 2022, pp. 173–193. ISBN: 978-3-030-87779-8. DOI: 10.1007/978-3-030-87779-8_8.

[2] A. Abildgaard, A. K. Witwit, J. S. Karlsen, E. A. Jacobsen, B. Tennøe, G. Ringstad, and P. Due-Tønnessen. "An autostereoscopic 3D display can improve visualization of 3D models from intracranial MR angiography." In: *Int J Comput Assist Radiol Surg* 5.5 (2010), pp. 549–554.

[3] Dor Abrahamson and Arthur Bakker. "Making sense of movement in embodied design for mathematics learning." In: *Cognitive research: principles and implications* 1 (2016), pp. 1–13.

[4] G Adam, A Bücker, C Nolte-Ernsting, J Tacke, and RW Günther. "Interventional MR imaging: percutaneous abdominal and skeletal biopsies and drainages of the abdomen." In: *European Radiology* 9 (1999), pp. 1471–1478.

[5] Alex Adkins, Ryan Canales, and Sophie Jörg. "Hands or Controllers? How Input Devices and Audio Impact Collaborative Virtual Reality." In: VRST '24. Trier, Germany: Association for Computing Machinery, 2024. ISBN: 9798400705359. DOI: 10.1145/3641825.3687718.

[6] Paola Aiello, Francesca D'Elia, Stefano Di Tore, and Maurizio Sibilio. "A constructivist approach to virtual reality for experiential learning." In: *E-learning and Digital Media* 9.3 (2012), pp. 317–324.

[7] Medhat Alaker, Greg R Wynn, and Tan Arulampalam. "Virtual reality training in laparoscopic surgery: a systematic review & meta-analysis." In: *International Journal of Surgery* 29 (2016), pp. 85–94.

[8] Amy L Alexander, Tad Brunyé, Jason Sidman, Shawn A Weil, et al. "From gaming to training: A review of studies on fidelity, immersion, presence, and buy-in and their effects on transfer in pc-based simulations and games." In: *DARWARS Training Impact Group* 5 (2005), pp. 1–14.

[9] Robin Alexander. *Essays on pedagogy*. Routledge, 2013.

[10] Salsabeel F. M. Alfalah, Jannat F. M. Falah, Tasneem Alfalah, Mutasem Elfalah, Nadia Muhaidat, and Orwa Falah. "A comparative study between a virtual reality heart anatomy system and traditional medical teaching modalities." In: *Virtual Reality* 23.3 (2019), pp. 229–234. ISSN: 1434-9957. DOI: 10.1007/s10055-018-0359-y.

[11] James Algina, H. J. Keselman, and Randall D. Penfield. "An Alternative to Cohen's Standardized Mean Difference Effect Size: A Robust Parameter and Confidence Interval in the Two Independent Groups Case." In: *Psychological Methods* 10.3 (2005), pp. 317–328. DOI: 10.1037/1082-989x.10.3.317.

[12] Devon Allcoat, Tim Hatchard, Freeha Azmat, Kim Stansfield, Derrick Watson, and Adrian von Mühlenen. "Education in the digital age: Learning experience in virtual and mixed realities." In: *Journal of Educational Computing Research* 59.5 (2021), pp. 795–816.

[13] Robert H. Anderson, David Bolender, Shumpei Mori, and Justin T. Tretter. "Virtual Reality Perhaps, but Is this Real Cardiac Anatomy?" In: *Clinical Anatomy* 32.4 (2019), pp. 468–468. DOI: 10.1002/ca.23306.

[14] Jonas Auda, Uwe Gruenefeld, Sarah Faltaous, Sven Mayer, and Stefan Schneegass. "A scoping survey on cross-reality systems." In: *ACM Computing Surveys* 56.4 (2023), pp. 1–38.

[15] Laura Aymerich-Franch, Cody Karutz, Jeremy N Bailenson, et al. "Effects of facial and voice similarity on presence in a public speaking virtual environment." In: *Proceedings of the international society for presence research annual conference*. 2012, pp. 24–26.

[16] Ronald T Azuma. "A Survey of Augmented Reality." In: *Presence: Teleoperators and Virtual Environments/MIT press* (1997).

[17] Ido Badash, Karen Burtt, Carlos A. Solorzano, and Joseph N. Carey. "Innovations in surgery simulation: a review of past, current and future techniques." In: *Annals of Translational Medicine* 4.23 (2016).

[18] Bernadette S de Bakker, Kees H de Jong, Jaco Hagoort, Roelof-Jan Oostra, and Antoon F M Moorman. "Towards a 3-dimensional atlas of the developing human embryo: The Amsterdam experience." In: *Reproductive Toxicology* 34.2 (2012), pp. 225–236. ISSN: 0890-6238. DOI: 10.1016/j.reprotox.2012.05.087.

[19] S Balakrishnan, M Syed Shahul Hameed, K Venkatesan, and G Aswin. "Interaction of spatial computing in augmented reality." In: *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*. Vol. 1. IEEE. 2021, pp. 1900–1904.

[20] Aaron Bangor, Philip Kortum, and James Miller. "Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale." In: *Journal of Usability Studies* 4.3 (2009), pp. 114–123.

[21] Jörg Barkhausen, Thomas Kahn, Gabriele A Krombach, Christiane K Kuhl, Joachim Lotz, David Maintz, Jens Ricke, Stefan O Schönberg, Thomas J Vogl, Frank K Wacker, et al. "White paper: interventional MRI: current status and potential for development considering economic perspectives, part 1: general application." In: *RöFo-Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren*. Vol. 189. 07. © Georg Thieme Verlag KG. 2017, pp. 611–623.

[22] Jörg Barkhausen, Thomas Kahn, Gabriele A Krombach, Christiane K Kuhl, Joachim Lotz, David Maintz, Jens Ricke, Stefan O Schönberg, Thomas J Vogl, Frank K Wacker, and German Association of Chairmen in Academic Radiology (KLR). "White Paper: Interventional MRI: Current Status and Potential for Development Considering Economic Perspectives, Part 2: Liver and Other Applications in Oncology." English. In: *RöFo-Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren* 189.11 (Nov. 2017). Epub 2017 Sep 1. PMID: 28863413, pp. 1047–1054. DOI: 10.1055/s-0043-112336.

[23] Sandra Barteit, Lucia Lanfermann, Till Bärnighausen, Florian Neuhann, Claudia Beiersmann, et al. "Augmented, mixed, and virtual reality-based head-mounted devices for medical education: systematic review." In: *JMIR serious games* 9.3 (2021), e29080.

[24] Mehmet Ilker Berkman and Ecehan Akan. "Presence and Immersion in Virtual Reality." In: *Encyclopedia of Computer Graphics and Games*. Ed. by Newton Lee. Cham: Springer International Publishing, 2019, pp. 1–10. ISBN: 978-3-319-08234-9. DOI: 10.1007/978-3-319-08234-9_162-1.

[25] Mehmet Ilker Berkman and Ecehan Akan. "Presence and immersion in virtual reality." In: *Encyclopedia of computer graphics and games*. Springer, 2024, pp. 1461–1470.

[26] Frank Biocca. "The cyborg's dilemma: Progressive embodiment in virtual environments." In: *Journal of computer-mediated communication* 3.2 (1997), JCMC324.

[27] Frank Biocca, Chad Harms, and Judee K Burgoon. "Toward a more robust theory and measure of social presence: Review and suggested criteria." In: *Presence: Teleoperators & virtual environments* 12.5 (2003), pp. 456–480.

[28] Frank Biocca, Chad Harms, and Jenn Gregg. "The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity." In: *4th annual international workshop on presence, Philadelphia, PA*. 2001, pp. 1–9.

[29] S. Birr, J. Mönch, D. Sommerfeld, U. Preim, and B. Preim. "The LiverAnatomyExplorer: a WebGL-based surgical teaching tool." In: *IEEE Comput Graph Appl* 33.5 (2013), pp. 48–58.

[30] Jim Blascovich. "Social Influence within Immersive Virtual Environments." In: *The Social Life of Avatars: Presence and Interaction in Shared Virtual Environments*. Ed. by Ralph Schroeder. London: Springer London, 2002, pp. 127–145. ISBN: 978-1-4471-0277-9. DOI: 10.1007/978-1-4471-0277-9_8.

[31] Susanne Bødker, Christian Dindler, Ole S. Iversen, and Rachel C. Smith. "What Is Participatory Design?" In: *Participatory Design*. Cham: Springer International Publishing, 2022, pp. 5–13. ISBN: 978-3-031-02235-7. DOI: 10.1007/978-3-031-02235-7_2. URL: https://doi.org/10.1007/978-3-031-02235-7_2.

[32] Anna Bolder, Stefan M Grünvogel, and Emanuel Angelescu. "Comparison of the usability of a car infotainment system in a mixed reality environment and in a real car." In: *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*. 2018, pp. 1–10.

[33] Michael Bonfert, Thomas Muender, Ryan P McMahan, Frank Steinicke, Doug Bowman, Rainer Malaka, and Tanja Döring. "The Interaction Fidelity Model: A Taxonomy to Distinguish the Aspects of Fidelity in Virtual Reality." In: *arXiv preprint arXiv:2402.16665* (2024).

[34] Felix Bork, Alexander Lehner, Ulrich Eck, Nassir Navab, Jens Waschke, and Daniela Kugelmann. "The effectiveness of collaborative augmented reality in gross anatomy teaching: a quantitative and qualitative pilot study." In: *Anatomical Sciences Education* 14.5 (2021), pp. 590–604.

[35] Doug A Bowman and Ryan P McMahan. "Virtual reality: how much immersion is enough?" In: *Computer* 40.7 (2007), pp. 36–43.

[36] Doug A. Bowman, Ryan P. McMahan, and Eric D. Ragan. "Questioning naturalism in 3D user interfaces." In: *Commun. ACM* 55.9 (Sept. 2012), pp. 78–88. ISSN: 0001-0782. DOI: 10.1145/2330667.2330687.

[37] Jennifer Brade, Mario Lorenz, Marc Busch, Niels Hammer, Manfred Tscheligi, and Philipp Klimant. "Being there again–Presence in real and virtual environments and its relation to usability and user experience using a mobile navigation task." In: *International Journal of Human-Computer Studies* 101 (2017), pp. 76–87.

[38] Harry Brenton, Juan Hernandez, Fernando Bello, Paul Strutton, Sanjay Purkayastha, Tony Firth, and Ara Darzi. "Using multimedia and Web3D to enhance anatomy teaching." In: *Computers & Education* 49.1 (2007). Web3D Technologies in Learning, Education and Training, pp. 32–53. ISSN: 0360-1315. DOI: 10.1016/j.compedu.2005.06.005.

[39] Emily Brown and Paul Cairns. "A grounded investigation of game immersion." In: *CHI'04 extended abstracts on Human factors in computing systems*. 2004, pp. 1297–1300.

[40] Judith R Brown. "Enabling educational collaboration—a new shared reality." In: *Computers & Graphics* 24.2 (2000), pp. 289–292.

[41] Richard KJ Brown, Sean Petty, Stephanie O'Malley, Jadranka Stojanovska, Matthew S Davenport, Ella A Kazerooni, and Daniel Fessahazion. "Virtual reality tool simulates MRI experience." In: *Tomography* 4.3 (2018), p. 95.

[42] Tim Brown et al. "Design thinking." In: *Harvard business review* 86.6 (2008), p. 84.

[43] Saniye Tugba Bulu. "Place presence, social presence, co-presence, and satisfaction in virtual worlds." In: *Computers & Education* 58.1 (2012), pp. 154–161.

[44] Oscar I Caldas, Natalia Sanchez, Mauricio Mauledoux, Oscar F Avilés, and Carlos Rodriguez-Guerrero. "Leading presence-based strategies to manipulate user experience in virtual reality environments." In: *Virtual Reality* 26.4 (2022), pp. 1507–1518.

[45] Nora Chekrouni, Roeland P Kleipool, and Bernadette S de Bakker. "The impact of using three-dimensional digital models of human embryos in the biomedical curriculum." In: *Annals of Anatomy - Anatomischer Anzeiger* 227 (2020), p. 151430. ISSN: 0940-9602. DOI: 10.1016/j.aanat.2019.151430.

[46] Chwen Jen Chen, Seong Chong Toh, and Wan Mohd Fauzy Wan Ismail. "Are Learning Styles Relevant To Virtual Reality?" In: *Journal of Research on Technology in Education* 38.2 (2005), pp. 123–141. DOI: 10.1080/15391523.2005.10782453.

[47] Long Chen, Thomas W Day, Wen Tang, and Nigel W John. "Recent developments and future challenges in medical mixed reality." In: *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE. 2017, pp. 123–135.

[48] V. Chheang, P. Saalfeld, T. Huber, F. Huettl, W. Kneist, B. Preim, and C. Hansen. "Collaborative Virtual Reality for Laparoscopic Liver Surgery Training." In: *Proc. of IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. 2019, pp. 1–17.

[49] Vuthea Chheang, Virve Fischer, Holger Buggenhagen, Tobias Huber, Florentine Huettl, Werner Kneist, Bernhard Preim, Patrick Saalfeld, and Christian Hansen. "Toward interprofessional team training for surgeons and anesthesiologists using virtual reality." In: *International journal of computer assisted radiology and surgery* 15 (2020), pp. 2109–2118.

[50] Vuthea Chheang, Patrick Saalfeld, Tobias Huber, Florentine Huettl, Werner Kneist, Bernhard Preim, and Christian Hansen. "Collaborative Virtual Reality for Laparoscopic Liver Surgery Training." In: *2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. 2019, pp. 1–17. DOI: 10.1109/AIVR46125.2019.00011.

[51] Vuthea Chheang, Patrick Saalfeld, Fabian Joeres, Christian Boedecker, Tobias Huber, Florentine Huettl, Hauke Lang, Bernhard Preim, and Christian Hansen. "A collaborative virtual reality environment for liver surgery planning." In: *Computers & Graphics* 99 (2021), pp. 234–246.

[54] Vuthea Chheang, Shayla Sharmin, Rommy Márquez-Hernández, Megha Patel, Danush Rajasekaran, Gavin Caulfield, Behdokht Kiafar, Jicheng Li, Pinar Kullu, and Roghayeh Leila Barmaki. "Towards anatomy education

with generative AI-based virtual assistants in immersive virtual reality environments." In: *2024 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR)*. IEEE. 2024, pp. 21–30.

[55] Kirsty Clarke, Aws Al-Mukhtar, Lina Alim, Amr Nimer, Emma Hatfield, Sanjeeve Sabharwal, and Sadie Syed. "Using Mixed Reality Simulation to Improve Junior Medical Trainees' Preparedness to Manage High-Acuity Trauma." In: *BMJ Open Quality* 13.2 (2024), e002575.

[56] Stephan Clasen and Philippe L Pereira. "Magnetic resonance guidance for radiofrequency ablation of liver tumors." In: *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine* 27.2 (2008), pp. 421–433.

[57] Murat Coban, Yusuf Islam Bolat, and Idris Goksu. "The potential of immersive virtual reality to enhance learning: A meta-analysis." In: *Educational Research Review* 36 (2022), p. 100452. ISSN: 1747-938X. DOI: https://doi.org/10.1016/j.edurev.2022.100452.

[58] Nigel Cross. *Design thinking: Understanding how designers think and work*. Bloomsbury Publishing, 2023.

[59] J. L. Crossingham, J. Jenkinson, N. Woolridge, S. Gallinger, G. A. Tait, and C. A. Moulton. "Interpreting three-dimensional structures from two-dimensional images: a web-based interactive 3D teaching model of surgical liver anatomy." In: *HPB (Oxford)* 11.6 (Sept. 2009), pp. 523–528.

[60] Barney Dalgarno and Mark JW Lee. "What are the learning affordances of 3-D virtual environments?" In: *British journal of educational technology* 41.1 (2010), pp. 10–32.

[61] Fred D Davis, RP Bagozzi, and PR Warshaw. "Technology acceptance model." In: *J Manag Sci* 35.8 (1989), pp. 982–1003.

[62] Fred D Davis, Andrina Granić, and Nikola Marangunić. *The technology acceptance model: 30 years of TAM*. Springer, 2024.

[63] Sara De Freitas, Genaro Rebolledo-Mendez, Fotis Liarokapis, George Magoulas, and Alexandra Poulovassilis. "Learning as immersive experiences: Using the four-dimensional framework for designing and evaluating immersive learning experiences in a virtual world." In: *British Journal of Educational Technology* 41.1 (2010), pp. 69–85. DOI: 10.1111/j.1467-8535.2009.01024.x.

[64] Das Hochschulforum Digitalisierung. *Digitale Transformation in der medizinischen Ausbildung, Handreichung Digitalisierung der Fachbereiche: Medizin*. Accessed: 2024-11-16. Nov. 2023. URL: https://hochschulforumdigitalisierung.de/wp-content/uploads/2023/11/HFD_AP_74_Medizin.pdf.

[65] Pierre Dillenbourg, Daniel Schneider, and Paraskevi Synteta. "Virtual Learning Environments." In: *3rd Hellenic Conference "Information & Communication Technologies in Education"*. Ed. by A. Dimitracopoulou. Rhodes, Greece: Kastaniotis Editions, Greece, 2002, pp. 3–18. URL: https://telearn.hal.science/hal-00190701.

[66] Huong Q Dinh, Neff Walker, Larry F Hodges, Chang Song, and Akira Kobayashi. "Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments." In: *Proceedings IEEE Virtual Reality (Cat. No. 99CB36316)*. IEEE. 1999, pp. 222–228.

[67] Alan Dix. "Human-Computer Interaction." In: *Encyclopedia of Database Systems*. Ed. by LING LIU and M. TAMER ÖZSU. Boston, MA: Springer US, 2009, pp. 1327–1331. ISBN: 978-0-387-39940-9. DOI: 10.1007/978-0-387-39940-9_192.

[68] Dinesh C Dobhal, Harendra Singh Negi, and Purushottam Das. "The Effect of Virtual Reality in the Modern System of Education." In: *2023 2nd International Conference for Innovation in Technology (INOCON)*. IEEE. 2023, pp. 1–4.

[69] Ralf Doerner, Wolfgang Broll, Paul Grimm, and Bernhard Jung. *Virtual and augmented reality (VR/AR): Foundations and methods of extended realities (XR)*. Springer Nature, 2022.

[70] Ralf Doerner, Wolfgang Broll, Bernhard Jung, Paul Grimm, Martin Göbel, and Rolf Kruse. "Introduction to virtual and augmented reality." In: *Virtual and Augmented Reality (VR/AR) Foundations and Methods of Extended Realities (XR)*. Springer, 2022, pp. 1–37.

[71] Kevin Doherty and Gavin Doherty. "The construal of experience in HCI: Understanding self-reports." In: *International Journal of Human-Computer Studies* 110 (2018), pp. 63–74.

[72] E. Doukakis, K. Debattista, T. Bashford-Rogers, A. Dhokia, A. Asadipour, A. Chalmers, and C. Harvey. "Audio-Visual-Olfactory Resource Allocation for Tri-

modal Virtual Environments." In: *IEEE Transactions on Visualization and Computer Graphics* 25.5 (2019), pp. 1865–1875. DOI: 10.1109/tvcg.2019.2898823.

[73] John V Draper, David B Kaber, and John M Usher. "Telepresence." In: *Human factors* 40.3 (1998), pp. 354–375.

[74] Tobias Drey. "Towards mixed reality in education: overcoming challenges of teaching, learning, and authoring." PhD thesis. Universität Ulm, 2023.

[75] Stephen R Ellis. "What are virtual environments?" In: *IEEE Computer Graphics and Applications* 14.1 (1994), pp. 17–22.

[76] Barrett Ens, Benjamin Bach, Maxime Cordeil, Ulrich Engelke, Marcos Serrano, Wesley Willett, Arnaud Prouzeau, Christoph Anthes, Wolfgang Büschel, Cody Dunne, et al. "Grand challenges in immersive analytics." In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 2021, pp. 1–17.

[77] Peggy A Ertmer and Timothy J Newby. "Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective." In: *Performance improvement quarterly* 6.4 (1993), pp. 50–72.

[78] Marta Fairén González, Mariona Farrés, Jordi Moyes Ardiaca, and Esther Insa. "Virtual Reality to teach anatomy." In: *Eurographics 2017: education papers*. European Association for Computer Graphics (Eurographics). 2017, pp. 51–58.

[79] Jannat Falah, Soheeb Khan, Tasneem Alfalah, Salsabeel F. M. Alfalah, Warren Chan, David K. Harrison, and Vassilis Charissis. "Virtual Reality medical training system for anatomy education." In: *Proc. of Science and Information Conference*. SAI '14. IEEE, 2014, pp. 752–758. DOI: 10.1109/SAI.2014.6918271.

[80] Abel A Fernandez, Sabrina Rodriguez Mora, and Ryan A Damerell. "Work-in-Progress—–Photogrammetry within Virtual Reality." In: *2021 7th International Conference of the Immersive Learning Research Network (iLRN)*. IEEE. 2021, pp. 1–2.

[81] Philip W Fink, Patrick S Foo, and William H Warren. "Obstacle avoidance during walking in real and virtual environments." In: *ACM Transactions on Applied Perception (TAP)* 4.1 (2007), 2–es.

[82] Frank Fischbach, Jürgen Bunke, Markus Thormann, Gunnar Gaffke, Kerstin Jungnickel, Jouke Smink, and Jens Ricke. "MR-guided freehand biopsy of liver lesions with fast continuous imaging using a 1.0-T open MRI scanner: experience in 50 patients." In: *Cardiovascular and interventional radiology* 34 (2011), pp. 188–192.

[83] John Fox and Sanford Weisberg. *An R Companion to Applied Regression*. Third. Thousand Oaks CA: Sage, 2019. ISBN: 9781544336473. URL: https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

[84] Julia Fröhlich and Ipke Wachsmuth. "The Visual, the Auditory and the Haptic – A User Study on Combining Modalities in Virtual Worlds." In: *Virtual Augmented and Mixed Reality. Designing and Developing Augmented and Virtual Environments*. Ed. by Randall Shumaker. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 159–168. ISBN: 978-3-642-39405-8.

[85] Tomohiro Fukuda, Hirokazu Ban, Katsuhito Yagi, and Junro Nishiie. "Development of high-definition virtual reality for historical architectural and urban digital reconstruction: A case study of Azuchi castle and old castle town in 1581." In: *Computer-Aided Architectural Design Futures. The Next City-New Technologies and the Future of the Built Environment: 16th International Conference, CAAD Futures 2015, São Paulo, Brazil, July 8-10, 2015. Selected Papers 16*. Springer. 2015, pp. 75–89.

[86] A. Fung, P. Kelly, G. Tait, P. D. Greig, and I. D. McGilvray. "Creating an animation-enhanced video library of hepato-pancreato-biliary and transplantation surgical procedures." In: *J Vis Commun Med* 39.1-2 (2016), pp. 27–32.

[87] L. Furcea, F. Graur, L. Scurtu, N. Plitea, D. Pîslă, C. Vaida, O. Deteşan, A. Szilaghy, H. Neagoş, A. Mureşan, and L. Vlad. "The advantages of implementing an e-learning platform for laparoscopic liver surgery." In: *Chirurgia (Bucur)* 106.6 (2011), pp. 799–806.

[88] Giorgos Ganias, Christos Lougiakis, Akrivi Katifori, Maria Roussou, Yannis Ioannidis, et al. "Comparing different grasping visualizations for object manipulation in vr using controllers." In: *IEEE Transactions on Visualization and Computer Graphics* 29.5 (2023), pp. 2369–2378.

[89] Giorgio Ganis and Rogier Kievit. "A New Set of Three-Dimensional Shapes for Investigating Mental Rotation Processes: Validation Data and Stimulus Set." In: *Journal of Open Psychology Data* 3 (2015). DOI: 10.5334/jopd.ai.

[90] Muriel Garreta-Domingo, Peter B Sloep, and Davinia Hernández-Leo. "Human-centred design to empower "teachers as designers"." In: *British Journal of Educational Technology* 49.6 (2018), pp. 1113–1130.

[91] Juan Garzón. "An Overview of Twenty-Five Years of Augmented Reality in Education." In: *Multimodal Technologies and Interaction* 5.7 (July 2021), p. 37. DOI: 10.3390/mti5070037.

[92] Yiannis Georgiou and Eleni A Kyza. "Relations between student motivation, immersion and learning outcomes in location-based augmented reality settings." In: *Computers in Human Behavior* 89 (2018), pp. 173–181.

[93] Dianne Gerrelli, Steven Lisgo, Andrew J Copp, and Susan Lindsay. "Enabling research with human embryonic and fetal tissue resources." In: *Development* 142.18 (2015), pp. 3073–3076. ISSN: 0950-1991. DOI: 10.1242/dev.122820.

[94] Marnix van Gisbergen, Michelle Kovacs, Fabio Campos, Malou van der Heeft, and Valerie Vugts. "What we don't know. the effect of realism in virtual reality on experience and behaviour." In: *Augmented Reality and Virtual Reality: The Power of AR and VR for Business* (2019), pp. 45–57.

[95] Guilherme Gonçalves, Hugo Coelho, Pedro Monteiro, Miguel Melo, and Maximino Bessa. "Systematic Review of Comparative Studies of the Impact of Realism in Immersive Virtual Experiences." In: *ACM Computing Surveys* 55.6 (2022), pp. 1–36. DOI: 10.1145/3533377.

[96] Charles R Graham. "Blended learning systems." In: *The handbook of blended learning: Global perspectives, local designs* 1 (2006), pp. 3–21.

[97] Simon Greenwold. "Spatial computing." In: *Massachusetts Institute of Technology, Master* (2003).

[98] Avinash Gupta, J Cecil, and Miguel Pirela-Cruz. "Immersive virtual reality based training and assessment of an orthopedic surgical process." In: *2020 IEEE International Systems Conference (SysCon)*. IEEE. 2020, pp. 1–7.

[99] Katherine S Gustilo, Jennifer Stratford, Marissa Hight, and Lisa M Lee. "3D Embryonic Heart Goes Mobile: Efficacy of Mobile Application in Congenital Cardiac Embryology Education." In: *The FASEB Journal* 36.S1 (2022). DOI: 10.1096/fasebj.2022.36.S1.R5399.

[100] Antal Haans, Wijnand A IJsselsteijn, and Yvonne AW de Kort. "The effect of similarities in skin texture and hand shape on perceived ownership of a fake limb." In: *Body Image* 5.4 (2008), pp. 389–394.

[101] Matthew Hackett and Michael Proctor. "Three-dimensional display technologies for anatomical education: a literature Review." In: *Journal of Science Education and Technology* 25.4 (2016), pp. 641–654.

[102] David Hamilton, Jim McKechnie, Edward Edgerton, and Claire Wilson. "Immersive virtual reality as a pedagogical tool in education: a systematic literature review of quantitative learning outcomes and experimental design." In: *Journal of Computers in Education* 8.1 (2021), pp. 1–32.

[103] Joel Harman, Ross Brown, Daniel Johnson, and Selen Turkay. "The Role of Visual Detail during Situated Memory Recall within a Virtual Reality Environment." In: *Proceedings of the 31st Australian Conference on Human-Computer-Interaction*. ACM, 2019. DOI: 10.1145/3369457.3369469.

[104] Sandra G. Hart. "Nasa-Task Load Index (NASA-TLX); 20 Years Later." In: *Proc. of the Human Factors and Ergonomics Society Annual Meeting* 50.9 (2006), pp. 904–908. DOI: 10.1177/154193120605000909.

[105] Sandra G. Hart and Lowell E. Staveland. "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research." In: *Human Mental Workload*. Ed. by Peter A. Hancock and Najmedin Meshkati. North-Holland, 1988, pp. 139–183. DOI: 10.1016/S0166-4115(08)62386-9.

[106] Jan Hendrik Plümer, Kevin Yu, Ulrich Eck, Denis Kalkofen, Philipp Steininger, Nassir Navab, and Markus Tatzgern. "XR Prototyping of Mixed Reality Visualizations: Compensating Interaction Latency for a Medical Imaging Robot." In: *arXiv e-prints* (2024), arXiv–2409.

[107] Herbert Hoijtink, Irene Klugkist, and Paul A. Boelen, eds. *Bayesian Evaluation of Informative Hypotheses*. Springer New York, 2008. DOI: 10.1007/978-0-387-09612-4.

[108] Matt C. Howard. "Investigating the simulation elements of environment and control: Extending the Uncanny Valley Theory to simulations." In: *Computers & Educa-*

tion 109 (2017), pp. 216–232. DOI: 10.1016/j.compedu.2017.03.005.

[109] Hsiu-Mei Huang, Ulrich Rauch, and Shu-Sheng Liaw. "Investigating learners' attitudes toward virtual reality learning environments: Based on a constructivist approach." In: *Computers & Education* 55.3 (2010), pp. 1171–1182. DOI: 10.1016/j.compedu.2010.05.014.

[110] Brian Hudson. "Holding complexity and searching for meaning: Teaching as reflective practice." In: *Journal of Curriculum Studies* 34.1 (2002), pp. 43–57.

[111] Andrew Hull, Ernesto Salcedo, Jennifer Stratford, Marissa Hight, Hailey Bunker, and Lisa Lee. "Embryorigami: The Educational Impact of 4D Virtual Folding Embryo Model on Student Learning and Confidence." In: *The FASEB Journal* 35.S1 (2021). DOI: 10.1096/fasebj.2021.35.S1.04542.

[112] Jonatan Hvass, Oliver Larsen, Kasper Vendelbo, Niels Nilsson, Rolf Nordahl, and Stefania Serafin. "Visual realism and presence in a virtual reality game." In: *2017 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*. IEEE, 2017. DOI: 10.1109/3dtv.2017.8280421.

[113] Wijnand A IJsselsteijn, Yvonne AW De Kort, and Karolien Poels. "The game experience questionnaire." In: (2013).

[114] Horace HS Ip and Chen Li. "Virtual reality-based learning environments: Recent developments and ongoing challenges." In: *Hybrid Learning: Innovation in Educational Practices: 8th International Conference, ICHL 2015, Wuhan, China, July 27-29, 2015, Proceedings 8*. Springer. 2015, pp. 3–14.

[115] *ISO 9241-11:2018: Ergonomics of human-system interaction - Part 11: Usability: Definitions and concepts*. Geneva, Switzerland: International Organization for Standardization, 2018.

[116] *ISO 9241-210:2019: Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems*. Geneva, Switzerland: International Organization for Standardization, 2019.

[117] Julie A Jacko. "Human computer interaction handbook: Fundamentals, evolving technologies, and emerging applications." In: (2012). DOI: 10.1201/b11963.

[118] Randolph L Jackson and Eileen Fagan. "Collaboration and learning within immersive virtual reality." In: *Proc. of Collaborative virtual environments*. ACM. 2000, pp. 83–92.

[119] Randolph L Jackson, Wayne Taylor, and William Winn. "Peer collaboration and virtual environments: a preliminary investigation of multi-participant virtual reality applied in science education." In: *Proc. of ACM symposium on Applied computing*. 1999, pp. 121–125.

[120] K. H. James, G. K. Humphrey, T. Vilis, B. Corrie, R. Baddour, and M. A. Goodale. ""Active" and "passive" learning of three-dimensional object structure within an immersive virtual reality environment." In: *Behavior Research Methods, Instruments, & Computers* 34.3 (2002), pp. 383–390. ISSN: 1532-5970. DOI: 10.3758/BF03195466.

[121] Susan Jang, Jonathan M. Vitale, Robert W. Jyung, and John B. Black. "Direct manipulation is better than passive viewing for learning anatomy in a three-dimensional virtual reality environment." In: *Computers & Education* 106 (2017), pp. 150–165. ISSN: 0360-1315. DOI: 10.1016/j.compedu.2016.12.009.

[122] Daniela Janßen, Christian Tummel, Anja Richert, and Ingrid Isenhardt. "Towards measuring user experience, activation and task performance in immersive virtual learning environments for students." In: *Immersive Learning Research Network: Second International Conference, iLRN 2016 Santa Barbara, CA, USA, June 27–July 1, 2016 Proceedings 2*. Springer. 2016, pp. 45–58.

[123] Hans-Christian Jetter, Jan-Henrik Schröder, Jan Gugenheimer, Mark Billinghurst, Christoph Anthes, Mohamed Khamis, and Tiare Feuchtner. "Transitional interfaces in mixed and cross-reality: A new frontier?" In: *Companion Proceedings of the 2021 Conference on Interactive Surfaces and Spaces*. 2021, pp. 46–49.

[124] Qiao Jin, Yu Liu, Svetlana Yarosh, Bo Han, and Feng Qian. "How Will VR Enter University Classrooms? Multi-Stakeholders Investigation of VR in Higher Education." In: *Proc. of CHI Conference on Human Factors in Computing Systems*. CHI '22. New Orleans, LA, USA: ACM, 2022. ISBN: 9781450391573. DOI: 10.1145/3491102.3517542.

[125] E. O. Johnson, A. V. Charchanti, and T. G. Troupis. "Modernization of an anatomy class: From conceptualization to implementation. A case for integrated

multimodal-multidisciplinary teaching." In: *Anat Sci Educ* 5.6 (2012), pp. 354–366.

[126] Seth Johnson, Daniel Orban, Hakizumwami Birali Runesha, Lingyu Meng, Bethany Juhnke, Arthur Erdman, Francesca Samsel, and Daniel F Keefe. "Bento box: An interactive and zoomable small multiples technique for visualizing 4d simulation ensembles in virtual reality." In: *Front. Robot. AI 6: 61. doi: 10.3389/frobt* (2019).

[127] Jinki Jung and Young Joong Ahn. "Effects of interface on procedural skill transfer in virtual training: Lifeboat launching operation study." In: *Computer Animation and Virtual Worlds* 29.3-4 (2018), e1812. DOI: 10.1002/cav.1812.

[128] Sungchul Jung and Robert W Lindeman. "Perspective: Does realism improve presence in vr? suggesting a model and metric for vr experience evaluation." In: *Frontiers in Virtual Reality* 2 (2021), p. 693327.

[129] Sungchul Jung, Christian Sandor, Pamela J Wisniewski, and Charles E Hughes. "Realme: The influence of body and hand representations on body ownership and presence." In: *Proceedings of the 5th Symposium on Spatial User Interaction*. 2017, pp. 3–11.

[130] Sungchul Jung, Andrew L. Wood, Simon Hoermann, Pramuditha L. Abhayawardhana, and Robert W. Lindeman. "The Impact of Multi-sensory Stimuli on Confidence Levels for Perceptual-cognitive Tasks in VR." In: *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2020. DOI: 10.1109/vr46266.2020.00067.

[131] Mohammed Kadri, Fatima-Ezzahra Boubakri, Timothy Teo, Fatima-Zahra Kaghat, Ahmed Azough, and Khalid Alaoui Zidani. "Virtual reality in medical education: Effectiveness of Immersive Virtual Anatomy Laboratory (IVAL) Compared to traditional learning approaches." In: *Displays* (2024), p. 102870.

[132] Dorota Kamińska, Tomasz Sapiński, Sławomir Wiak, Toomas Tikk, Rain Eric Haamer, Egils Avots, Ahmed Helmi, Cagri Ozcinar, and Gholamreza Anbarjafari. "Virtual Reality and Its Applications in Education: Survey." In: *Information* 10.10 (2019). ISSN: 2078-2489. DOI: 10.3390/info10100318.

[133] Elizabeth Kappler, Rosemarie Figueroa Jacinto, and Steve Arndt. "Evaluation of Visual Acuity and Perceptual Field of View Using the Varjo XR-3 Headset in a

Virtual Environment." In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 66.1 (2022), pp. 2193–2197. DOI: 10.1177/1071181322661195.

[134] Melina Klepsch, Florian Schmitz, and Tina Seufert. "Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load." In: *Frontiers in psychology* 8 (2017), p. 1997.

[135] Thomas Kollewe, Monika Sennekamp, and Falk Ochsendorf. *Medizindidaktik*. Springer, 2018.

[136] Karel Kreijns, Kate Xu, and Joshua Weidlich. "Social presence: Conceptualization and measurement." In: *Educational Psychology Review* 34.1 (2022), pp. 139–170.

[138] Ramazan Kurul, Muhammed Nur Ögün, Ayşe Neriman Narin, Şebnem Avci, and Beyza Yazgan. "An Alternative Method for Anatomy Training: Immersive Virtual Reality." In: *Anatomical Sciences Education* 13 (2020). DOI: 10.1002/ase.1959.

[140] Sheldon J. Lachman. "Learning is a Process: Toward an Improved Definition of Learning." In: *The Journal of Psychology* 131.5 (1997), pp. 477–480. DOI: 10.1080/00223989709603535. eprint: https://doi.org/10.1080/00223989709603535.

[141] Jérémy Lacoche, Morgan Le Chénéchal, Sébastien Chalmé, Jérôme Royan, Thierry Duval, Valérie Gouranton, Éric Maisel, and Bruno Arnaldi. "Dealing with frame cancellation for stereoscopic displays in 3d user interfaces." In: *2015 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE. 2015, pp. 73–80.

[142] Marc Erich Latoschik, Florian Kern, Jan-Philipp Stauffert, Andrea Bartl, Mario Botsch, and Jean-Luc Lugrin. "Not Alone Here?! Scalability and User Experience of Embodied Ambient Crowds in Distributed Social Virtual Reality." In: *IEEE Transactions on Visualization and Computer Graphics* 25.5 (2019), pp. 2134–2144. DOI: 10.1109/tvcg.2019.2899250.

[143] Marc Erich Latoschik, Daniel Roth, Dominik Gall, Jascha Achenbach, Thomas Waltemate, and Mario Botsch. "The effect of avatar realism in immersive social virtual realities." In: *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*. ACM, 2017. DOI: 10.1145/3139131.3139156.

[144] Marc Erich Latoschik and Carolin Wienrich. "Congruence and Plausibility, Not Presence: Pivotal Conditions for XR Experiences and Effects, a Novel Approach." In: *Frontiers in Virtual Reality* 3 (2022). DOI: 10.3389/frvir.2022.694433.

[145] Bettina Laugwitz, Theo Held, and Martin Schrepp. "Construction and evaluation of a user experience questionnaire." In: *HCI and Usability for Education and Work: 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society, USAB 2008, Graz, Austria, November 20-21, 2008. Proceedings 4*. Springer. 2008, pp. 63–76.

[146] Diana Laurillard and Michael Derntl. "Learner Centred Design - Overview." In: *Practical Design Patterns for Teaching and Learning with Technology*. Ed. by Yishay Mor, Harvey Mellar, Steven Warburton, and Niall Winters. Rotterdam: SensePublishers, 2014, pp. 13–16. ISBN: 978-94-6209-530-4. DOI: 10.1007/978-94-6209-530-4_2.

[147] Joseph J LaViola Jr, Ernst Kruijff, Ryan P McMahan, Doug Bowman, and Ivan P Poupyrev. *3D user interfaces: theory and practice*. Addison-Wesley Professional, 2017.

[148] Elinda Ai-Lim Lee, Kok Wai Wong, and Chun Che Fung. "How does desktop virtual reality enhance learning outcomes? A structural equation modeling approach." In: *Computers & Education* 55.4 (2010), pp. 1424–1442.

[149] Elinda Ai-Lim Lee, Kok Wai Wong, and Chun Che Fung. "Learning with Virtual Reality: Its Effects on Students with Different Learning Styles." In: *Transactions on Edutainment IV*. Ed. by Zhigeng Pan, Adrian David Cheok, Wolfgang Müller, Xiaopeng Zhang, and Kevin Wong. Berlin, Heidelberg: Springer, 2010, pp. 79–90. ISBN: 978-3-642-14484-4. DOI: 10.1007/978-3-642-14484-4_8.

[150] Kwan Min Lee. "Presence, explicated." In: *Communication theory* 14.1 (2004), pp. 27–50.

[151] Kwan Min Lee, Younbo Jung, Jaywoo Kim, and Sang Ryong Kim. "Are physically embodied social agents better than disembodied social agents?: The effects of physical embodiment, tactile interaction, and people's loneliness in human–robot interaction." In: *International journal of human-computer studies* 64.10 (2006), pp. 962–973.

[152] Michael D. Lee and Eric-Jan Wagenmakers. *Bayesian Cognitive Modeling*. Cambridge University Press, 2014. DOI: 10.1017/cbo9781139087759.

[153] Jimmie Leppink, Fred Paas, Cees PM Van der Vleuten, Tamara Van Gog, and Jeroen JG Van Merriënboer. "Development of an instrument for measuring different types of cognitive load." In: *Behavior research methods* 45 (2013), pp. 1058–1072.

[154] James R Lewis. "The system usability scale: past, present, and future." In: *International Journal of Human–Computer Interaction* 34.7 (2018), pp. 577–590.

[155] James R Lewis and Jeff Sauro. "Item benchmarks for the system usability scale." In: *Journal of Usability Studies* 13.3 (2018).

[156] Robb Lindgren and Mina Johnson-Glenberg. "Emboldened by Embodiment: Six Precepts for Research on Embodied Learning and Mixed Reality." In: *Educational Researcher* 42.8 (2013), pp. 445–452. DOI: 10.3102/0013189X13511661.

[157] M Lombard. "At the heart of it all: The concept of presence." In: *Journal of Computer-Mediated Communication* 3.2 (1997).

[158] Jean-Luc Lugrin, Johanna Latt, and Marc Erich Latoschik. *Anthropomorphism and Illusion of Virtual Body Ownership*. 2015. DOI: 10.2312/EGVE.20151303.

[159] Tiffany Luong, Yi Fei Cheng, Max Mobus, Andreas Fender, and Christian Holz. " Controllers or Bare Hands? A Controlled Evaluation of Input Techniques on Interaction Performance and Exertion in Virtual Reality." In: *IEEE Transactions on Visualization & Computer Graphics* 29.11 (Nov. 2023), pp. 4633–4643. ISSN: 1941-0506. DOI: 10.1109/TVCG.2023.3320211.

[160] Edris AF Mahtab and Anastasia D Egorova. "Current and future applications of virtual reality technology for cardiac interventions." In: *Nature Reviews Cardiology* 19.12 (2022), pp. 779–780.

[161] Patrick Mair and Rand Wilcox. "Robust Statistical Methods in R Using the WRS2 Package." In: *Behavior Research Methods* 52 (2020), pp. 464–488.

[162] Guido Makransky, Thomas S Terkildsen, and Richard E Mayer. "Adding immersive virtual reality to a science lab simulation causes more presence but less learning." In: *Learning and instruction* 60 (2019), pp. 225–236.

[163] Sharmistha Mandal. "Brief introduction of virtual reality & its challenges." In: *International Journal of Scientific & Engineering Research* 4.4 (2013), pp. 304–309.

[164] Steve Mann. “Mediated Reality.” In: *Linux J.* 1999.59es (Mar. 1999), 5–es. ISSN: 1075-3583.

[165] Steve Mann, Tom Furness, Yu Yuan, Jay Iorio, and Zixin Wang. “All reality: Virtual, augmented, mixed (x), mediated (x, y), and multimediated reality.” In: *arXiv preprint arXiv:1804.08386* (2018).

[166] H. S. Maresky, A. Oikonomou, I. Ali, N. Ditkofsky, M. Pakkal, and B. Ballyk. “Virtual reality and cardiac anatomy: Exploring immersive three-dimensional cardiac imaging, a pilot study in undergraduate medical anatomy education.” In: *Clinical Anatomy* 32.2 (2019), pp. 238–243. DOI: 10.1002/ca.23292.

[167] Kazushi Maruya and Tomoko Ohtani. “Shadows can change the shape appearances of real and virtual objects.” In: *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE. 2019, pp. 1076–1077.

[168] Antonella Maselli and Mel Slater. “The building blocks of the full body ownership illusion.” In: *Frontiers in Human Neuroscience* 7 (2013). DOI: 10.3389/fnhum.2013.00083.

[169] Frank Maurer, Craig Anslow, Joaquim Jorge, and Mauricio Sousa. “Enhancing cross-reality applications and user experiences.” In: *Proceedings of the 2022 International Conference on Advanced Visual Interfaces*. AVI ’22. Frascati, Rome, Italy: Association for Computing Machinery, 2022. ISBN: 9781450397193. DOI: 10.1145/3531073.3535256.

[170] Richard E Mayer, Guido Makransky, and Jocelyn Parong. “The promise and pitfalls of learning in immersive virtual reality.” In: *International Journal of Human–Computer Interaction* 39.11 (2023), pp. 2229–2238.

[171] R. P. McMahan, D. A. Bowman, D. J. Zielinski, and R. B. Brady. “Evaluating Display Fidelity and Interaction Fidelity in a Virtual Reality Game.” In: *IEEE Transactions on Visualization and Computer Graphics* 18.4 (2012), pp. 626–633. DOI: 10.1109/tvcg.2012.43.

[172] Ryan P. McMahan, Chengyuan Lai, and Swaroop K. Pal. “Interaction Fidelity: The Uncanny Valley of Virtual Reality Interactions.” In: *Lecture Notes in Computer Science*. Springer International Publishing, 2016, pp. 59–70. DOI: 10.1007/978-3-319-39907-2_6.

[173] Pratham Darrpan Mehta, Harsha Karanth, Haoyang Yang, Timothy C Slesnick, Fawwaz Shaw, and Duen Horng Chau. "ARCollab: Towards Multi-User Interactive Cardiovascular Surgical Planning in Mobile Augmented Reality." In: *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 2024, pp. 1–6.

[174] Miguel Melo, Guilherme Gonçalves, Pedro Monteiro, Hugo Coelho, José Vasconcelos-Raposo, and Maximino Bessa. "Do Multisensory Stimuli Benefit the Virtual Reality Experience? A Systematic Review." In: *IEEE Transactions on Visualization and Computer Graphics* 28.2 (2022), pp. 1428–1442. DOI: 10.1109/TVCG.2020.3010088.

[175] E. Messier, J. Wilcox, A. Dawson-Elli, G. Diaz, and C. A. Linte. "An Interactive 3D Virtual Anatomy Puzzle for Learning and Simulation - Initial Demonstration and Evaluation." In: *Stud Health Technol Inform* 220 (2016), pp. 233–240.

[176] Ines Miguel-Alonso, David Checa, Henar Guillen-Sanz, and Andres Bustillo. "Evaluation of the novelty effect in immersive Virtual Reality learning experiences." In: *Virtual Reality* 28.1 (2024), p. 27.

[177] Paul Milgram and Fumio Kishino. "A taxonomy of mixed reality visual displays." In: *IEICE TRANSACTIONS on Information and Systems* 77.12 (1994), pp. 1321–1329.

[178] Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. "Augmented reality: A class of displays on the reality-virtuality continuum." In: *Telemanipulator and telepresence technologies*. Vol. 2351. Spie. 1995, pp. 282–292.

[179] Natalie B Milman. "Defining and conceptualizing mixed reality, augmented reality, and virtual reality." In: *Distance Learning* 15.2 (2018), pp. 55–58.

[180] Xin Min, Wenqiao Zhang, Shouqian Sun, Nan Zhao, Siliang Tang, and Yueting Zhuang. "VPModel: High-Fidelity Product Simulation in a Virtual-Physical Environment." In: *IEEE Transactions on Visualization and Computer Graphics* 25.11 (2019), pp. 3083–3093. DOI: 10.1109/tvcg.2019.2932276.

[181] Takato Mizuho, Takuji Narumi, and Hideaki Kuzuoka. "Effects of the Visual Fidelity of Virtual Environments on Presence, Context-dependent Forgetting, and Source-

monitoring Error." In: *IEEE Transactions on Visualization and Computer Graphics* 29.5 (2023), pp. 2607–2614.

[182] Tom Moher, Andrew Johnson, Stellan Ohlsson, and Mark Gillingham. "Bridging strategies for VR-based learning." In: *Proc. of ACM SIGCHI conference on Human Factors in Computing Systems*. 1999, pp. 536–543.

[183] J. Mönch, K. Mühler, C. Hansen, K. J. Oldhafer, G. Stavrou, C. Hillert, C. Logge, and B. Preim. "The LiverSurgeryTrainer: training of computer-based planning in liver resection surgery." In: *Int J Comput Assist Radiol Surg* 8.5 (Sept. 2013), pp. 809–818.

[184] Stephen J Moorman. "Prof-in-a-Box: using internet-videoconferencing to assist students in the gross anatomy laboratory." In: *BMC medical education* 6.1 (2006), pp. 1–7.

[185] Suzana Guimar Moraes and Luis Antonio Violin Pereira. "A multimedia approach for teaching human embryology: Development and evaluation of a methodology." In: *Annals of Anatomy - Anatomischer Anzeiger* 192.6 (2010), pp. 388–395. ISSN: 0940-9602. DOI: 10.1016/j.aanat.2010.05.005.

[186] Masahiro Mori, Karl F. MacDorman, and Norri Kageki. "The Uncanny Valley [From the Field]." In: *IEEE Robotics & Automation Magazine* 19.2 (2012), pp. 98–100. DOI: 10.1109/MRA.2012.2192811.

[187] Jesper Mortensen, Insu Yu, Pankaj Khanna, Franco Tecchia, Bernhard Spanlang, Giuseppe Marino, and Mel Slater. "Real-time global illumination for vr applications." In: *IEEE Computer Graphics and Applications* 28.6 (2008), pp. 56–64.

[188] Stylianos Mystakidis. "Metaverse." In: *Encyclopedia* 2.1 (2022), pp. 486–497.

[189] I Nakarada-Kordic, S Reay, G Bennett, J Kruse, A-M Lydon, and J Sim. "Can virtual reality simulation prepare patients for an MRI experience?" In: *Radiography* 26.3 (2020), pp. 205–213.

[190] Kim JL Nevelsteen. "Virtual world, defined from a technological perspective and applied to video games, mixed reality, and the Metaverse." In: *Computer animation and virtual worlds* 29.1 (2018), e1752.

[191] M. Newman, B. Gatersleben, K.J. Wyles, and E. Ratcliffe. "The use of virtual reality in environment experiences and the importance of realism." In: *Journal of Environmental Psychology* 79 (2022), p. 101733. ISSN: 0272-4944. DOI: 10.1016/j.jenvp.2021.101733.

[192] D. Nobuoka, T. Fuji, K. Yoshida, K. Takagi, T. Kuise, M. Utsumi, R. Yoshida, Y. Umeda, S. Shinoura, Y. Takeda, and A. Ohtsuka. "Surgical education using a multi-viewpoint and multi-layer three-dimensional atlas of surgical anatomy (with video)." In: *J Hepatobiliary Pancreat Sci* 21.8 (2014), pp. 556–561.

[193] Steffen Oeltze-Jafra and Bernhard Preim. "Survey of labeling techniques in medical visualizations." In: *Proceedings of the 4th Eurographics Workshop on Visual Computing for Biology and Medicine*. 2014, pp. 199–208.

[194] Catherine S Oh, Jeremy N Bailenson, and Gregory F Welch. "A systematic review of social presence: Definition, antecedents, and implications." In: *Frontiers in Robotics and AI* 5 (2018), p. 409295.

[195] Jaisa Olasky, Ganesh Sankaranarayanan, Neal E Seymour, J Harvey Magee, Andinet Enquobahrie, Ming C Lin, Rajesh Aggarwal, L Michael Brunt, Steven D Schwaitzberg, Caroline GL Cao, et al. "Identifying opportunities for virtual reality simulation in surgical education: a review of the proceedings from the innovation, design, and emerging alliances in surgery (IDEAS) conference: VR surgery." In: *Surgical innovation* 22.5 (2015), pp. 514–521.

[196] Y. Oshiro, H. Yano, J. Mitani, S. Kim, J. Kim, K. Fukunaga, and N. Ohkohchi. "Novel 3-dimensional virtual hepatectomy simulation combined with real-time deformation." In: *World J Gastroenterol* 21.34 (2015), pp. 9982–9992.

[197] Joshua Owolabi and Abebe Bekele. "Implementation of innovative educational technologies in teaching of anatomy and basic medical sciences during the COVID-19 pandemic in a developing country: the COVID-19 silver lining?" In: *Advances in medical education and practice* (2021), pp. 619–625.

[198] J. Pan, L. Zhang, P. Yu, Y. Shen, H. Wang, H. Hao, and H. Qin. "Real-time VR Simulation of Laparoscopic Cholecystectomy based on Parallel Position-based Dynamics in GPU." In: *IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 2020, pp. 548–556.

[199] Enrico Pannicke. "Empfangsspulen für bildgeführte Eingriffe mittels Magnetresonanztomographie." PhD thesis. Otto-von-Guericke-Universität Magdeburg, Magdeburg, 2021. DOI: 10.25673/39038.

[200] Indrajeet Patil. "Visualizations with statistical details: The 'ggstatsplot' approach." In: *Journal of Open Source Software* 6.61 (2021), p. 3167. DOI: 10.21105/joss.03167.

[201] Manjula Patil. "Innovative Method of Teaching Embryology Using Three-Dimensional Playdough Model Construction: A Constructivist Teaching." In: *National Journal of Clinical Anatomy* 9 (Jan. 2020), p. 135. DOI: 10.4103/NJCA.NJCA_67_20.

[202] Vaidehi Patil, Jyotindra Narayan, Kamalpreet Sandhu, and Santosha K Dwivedy. "Integration of virtual reality and augmented reality in physical rehabilitation: a state-of-the-art review." In: *Revolutions in Product Design for Healthcare: Advances in Product Design and Design Methods for Healthcare* (2022), pp. 177–205.

[203] Apurba Patra, Adil Asghar, Priti Chaudhary, and Kumar Satish Ravi. "Integration of innovative educational technologies in anatomy teaching: new normal in anatomy education." In: *Surgical and Radiologic Anatomy* 44.1 (2022), pp. 25–32.

[204] Shiva Pedram, Grace Kennedy, and Sal Sanzone. "Toward the validation of VR-HMDs for medical education: a systematic literature review." In: *Virtual Reality* 27.3 (2023), pp. 2255–2280.

[205] Shiva Pedram, Grace Kennedy, and Sal Sanzone. "Assessing the validity of VR as a training tool for medical students." In: *Virtual Reality* 28.1 (2024), p. 15.

[206] Effi Freya Picka, Annika Vogel, Marie-Sophie Roder, Jonas Birkle, Daniela Schrenk, Jessica Linnemann, Julia Moritz, and Stefan Pfeffer. "Virtual Usability Testing (VIRTUSE)-Development of a Methodical Approach for Usability Testing in VR." In: *HCI International 2022 Posters: 24th International Conference on Human-Computer Interaction, HCII 2022, Virtual Event, June 26–July 1, 2022, Proceedings, Part I*. Springer. 2022, pp. 109–116.

[207] Sandra Poeschl and Nicola Doering. "Measuring co-presence and social presence in virtual environments–psychometric construction of a german scale for a fear of public speaking scenario." In: *Annual Review of Cybertherapy and Telemedicine 2015* (2015), pp. 58–63.

[208] Sandra Poeschl, Konstantin Wall, and Nicola Doering. "Integration of spatial sound in immersive virtual environments an experimental study on effects of spatial sound on presence." In: *2013 IEEE Virtual Reality (VR)*. IEEE. 2013, pp. 129–130.

[209] Daniel Pohlandt, Bernhard Preim, and Patrick Saalfeld. "Supporting Anatomy Education with a 3D Puzzle in a Virtual Reality Environment – Results from a Pilot Study." In: *Mensch und Computer*. 2019, pp. 91–102.

[210] Vahid Pooryousef, Maxime Cordeil, Lonni Besançon, Christophe Hurter, Tim Dwyer, and Richard Bassed. "Working with forensic practitioners to understand the opportunities and challenges for mixed-reality digital autopsy." In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 2023, pp. 1–15.

[211] Jack Pottle. "Virtual reality and the transformation of medical education." In: *Future healthcare journal* 6.3 (2019), p. 181.

[212] Bernhard Preim, Monique Meuschke, and Veronika Weiß. "A Survey of Medical Visualization Through the Lens of Metaphors." In: *IEEE Transactions on Visualization and Computer Graphics* 30.10 (2024), pp. 6639–6664. DOI: 10.1109/TVCG.2023.3330546.

[213] Bernhard Preim and Patrick Saalfeld. "A survey of virtual human anatomy education systems." In: *Computers & Graphics* 71 (2018), pp. 132–153. ISSN: 0097-8493. DOI: 10.1016/j.cag.2018.01.005.

[214] Bernhard Preim, Patrick Saalfeld, and Christian Hansen. "Virtual and Augmented Reality for Educational Anatomy." In: *Digital Anatomy : Applications of Virtual, Mixed and Augmented Reality*. Ed. by Jean-François Uhl, Joaquim Jorge, Daniel Simões Lopes, and Pedro F Campos. Cham: Springer, 2021, pp. 299–324. ISBN: 978-3-030-61905-3. DOI: 10.1007/978-3-030-61905-3_16.

[215] Jia Qu, Albert Fung, Paul Kelly, Gordon Tait, Paul D. Greig, Anne Agur, Ian D. McGilvray, and Jodie Jenkinson. "Visualising a rare and complex case of advanced hilar cholangiocarcinoma." In: *Journal of Visual Communication in Medicine* 40.1 (2017), pp. 26–31.

[216] Ed de Quincey, Chris Briggs, Theocharis Kyriacou, and Richard Waller. "Student centred design of a learning analytics system." In: *Proceedings of the 9th international conference on learning analytics & knowledge*. 2019, pp. 353–362.

[217] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2022. URL: https://www.R-project.org/.

[218] Eric D. Ragan, Doug A. Bowman, Regis Kopper, Cheryl Stinson, Siroberto Scerbo, and Ryan P. McMahan. "Effects of Field of View and Visual Complexity on Virtual Reality Training Effectiveness for a Visual Scanning Task." In: *IEEE Transactions on Visualization and Computer Graphics* 21.7 (2015), pp. 794–807. DOI: 10.1109/tvcg.2015.2403312.

[219] N. Rashidian, M. Vierstraete, A. Alseidi, R. I. Troisi, and W. Willaert. "Surgical education interventions in liver surgery: a systematic review." In: *Updates Surg* 72.3 (2020), pp. 583–594.

[220] Rabindra Ayyan Ratan and Béatrice Hasler. "Self-presence standardized: Introducing the self-presence questionnaire (SPQ)." In: *Proceedings of the 12th annual international workshop on presence*. Vol. 81. 2009.

[221] Philipp A Rauschnabel, Reto Felix, Chris Hinsch, Hamza Shahab, and Florian Alt. "What is XR? Towards a framework for augmented and virtual reality." In: *Computers in human behavior* 133 (2022), p. 107289.

[222] Francisco Rebelo, Paulo Noriega, Emília Duarte, and Marcelo Soares. "Using virtual reality to assess user experience." In: *Human factors* 54.6 (2012), pp. 964–982.

[223] Manuel Rebol, Krzysztof Pietroszek, Neal Sikka, Claudia Ranniger, Colton Hood, Adam Rutenberg, Puja Sasankan, and Christian Gütl. "Evaluating Augmented Reality Communication: How Can We Teach Procedural Skill in AR?" In: *Proceedings of the 29th ACM Symposium on Virtual Reality Software and Technology*. 2023, pp. 1–11.

[224] Holger Regenbrecht and Thomas Schubert. "Real and Illusory Interactions Enhance Presence in Virtual Environments." In: *Presence: Teleoperators and Virtual Environments* 11.4 (2002), pp. 425–434. DOI: 10.1162/105474602760204318.

[225] Charles M Reigeluth. *Instructional design theories and models: An overview of their current status*. Routledge, 1983.

[226] Anke Verena Reinschluessel, Joern Teuber, Marc Herrlich, Jeffrey Bissel, Melanie van Eikeren, Johannes Ganser, et al. "Virtual reality for user-centered design and evaluation of touch-free interaction techniques for

navigating medical images in the operating room." In: *Ext Abstr Hum Factors Computing Syst*. 2017, pp. 2001–2009.

[227] Maximilian F Reiser, Wolfhard Semmler, and Hedvig Hricak. *Magnetic resonance tomography*. Springer Science & Business Media, 2007.

[228] Bernhard Reitinger, Alexander Bornik, Reinhard Beichel, and Dieter Schmalstieg. "Liver surgery planning using virtual reality." In: *IEEE Computer Graphics and Applications* 26.6 (2006), pp. 36–47.

[229] J Reyes, J Santana, Agustín Trujillo-Pino, M Maynar, and M Rodriguez. "Virtual Reality for Studying Cardiovascular Anatomy by Assembling Elements Concept and Components." In: *Proc. of Eurographics Workshop on Visual Computing for Biology and Medicine*. VCBM '21. Paris, 2021. DOI: 10.13140/RG.2.2.28432.48645.

[230] A. Richardson, M. Hazzard, S. D. Challman, A. M. Morgenstein, and J. K. Brueckner. "A "Second Life" for gross anatomy: applications for multiuser virtual environments in teaching the anatomical sciences." In: *Anat Sci Educ* 4.1 (2011), pp. 39–43.

[231] Robert Riener, Matthias Harders, Robert Riener, and Matthias Harders. *Introduction to virtual reality in medicine*. Springer, 2012.

[232] Gabriela Ripka, Jennifer Tiede, Silke Grafe, and Marc Erich Latoschik. "Teaching and Learning Processes in Immersive VR – Comparing Expectations of Preservice Teachers and Teacher Educators." In: *Society for Information Technology & Teacher Education (SITE) International Conference*. 2020.

[233] KA Ritter III and Terrence L Chambers. "Three-dimensional modeled environments versus 360 degree panoramas for mobile virtual reality training." In: *Virtual Reality* 26.2 (2022), pp. 571–581.

[234] Katja Rogers, Jana Funke, Julian Frommel, Sven Stamm, and Michael Weber. "Exploring interaction fidelity in virtual reality: Object manipulation and whole-body movements." In: *Proceedings of the 2019 CHI conference on human factors in computing systems*. 2019, pp. 1–14.

[235] Yvonne Rogers, Helen Sharp, and Jenny Preece. *Interaction Design: Beyond Human-Computer Interaction*. 6th. John Wiley & Sons, 2023. URL: http://id-book.com.

[236] Mario A Rojas-Sánchez, Pedro R Palos-Sánchez, and José A Folgado-Fernández. "Systematic literature review and bibliometric analysis on virtual reality and education." In: *Education and Information Technologies* 28.1 (2023), pp. 155–192.

[237] Eva Rothgang, Wesley D Gilson, Frank Wacker, Joachim Hornegger, Christine H Lorenz, and Clifford R Weiss. "Rapid freehand MR-guided percutaneous needle interventions: An image-based approach to improve workflow and feasibility." In: *Journal of Magnetic Resonance Imaging* 37.5 (2013), pp. 1202–1212.

[238] P. Saalfeld, S. Oeltze-Jafra, S. Saalfeld, U. Preim, O. Beuing, and B. Preim. "Sketching and Annotating Vascular Structures to Support Medical Teaching, Treatment Planning and Patient Education." In: *Proc. of Eurographics: Dirk Bartz Prize for Visual Computing in Medicine*. EG '17. Lyon, France: Eurographics, 2017, pp. 5–8. DOI: 10.2312/egm.20171043.

[239] P. Saalfeld, A. Stojnic, B. Preim, and S. Oeltze-Jafra. "Semi-Immersive 3D Sketching of Vascular Structures for Medical Education." In: *Proc. of Eurographics Workshop on Visual Computing for Biology and Medicine*. VCBM '16. Bergen, Norway: Eurographics, 2016, pp. 123–132. ISBN: 9783038680109.

[240] Patrick Saalfeld, Anna Schmeier, Wolfgang D'Hanis, Hermann-Josef Rothkötter, and Bernhard Preim. "Student and Teacher Meet in a Shared Virtual Reality: A one-on-one Tutoring System for Anatomy Education." In: *Proc. of Eurographics Workshop on Visual Computing for Biology and Medicine (EG VCBM)*. 2020.

[241] Riza Salar, Faruk Arici, Seyma Caliklar, and Rabia M Yilmaz. "A model for augmented reality immersion experiences of university students studying in science education." In: *Journal of Science Education and Technology* 29 (2020), pp. 257–271.

[242] Marilyn C Salzman, Chris Dede, R Bowen Loftin, and Jim Chen. "A model for understanding how virtual reality aids complex conceptual learning." In: *Presence: Teleoperators & Virtual Environments* 8.3 (1999), pp. 293–316.

[243] P. Sánchez-González, A. M. Cano, I. Oropesa, F. M. Sánchez-Margallo, F. Del Pozo, P. Lamata, and E. J. Gómez. "Laparoscopic video analysis for training and image-guided surgery." In: *Minim Invasive Ther Allied Technol* 20.6 (2011), pp. 311–320.

[244] Maria V Sanchez-Vives and Mel Slater. "From presence to consciousness through virtual reality." In: *Nature Reviews Neuroscience* 6.4 (2005), pp. 332–339.

[245] Naveen Kumar Sankaran, Harris J Nisar, Ji Zhang, Kyle Formella, Jennifer Amos, Lisa T Barker, John A Vozenilek, Steven M LaValle, and Thenkurussi Kesavadas. "Efficacy Study on Interactive Mixed Reality (IMR) Software with Sepsis Prevention Medical Education." In: *IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 2019, pp. 664–670.

[246] Raimund Schatz, Georg Regal, Stephanie Schwarz, Stefan Suettc, and Marina Kempf. "Assessing the QoE impact of 3D rendering style in the context of VR-based training." In: *2018 Tenth international conference on quality of multimedia experience (QoMEX)*. IEEE. 2018, pp. 1–6.

[247] D Schmalstieg. *Augmented Reality, Principles and Practice*. Addison-Wesley Professional, 2016.

[248] Rahel Louise Schmied-Kowarzik and Volker Paelke. "Examining the Importance of Realism in Virtual Reality Therapy Environments for People With Specific Phobias." In: *GI VR / AR Workshop*. Ed. by Martin Weier, Matthias Bues, and Reto Wechner. Gesellschaft für Informatik e.V., 2021. DOI: 10.18420/vrar2021_16.

[258] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. "Applying the user experience questionnaire (UEQ) in different evaluation scenarios." In: *Design, User Experience, and Usability. Theories, Methods, and Tools for Designing the User Experience: Third International Conference, DUXU 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014, Proceedings, Part I 3*. Springer. 2014, pp. 383–392.

[259] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. "Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S)." In: *International Journal of Interactive Multimedia and Artificial Intelligence* 4.6 (2017), p. 103. DOI: 10.9781/ijimai.2017.09.001.

[260] Jan-Henrik Schröder, Daniel Schacht, Niklas Peper, Anita Marie Hamurculu, and Hans-Christian Jetter. "Collaborating across realities: Analytical lenses for understanding dyadic collaboration in transitional interfaces." In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 2023, pp. 1–16.

[261] Helmut Schrom-Feiertag, Georg Regal, Jaison Puthenkalam, and Stefan Suette. “Immersive Experience Prototyping: Using Mixed Reality to Integrate Real Devices in Virtual Simulated Contexts to Prototype Experiences with Mobile Apps.” In: *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE. 2021, pp. 75–81.

[262] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. “The Experience of Presence: Factor Analytic Insights.” In: *Presence: Teleoperators and Virtual Environments* 10.3 (2001), pp. 266–281. DOI: 10.1162/105474601300343603.

[263] D. H. Schunk. *Learning Theories: An Educational Perspective*. Pearson, 2012. ISBN: 9780137071951.

[264] Manfred Schwab. *Encyclopedia of cancer*. Springer Science & Business Media, 2008.

[265] Valentin Schwind, Pascal Knierim, Nico Haas, and Niels Henze. “Using presence questionnaires in virtual reality.” In: *Proceedings of the 2019 CHI conference on human factors in computing systems*. 2019, pp. 1–12.

[266] Mickael Sereno, Xiyao Wang, Lonni Besançon, Michael J Mcguffin, and Tobias Isenberg. “Collaborative work in augmented reality: A survey.” In: *IEEE Transactions on Visualization and Computer Graphics* 28.6 (2020), pp. 2530–2549.

[267] W. Si, X. Liao, Q. Wang, and P. Heng. “Augmented Reality-Based Personalized Virtual Operative Anatomy for Neurosurgical Guidance and Training.” In: *IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 2018, pp. 683–684.

[268] Stuart G Silverman, Bruce D Collick, Michael R Figueira, Ramin Khorasani, Douglass F Adams, Robert W Newman, George P Topulos, and Ferenc A Jolesz. “Interactive MR-guided biopsy in an open-configuration MR imaging system.” In: *Radiology* 197.1 (1995), pp. 175–181.

[269] Adalberto L. Simeone, Mohamed Khamis, Augusto Esteves, Florian Daiber, Matjaž Kljun, Klen Čopič Pucihar, Poika Isokoski, and Jan Gugenheimer. “International Workshop on Cross-Reality (XR) Interaction.” In: *Companion Proceedings of the 2020 Conference on Interactive Surfaces and Spaces*. ISS Companion ’20. Virtual Event, Portugal: Association for Computing Machinery, 2020, pp. 111–114. ISBN: 9781450375269. DOI: 10.1145/3380867.3424551.

[270] Henrik Singmann, Ben Bolker, Jake Westfall, Frederik Aust, and Mattan S. Ben-Shachar. *afex: Analysis of Factorial Experiments*. R package version 1.2-1. 2023. URL: https://CRAN.R-project.org/package=afex.

[271] Sakshi Sinha, Veronica DeYoung, Amit Nehru, Danielle Brewer-Deluce, and Bruce C. Wainman. "Determinants of Learning Anatomy in an Immersive Virtual Reality Environment – A Scoping Review." In: *Medical Science Educator* (2022). ISSN: 2156-8650. DOI: 10.1007/s40670-022-01701-y.

[272] Richard Skarbez, Frederick P Brooks Jr, and Mary C Whitton. "A survey of presence and related concepts." In: *ACM computing surveys (CSUR)* 50.6 (2017), pp. 1–39.

[273] Richard Skarbez, Solene Neyret, Frederick P Brooks, Mel Slater, and Mary C Whitton. "A psychophysical experiment regarding components of the plausibility illusion." In: *IEEE transactions on visualization and computer graphics* 23.4 (2017), pp. 1369–1378.

[274] Richard Skarbez, Missie Smith, and Mary C Whitton. "Revisiting Milgram and Kishino's reality-virtuality continuum." In: *Frontiers in Virtual Reality* 2 (2021), p. 647997.

[275] Rick Skarbez, Missie Smith, and Mary Whitton. "It Is Time to Let Go of'Virtual Reality'." In: *Communications of the ACM* 66.10 (2023), pp. 41–43.

[276] Mel Slater. "A note on presence terminology." In: *Presence connect* 3.3 (2003), pp. 1–5.

[277] Mel Slater. "Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments." In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 364.1535 (2009), pp. 3549–3557.

[278] Mel Slater. "Immersion and the illusion of presence in virtual reality." In: *British journal of psychology* 109.3 (2018), pp. 431–433.

[279] Mel Slater, Domna Banakou, Alejandro Beacco Porres, Jaime Gallego Vila, Francisco Macia Varela, and Ramon Oliva Martínez. "A separate reality: an update on place illusion and plausibility in virtual reality." In: *Frontiers in virtual reality* 3.Article 914392 (2022), pp. 1–16.

[280] Mel Slater, Pankaj Khanna, Jesper Mortensen, and Insu Yu. "Visual realism enhances realistic response in an immersive virtual environment." In: *IEEE computer graphics and applications* 29.3 (2009), pp. 76–84.

[281] Mel Slater and Sylvia Wilbur. "A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments." In: *Presence: Teleoperators & Virtual Environments* 6.6 (1997), pp. 603–616.

[282] Y. Slavova and M. Mu. "A Comparative Study of the Learning Outcomes and Experience of VR in Education." In: *IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 2018, pp. 685–686.

[283] Jennifer Smerling, Charles C Marboe, Jay H Lefkowitch, Martina Pavlicova, Emile Bacha, Andrew J Einstein, Yoshifumi Naka, Julie Glickstein, and Kanwal M Farooqi. "Utility of 3D Printed Cardiac Models for Medical Student Education in Congenital Heart Disease: Across a Spectrum of Disease Severity." In: *Pediatric Cardiology* 40.6 (2019), pp. 1258–1265. ISSN: 1432-1971 (Electronic). DOI: 10.1007/s00246-019-02146-8.

[284] Håvard Snarby, Tarald Gåsbakk, Ekaterina Prasolova-Førland, Aslak Steinsbekk, and Frank Lindseth. "Procedural Medical Training in VR in a Smart Virtual University Hospital." In: *Smart Education and e-Learning 2018 5*. Springer. 2019, pp. 132–141.

[285] Maarten Someren, Yvonne Barnard, and Jacobijn Sandberg. *The Think Aloud Method - A Practical Guide to Modelling CognitiveProcesses*. Academic Press Inc, 1994.

[286] Andrej Somrak, Matevž Pogačnik, and Jože Guna. "Suitability and comparison of questionnaires assessing virtual reality-induced symptoms and effects and user experience in virtual environments." In: *Sensors* 21.4 (2021), p. 1185.

[287] Maximilian Speicher, Brian D Hall, and Michael Nebeling. "What is mixed reality?" In: *Proceedings of the 2019 CHI conference on human factors in computing systems*. 2019, pp. 1–15.

[288] Tania Steren. "From intellectual craftsmanship to virtual context: methodological tools for social research." In: *Sociologias* 5.SE (2010), pp. 0-0.

[289] J Steuer. "Defining virtual reality: Dimensions determining telepresence." In: *Communication in the age of virtual reality/Lawrence Erlbaum and Associates* (1995).

[290] Jonathan A. Stevens and J. Peter Kincaid. "The Relationship between Presence and Performance in Virtual Simulation Training." In: *Open Journal of Modelling and Simu-*

lation 03.02 (2015), pp. 41–48. DOI: 10.4236/ojmsi.2015.32005.

[291] Zane Stromberga, Charlotte Phelps, Jessica Smith, and Christian Moro. "Teaching with disruptive technology: the use of augmented, virtual, and mixed reality (HoloLens) for disease education." In: *Biomedical Visualisation: Volume 9*. Springer, 2021, pp. 147–162.

[292] John Sweller, Paul Ayres, Slava Kalyuga, John Sweller, Paul Ayres, and Slava Kalyuga. "Measuring cognitive load." In: *Cognitive load theory* (2011), pp. 71–85.

[293] Keiran Tait, Matthieu Poyade, and Jennifer A Clancy. "eLearning and Embryology: Designing an Application to Improve 3D Comprehension of Embryological Structures." In: *Biomedical Visualisation : Volume 7*. Ed. by Paul M Rea. Cham: Springer, 2020, pp. 19–38. ISBN: 978-3-030-43961-3. DOI: 10.1007/978-3-030-43961-3_2.

[294] Yuk Ming Tang, Ka Yin Chau, Alex Pak Ki Kwok, Tongcun Zhu, and Xiangdong Ma. "A systematic review of immersive technology applications for medical practice and education-trends, application areas, recipients, teaching contents, evaluation methods, and performance." In: *Educational Research Review* 35 (2022), p. 100429.

[295] Catherine Taylor and Darren Cosker. "Interacting with real objects in virtual worlds." In: *Real VR–Immersive Digital Reality: How to Import the Real World into Head-Mounted Immersive Displays* (2020), pp. 337–353.

[296] Talia Tene, Diego Fabián Vique López, Paulina Elizabeth Valverde Aguirre, Luz María Orna Puente, and Cristian Vacacela Gomez. "Virtual reality and augmented reality in medical education: an umbrella review." In: *Frontiers in Digital Health* 6 (2024), p. 1365345.

[297] Theophilus Teo, Mitchell Norman, Gun A Lee, Mark Billinghurst, and Matt Adcock. "Exploring interaction techniques for 360 panoramas inside a 3D reconstructed scene for mixed reality remote collaboration." In: *Journal on Multimodal User Interfaces* 14 (2020), pp. 373–385.

[298] Gaetano Tieri, Emmanuele Tidoni, Enea Francesco Pavone, and Salvatore Maria Aglioti. "Mere observation of body discontinuity affects perceived ownership and vicarious agency over a virtual hand." In: *Experimental brain research* 233 (2015), pp. 1247–1259.

[299] Alexander Toet, Tina Mioch, Simon Gunkel, Omar Niamut, and Jan BF van Erp. "Assessment of presence in augmented and mixed reality." In: (2021).

[300] Tanh Quang Tran, Tobias Langlotz, and Holger Regenbrecht. "A Survey On Measuring Presence in Mixed Reality." In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 2024, pp. 1–38.

[301] Wan-Lun Tsai, Li-wen Su, Tsai-Yen Ko, Cheng-Ta Yang, and Min-Chun Hu. "Improve the decision-making skill of basketball players by an action-aware VR training system." In: *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE. 2019, pp. 1193–1194.

[302] Manos Tsakiris and Patrick Haggard. "The Rubber Hand Illusion Revisited: Visuotactile Integration and Self-Attribution." In: *Journal of Experimental Psychology: Human Perception and Performance* 31.1 (2005), pp. 80–91. DOI: 10.1037/0096-1523.31.1.80.

[303] Duc Van Nguyen, Safa Ben Lakhal, and Amine Chellali. "Preliminary evaluation of a virtual needle insertion training system." In: *2015 IEEE Virtual Reality (VR)*. IEEE. 2015, pp. 247–248.

[304] Steven G. Vandenberg and Allan R. Kuse. "Mental rotations, a group test of three-dimensional spatial visualization." In: *Perceptual and Motor Skills* 47.2 (1978), pp. 599–604. DOI: 10.2466/pms.1978.47.2.599.

[305] Vineesha Veer, Charlotte Phelps, and Christian Moro. "Incorporating mixed reality for knowledge retention in physiology, anatomy, pathology, and pharmacology interdisciplinary education: a randomized controlled trial." In: *Medical Science Educator* 32.6 (2022), pp. 1579–1586.

[306] Caroline Villard, Luc Soler, and Afshin Gangi. "Radiofrequency ablation of hepatic tumors: simulation, planning, and contribution of virtual reality and haptics." In: *Computer Methods in Biomechanics and Biomedical Engineering* 8.4 (2005), pp. 215–227.

[307] Prokopia Vlachogianni and Nikolaos Tselios. "Perceived usability evaluation of educational technology using the System Usability Scale (SUS): A systematic review." In: *Journal of Research on Technology in Education* 54.3 (2022), pp. 392–409.

[308] Jan-Niklas Voigt-Antons, Tanja Kojic, Danish Ali, and Sebastian Möller. "Influence of hand tracking as a way of interaction in virtual reality on user experience." In: *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE. 2020, pp. 1–4.

[309] Alexandra Voit, Sven Mayer, Valentin Schwind, and Niels Henze. "Online, VR, AR, Lab, and In-Situ: Comparison of Research Methods to Evaluate Smart Artifacts." In: *Proc. of CHI Conference on Human Factors in Computing Systems*. CHI '19. Glasgow, Scotland Uk: ACM, 2019, pp. 1–12. ISBN: 9781450359702. DOI: 10.1145/3290605.3300737.

[311] Jing-Yuan Wang, Yujie Gong, Jie Zhou, Fenfen Qi, Yingzhi Liu, Zaiheng Cheng, Rongliang Chen, Xinhong Wang, Li Luo, and Xiao-Chuan Cai. "A Collaborative Mixed Reality Platform for the Planning of Vascular Surgery." In: *2024 IEEE International Conference on Metaverse Computing, Networking, and Applications (MetaCom)*. IEEE. 2024, pp. 9–16.

[312] Nanjia Wang, Yue Li, Francesco Chiossi, Fabian Pointecker, Lixiang Zhao, and Daniel Zielasko. *The Second Joint Workshop on Cross Reality*. 2024. arXiv: 2407.19843 [cs.HC]. URL: https://arxiv.org/abs/2407.19843.

[313] Yu Wang, Yuanjie Wu, Sungchul Jung, Simon Hoermann, Shouwen Yao, and Robert W Lindeman. "Enlarging the usable hand tracking area by using multiple leap motion controllers in VR." In: *IEEE Sensors Journal* 21.16 (2021), pp. 17947–17961.

[314] Colin Ware, Kevin Arthur, and Kellogg S Booth. "Fish tank virtual reality." In: *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*. 1993, pp. 37–42.

[315] Chris Watkins and Peter Mortimore. "Pedagogy: What do we know." In: *Understanding pedagogy and its impact on learning* (1999), pp. 1–19.

[316] Stefan Weber, David Weibel, and Fred W Mast. "How to get there when you are there already? Defining presence in virtual reality and the importance of perceived realism." In: *Frontiers in psychology* 12 (2021), p. 628298.

[317] David Weibel and Bartholomäus Wissmath. "Immersion in computer games: The role of spatial presence and flow." In: *International Journal of Computer Games Technology* 2011.1 (2011), p. 282345.

[318] Dominik Weishaupt, Victor D. Koechli, and Borut Marincek. *How does MRI work? An introduction into physics and functionality of magnetic resonance imaging*. 6th ed. Wie funktioniert MRI? Eine Einfuehrung in Physik und Funktionsweise der Magnetresonanzbildgebung. Germany: Springer Medizin Verlag, 2009. ISBN: 978-3-540-89572-5. DOI: 10.1007/978-3-540-89573-2.

[319] Clifford R Weiss, Sherif Gamal Nour, and Jonathan S Lewin. "MR-guided biopsy: a review of current techniques and applications." In: *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine* 27.2 (2008), pp. 311–325.

[320] Tim Weißker, André Kunert, Bernd Fröhlich, and Alexander Kulik. "Spatial updating and simulator sickness during steering and jumping in immersive virtual environments." In: *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE. 2018, pp. 97–104.

[321] Ian Westbury, Stefan Hopmann, and Kurt Riquarts. *Teaching as a reflective practice: The German Didaktik tradition*. Routledge, 2012.

[322] Dirk Weyhe, Verena Uslar, Felix Weyhe, Maximilian Kaluschke, and Gabriel Zachmann. "Immersive anatomy Atlas—Empirical study investigating the usability of a virtual reality environment as a learning tool for anatomy." In: *Frontiers in surgery* 5 (2018), p. 73.

[323] Carolin Wienrich, Philipp Komma, Stephanie Vogt, and Marc E Latoschik. "Spatial presence in mixed realities–considerations about the concept, measures, design, and experiments." In: *Frontiers in Virtual Reality* 2 (2021), p. 694315.

[324] Carolin Wienrich, Christine Katharina Weidner, Celina Schatto, David Obremski, and Johann Habakuk Israel. "A virtual nose as a rest-frame-the impact on simulator sickness and game experience." In: *2018 10th international conference on virtual worlds and games for serious applications (VS-Games)*. IEEE. 2018, pp. 1–8.

[325] Robert S Wigton. "See one, do one, teach one." In: *Academic Medicine* 67.11 (1992), p. 743.

[326] Ignatius Alex Wijayanto, Sabarish V Babu, Christopher C Pagano, and Jung Hong Chuang. "Comparing the Effects of Visual Realism on Size Perception in VR versus Real World Viewing through Physical and Verbal Judg-

ments." In: *IEEE Transactions on Visualization and Computer Graphics* 29.5 (2023), pp. 2721–2731.

[327] Rand R. Wilcox. "Chapter 8 - Comparing Multiple Dependent Groups." In: *Introduction to Robust Estimation and Hypothesis Testing*. Ed. by Rand R. Wilcox. Fifth Edition. Academic Press, 2022, pp. 467–539. ISBN: 978-0-12-820098-8. DOI: 10.1016/B978-0-12-820098-8.00014-2.

[328] Michael Wilkinson, Sean Brantley, and Jing Feng. "A mini review of presence and immersion in virtual reality." In: *Proceedings of the human factors and ergonomics society annual meeting*. Vol. 65. 1. SAGE Publications Sage CA: Los Angeles, CA. 2021, pp. 1099–1103.

[329] Bob G Witmer and Michael J Singer. "Measuring presence in virtual environments: A presence questionnaire." In: *Presence* 7.3 (1998), pp. 225–240.

[330] Hsin-Kai Wu, Silvia Wen-Yu Lee, Hsin-Yi Chang, and Jyh-Chong Liang. "Current status, opportunities and challenges of augmented reality in education." In: *Computers & education* 62 (2013), pp. 41–49.

[331] Insu Yu, Jesper Mortensen, Pankaj Khanna, Bernhard Spanlang, and Mel Slater. "Visual realism enhances realistic response in an immersive virtual environment-part 2." In: *IEEE Computer Graphics and Applications* 32.6 (2012), pp. 36–45.

[332] Ambreen Zaman, Anup Roy, Kanij Fatema, Nusrat Jahan Farin, Tanja Doring, and Rainer Malaka. "Explore Voice and Foot-based Interaction Techniques to Navigate 2D Radiological Images in the Virtual Reality Operation Theatre." In: *22nd Int Conf Comput Inf Technol*. IEEE. 2019, pp. 1–7.

[333] Bo Zhang, Nigel Robb, et al. "Immersion experiences in a tablet-based markerless augmented reality working memory game: randomized controlled trial and user experience study." In: *JMIR serious games* 9.4 (2021), e27036.

[334] Nan Zhang, Hongkai Wang, Tianqi Huang, Xinran Zhang, and Hongen Liao. "A VR Environment for Human Anatomical Variation Education: Modeling, Visualization and Interaction." In: *IEEE Transactions on Learning Technologies* (2022), pp. 1–13. DOI: 10.1109/TLT.2022.3227100.

[335] Guanjie Zhao, Minjie Fan, Yibiao Yuan, Fei Zhao, and Huaxing Huang. "The comparison of teaching efficiency between virtual reality and traditional education in medical education: a systematic review and meta-analysis." In: *Annals of Translational Medicine* 9.3 (2021). ISSN: 2305-5847. URL: https://atm.amegroups.org/article/view/61183.

[336] Jingjie Zhao, Xinliang Xu, Hualin Jiang, and Yi Ding. "The effectiveness of virtual reality-based technology on anatomy teaching: a meta-analysis of randomized controlled studies." In: *BMC medical education* 20 (2020), pp. 1–10.

[337] Xingchen Zhou and Pei-Luen Patrick Rau. "Determining fidelity of mixed prototypes: Effect of media and physical interaction." In: *Applied Ergonomics* 80 (Oct. 2019), pp. 111–118. DOI: 10.1016/j.apergo.2019.05.007.

[338] Egui Zhu, Arash Hadadgar, Italo Masiello, and Nabil Zary. "Augmented reality in healthcare education: an integrative review." In: *PeerJ* 2 (2014), e469.

[339] Paul Zimmons and Abigail Panter. "The influence of rendering quality on presence and task performance in a virtual environment." In: *IEEE Virtual Reality, 2003. Proceedings*. IEEE. 2003, pp. 293–294.

EHRENERKLÄRUNG

"Ich versichere hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; verwendete fremde und eigene Quellen sind als solche kenntlich gemacht. Insbesondere habe ich nicht die Hilfe einer kommerziellen Promotionsberaterin/eines kommerziellen Promotionsberaters in Anspruch genommen. Dritte haben von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Ich habe insbesondere nicht wissentlich:

- Ergebnisse erfunden oder widersprüchliche Ergebnisse verschwiegen,
- statistische Verfahren absichtlich missbraucht, um Daten in ungerechtfertigter Weise zu interpretieren,
- fremde Ergebnisse oder Veröffentlichungen plagiiert,
- fremde Forschungsergebnisse verzerrt wiedergegeben.

Mir ist bekannt, dass Verstöße gegen das Urheberrecht Unterlassungs- und Schadensersatzansprüche der Urheberin / des Urhebers sowie eine strafrechtliche Ahndung durch die Strafverfolgungsbehörden begründen kann. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form als Dissertation eingereicht und ist als Ganzes auch noch nicht veröffentlicht."

Magdeburg, den 03.04.2025

Danny Schott

colophon

This document was typeset using the typographical look-and-feel `classicthesis`, developed by Andr'e Miede. The style draws inspiration from Robert Bringhurst's seminal work on typography, "*The Elements of Typographic Style*." The `classicthesis` package is available for both LaTeX and LyX:

https://bitbucket.org/amiede/classicthesis/

application of natural language processing

Generative AI systems (ChatGPT version 4[1]) were purposefully employed to support the linguistic refinement of this work. The decision to incorporate these technologies was based on their capability to enhance the quality and precision of the document's presentation. All AI-generated content was rigorously reviewed to ensure compliance with the principles of good scientific practice.

1 OpenAI, ChatGPT, https://chat.openai.com