The Design Space of Augmented Reality Authoring Tools and its Exploration for the Procedural Training Context

by
Jonas Blattgerste

A doctoral thesis submitted in partial fulfillment of
the requirements for the degree of

Doktor-Ingenieur (Dr.-Ing.)

at
Bielefeld University

Reviewers
PD Dr. Sven Wachsmuth (Bielefeld University)
Prof. Dr. Thies Pfeiffer (University of Applied Sciences Emden/Leer)
Prof. Dr. Raphael Zender (Humboldt University of Berlin)

Examination board
Prof. Dr. Barbara Hammer (Bielefeld University)
Dr. Olya Hakobyan (Bielefeld University)

Defended and approved
22.12.2023

This thesis is printed on age-resistant paper in accordance with ISO 9706.

Abstract

Augmented Reality (AR) has great potential for assistance, training, educational, and prototyping purposes. This is already well established through the scientific literature of the past two decades. Now, suitable hardware is increasingly available, novel network protocols make location-independent usage possible, and exemplary AR use cases are increasingly explored and evaluated by researchers and industry alike. One challenge remaining for AR's widespread adoption is the creation of AR content to utilize this infrastructure and available hardware at scale. As the majority of current AR content, developed through conventional development methods, does not scale appropriately, it is important to enable users themselves to create their own AR content. Ideally, this would utilize ideas, benefits, and insights gained from these specific exemplary AR use cases and make the learnings widely available, comprehensible, and applicable by non-programmers.

As the overall inquiry, this thesis contributes towards filling this gap by systematically establishing the design space of AR authoring tools and then exploring it with the creation of a novel AR authoring tool for the creation of handheld AR procedural trainings. Hereby, it deliberately draws upon real learnings from first developing AR trainings in the same context, the conventional way. To accomplish this overall inquiry, the thesis is split into three major parts:

In the first part, a comprehensive systematic scoping review of all research publications contributing AR authoring tools published between 2000 and 2020 is presented. The 293 articles included in this systematic review are then mapped onto 26 dimensions, and a literature map of AR authoring tools is contributed. Furthermore, the current scopes and gaps of efforts in the field of AR authoring tools are elaborated. Afterward, a first proposal of the design space of AR authoring tools is contributed and discussed based on this systematically established literature map.

In the second part, a detailed account of the conventional development process for complex procedural AR trainings and their evaluation is presented as a case study of developing, evaluating and publishing the Heb@AR App. This part describes the professional, didactic and technical development of the Heb@AR App from the Human-Computer-Interaction perspective and its evaluations in terms of usability, utility, and usefulness. It furthermore discusses these efforts contextualized in a vision of AR-based training and particularly the aspects of AR trainings scalability. It contributes the Heb@AR App itself as an Open Educational Resource that was implemented as a successful learning scenario, but in this also provides evidence that realistically scalable AR training concepts on handheld devices can elicit learning benefits.

In the third and finally part, the TrainAR authoring tool is contributed as an open-source framework based on exploring the previously established design space with learnings from the conventional development of the Heb@AR App. TrainAR is a holistic framework to create scalable procedural handheld AR trainings that enables the development of AR trainings without programming expertise, that includes a full documentation and was evaluated for the usefulness of the authoring tool and the created AR trainings holistically.

Contents

1 Introduction 1
- 1.1 Research Inquiries . . . 3
  - 1.1.1 Research Objective & Questions . . . 4
- 1.2 Structure & Content of this Thesis . . . 4
  - 1.2.1 The Reader's Guide to this PhD Thesis . . . 5
  - 1.2.2 Previously Published Research Articles . . . 6
  - 1.2.3 Open-Source Publications . . . 7
  - 1.2.4 Open Educational Resource Publications . . . 7
  - 1.2.5 Data Publications . . . 7
  - 1.2.6 Acknowledgement of External Contributions . . . 7
  - 1.2.7 Acknowledgement of Funding . . . 8

2 A Systematic Scoping Review and Mapping Study of AR Authoring Tools 9
- 2.1 Previous Classification Efforts . . . 11
  - 2.1.1 Systematic Classification Efforts . . . 11
  - 2.1.2 Non-Systematic Classification Efforts & Proposals . . . 12
- 2.2 Review Goal & Research Questions . . . 13
- 2.3 Review Methodology & Procedure . . . 14
  - 2.3.1 Database Selection . . . 15
  - 2.3.2 Search String Construction . . . 15
  - 2.3.3 Database Results . . . 16
  - 2.3.4 Inclusion & Exclusion Criteria . . . 17
  - 2.3.5 Screening & Review Process . . . 18
  - 2.3.6 Snowballing Procedure . . . 19
  - 2.3.7 Snowballing Results . . . 19
- 2.4 Search & Review Assessment . . . 21
  - 2.4.1 PRISMA & Other SLR Guidance Conformity . . . 21
  - 2.4.2 Threats to Validity . . . 23
  - 2.4.3 Comparison to Overlapping Systematic Reviews . . . 24
- 2.5 Scoping & Mapping Study Methodology . . . 26
  - 2.5.1 Mapping Procedure . . . 26
- 2.6 The Literature Map of AR Authoring Tools . . . 29
  - 2.6.1 Publication Year . . . 29
  - 2.6.2 Authors . . . 30
  - 2.6.3 Publication Format/Type . . . 30
  - 2.6.4 Publisher . . . 31
  - 2.6.5 Contribution . . . 32
  - 2.6.6 Citations . . . 33

2.6.7 References . . . 34
2.6.8 Deployment Purpose . . . 35
2.6.9 Deployment Context . . . 36
2.6.10 Usability Evaluation . . . 39
2.6.11 Availability . . . 39
2.6.12 Construct Author . . . 41
2.6.13 Authoring Hardware . . . 43
2.6.14 Markup Notation . . . 44
2.6.15 Modularity . . . 45
2.6.16 Scene Preview . . . 46
2.6.17 In-Situ Authoring . . . 47
2.6.18 Authoring Interactions . . . 47
2.6.19 App Relationship . . . 48
2.6.20 Construct Distribution . . . 49
2.6.21 Construct User . . . 50
2.6.22 Usage Hardware . . . 51
2.6.23 User Interactions . . . 52
2.6.24 Content Type . . . 53
2.6.25 Content Sequentiality . . . 54
2.6.26 Tracking Type . . . 55
2.7 Overall Scope & Gaps of the Field . . . 55
2.8 Limitations & Threats to the Mapping Study . . . 60
2.9 Current & Future Work . . . 61
2.9.1 Expansion of the Review Timeframe . . . 61
2.9.2 In-Depth Review of Evaluation Efforts . . . 61
2.9.3 Review & Synthesizing of Markup Language Usage . . . 63
2.9.4 Mapping Commercial AR Authoring Tools . . . 65
2.10 Summary . . . 67

3 Constructing the Design Space for AR Authoring Tools 69
3.1 A Taxonomic, Typologic or Categorized Understanding . . . 69
3.2 Constructing a Design Space to Explore . . . 72
3.3 The Design Space of AR Authoring Tools . . . 73
3.3.1 A Systematic Understanding: Neither Comprehensive Nor Definitive . 74
3.3.2 How to Explore the Design Space: A Guiding Framework . . . 74
3.4 Visual Summary . . . 75

4 Heb@AR—Augmented Reality Trainings for Midwifery Education 77
4.1 Toward Augmented Reality-Based Trainings . . . 78
4.1.1 Computer- and Web-Based Trainings . . . 78
4.1.2 Augmented Reality-Based Trainings . . . 79

4.1.3 Acceptance & Scalability of ARBTs . . . 79
4.2 Academization of Midwifery Education & Project Heb@AR . . . 81
4.2.1 Project Heb@AR . . . 81
4.2.2 The Project's Research Questions . . . 82
4.3 Development of the AR Trainings . . . 83
4.3.1 Observation & Documentation of the Procedure . . . 84
4.3.2 Formalization of the Task Process Analysis . . . 85
4.3.3 Establishment of Learning Goals . . . 85
4.3.4 Procedure Transfer into Stateflow Diagrams . . . 85
4.3.5 Technical Development of the Initial Prototypes . . . 85
4.3.6 Iterative Improvements & Amalgamation of AR Trainings . . . 89
4.3.7 Outline of the Development Timeline . . . 91
4.4 The Heb@AR App . . . 93
4.4.1 Central Design Concepts . . . 93
4.4.2 Menus & Peripheral Features of the App . . . 96
4.4.3 Training 1: Emergency Tocolysis . . . 97
4.4.4 Training 2: Preparation for a Cesarean Section . . . 99
4.4.5 Training 3: Reanimation of a Newborn . . . 100
4.4.6 Training 4: Reanimation of a Newborn (Virtual) . . . 104
4.4.7 Training 5: Female Pelvis . . . 105
4.4.8 Supplementary Heb@AR Material . . . 106
4.4.9 Long-term Vision for the App & Analytics Functionality . . . 107
4.5 Development of Efficient Usability Benchmarking Utility . . . 108
4.6 Evaluation of the Heb@AR App . . . 111
4.6.1 Disclaimer . . . 112
4.6.2 Evaluation 1: Preparation of an Emergency Tocolysis . . . 113
4.6.3 Evaluation 2: Permanent Catheter Placement . . . 114
4.6.4 Evaluation 3: Reanimation of a Newbown . . . 116
4.6.5 Evaluation 5: The Female Pelvis . . . 120
4.6.6 Evaluation 4: Virtual Reanimation of a Newborn . . . 126
4.6.7 Evaluation of the Lecturers Perception . . . 130
4.6.8 Summative Evaluation: Objective Final Exam Comparisons . . . 134
4.6.9 Summative Evaluation: Students' Retrospective Feedback . . . 138
4.7 Discussion . . . 141
4.7.1 The Heb@AR App and Our Vision of ARBTs . . . 142
4.7.2 Challenges & Opportunities of Developing the AR Trainings . . . 144
4.7.3 Insights Based on the Evaluation Efforts of Project Heb@AR . . . 147
4.7.4 The Usefulness of the Heb@AR App as a Learning Resource . . . 154
4.7.5 Limitations of the Results . . . 155
4.7.6 Current & Future Work . . . 156
4.8 Summary . . . 157

5 Design Space Exploration: Creating a Heb@AR Authoring Tool? 159
5.1 The 5w-1H-Guided Reflection . . . 159
5.2 Exploration of Potential Design Decisions . . . 161
5.2.1 Content Type, Content Sequentiality, Construct User Hardware, User Interaction Concept, and Tracking Type . . . 163
5.2.2 Authoring Hardware & Internal/External Authoring . . . 163
5.2.3 Authoring Tool Modularity & Authoring Interaction Concept . . . 163
5.2.4 Authoring Contextualization & Authoring Preview . . . 164
5.2.5 Markup Notation & Distribution . . . 164
5.2.6 Additional Decisions & Novel Perspectives . . . 164

6 The TrainAR Framework 165
6.1 The TrainAR Interaction Concept . . . 168
6.1.1 Virtual Training Assembly . . . 168
6.1.2 Adaptive Instructions . . . 169
6.1.3 User Actions . . . 170
6.1.4 Layered Feedback . . . 171
6.1.5 Insights . . . 171
6.2 Exemplary Implementation: A TrainAR Midwifery Training . . . 172
6.2.1 General Considerations and Design Decisions . . . 172
6.2.2 Conceptual & Technical Onboarding . . . 173
6.2.3 Instructions . . . 173
6.2.4 User Actions . . . 174
6.2.5 Guidance & Feedback . . . 176
6.2.6 Professional Midwife Insights . . . 176
6.2.7 Training Assessment . . . 177
6.3 Formative Evaluation & Iterative Improvements . . . 177
6.3.1 Participants . . . 178
6.3.2 Results: Usability . . . 178
6.3.3 Results: User Experience . . . 178
6.3.4 Results: Qualitative Feedback . . . 180
6.3.5 Subsequent Improvements . . . 181
6.4 Didactic Framework & Utility . . . 183
6.4.1 Training Contextualisation & Structure . . . 183
6.4.2 Integration in Curricular Teaching . . . 185
6.4.3 Applying TrainAR to Procedural Training Tasks . . . 186
6.5 TrainAR Components from a Technical Perspective . . . 189
6.5.1 Combining TrainAR Components Towards an AR Authoring Tool . 190
6.5.2 Components of TrainAR from a Technical Perspective . . . 191
6.6 The TrainAR Authoring Tool . . . 193
6.6.1 Design Considerations for the TrainAR Authoring Tool . . . 195

6.6.2 Open-Source Availability and Documentation 198
6.6.3 Envisioned Workflow for Authoring TrainAR Trainings 198
6.6.4 Content Generation through 3D Scanning and Natural Language Prompts 199
6.6.5 Key Aspects of the Technical Implementation 200
6.6.6 Distribution of TrainAR Trainings 203
6.6.7 Beyond TrainARs Statemachine: Expanding on the TrainAR Framework 203
6.7 Discussion 204
6.7.1 Current & Future Work 205
6.8 Summary 207

7 TrainAR Evaluations 209
7.1 Utility and Usability of TrainAR Trainings 210
7.1.1 Preparation of a Tocolytic Injection 211
7.1.2 German–Latin Denomination of the Female Pelvis 212
7.1.3 MARLabs Titration Experiment 213
7.1.4 Exploration of Fruit Ripeness and Sourness 213
7.2 Usability of the TrainAR Authoring Tool 214
7.2.1 Pre-Study and Non-Representative Observations 214
7.2.2 Systematic Usability Study Design 215
7.2.3 Setup and Procedure 216
7.2.4 Participants 217
7.2.5 Results 217
7.3 Domain Expert Requirements for the TrainAR Framework 223
7.3.1 Methodology 223
7.3.2 Results 224
7.4 An Initial Vertical Deployment & Evaluation of TrainAR 226
7.4.1 A TrainAR Training for Radio Relay Assembly 226
7.4.2 Author's Reflection on the Authoring Process 227
7.5 Discussion 228
7.5.1 Evaluations of TrainAR Trainings 228
7.5.2 Evaluation of the TrainAR Authoring Tool 229
7.5.3 Impressions & Requirements from Domain Experts 230
7.5.4 Insights From the First Vertical Evaluation Efforts 231
7.5.5 In-The-Wild Testing Approach & Ongoing Evaluations 232
7.6 Summary 233

8 Conclusion 235
8.1 Additional Contributions of this Thesis 235
8.2 Limitations & Future Work 236

List of Figures 239

List of Tables 251

Bibliography 255

Appendix 307
1 PRISMA 2020 Checklist 308
2 PRISMA-ScR 2018 Checklist 310
3 Systematic Review — Mapping Result Tables 312
4 System Usability Scale Results in Table Format 337
5 TrainAR Authoring Tool Evaluation — Authoring Tasks 342

1 Introduction

"If I have seen further, it is by standing on the shoulders of giants."
— Isaac Newton

In 1968, Sutherland published his paper called "A head-mounted three dimensional display" in the "Fall Joint Computer Conference", where he described a technical project using "half-silvered mirrors in the prisms [...] to see both the images from the cathode ray tubes and objects in the room simultaneously" [449]. Today, this paper is recognized as the beginning of endeavors toward the technical realization of Augmented Reality (AR), which, as a concept under different names, was already discussed in science fiction literature years prior, like the 1959 novel "Starship Troopers" by Robert Heinlein [191]. In the early 1990s, widely accredited to Tom Caudell and David Mizell, the term "Augmented Reality" was coined when they explored the usage of the technology for assistance purposes in aircraft maintenance [45]. Then, in his field-defining work "A taxonomy of mixed reality visual displays" from 1994, Milgram contextualized the term "Augmented Reality" with parallel endeavors of the "Virtual Reality" domain on the Virtuality Continuum. As visualized in Figure 1.1, rooted in this Virtuality Continuum are the so-called "Mixed Reality" technologies that incorporate some amount of visual, auditory and haptic reality, and some amount of computer-generated content. This starts with subtly augmenting reality with visual hints, e.g., through projection-based approaches on the left side of the continuum and ends with fully immersive Virtual Reality (VR) environments, that ideally no longer incorporate aspects from physical reality, on the right end. Also visualized in this figure are exemplary MR technologies like projection-based approaches, handheld AR approaches, cave-based installations and immersive 360-degree videos. While those technologies are today understood to be part of the Virtuality Continuum, only head-mounted AR and immersive VR approaches were initial considered in Milgrams definition of Mixed Reality [321]. The most widely used and more inclusive definition of AR used today was coined by Azuma [22]. In his definition, AR has to follow three characteristics: Firstly, it combines the real world with virtual content. Secondly, it is interactive in real time. And, thirdly, content is registered in three-dimensional space.

Using this technology of AR to contextualize computer-generated information directly into a physical context could revolutionize many application areas like information visualization, task assistance, education, training or spatially contextualized entertainment applications. Having the information "in-situ", directly where it is needed, has self-evident advantages. While the scientific exploration of these advantages and AR's applicability in the mentioned areas has been increasingly explored throughout the last 20 years, currently most AR applications in the literature are carefully developed applications based on predefined tasks and requirements. While this is a feasible way of exploring benefits of technological differences, implementation decisions or context-specific needs in the early research stages, these approaches are time-consuming, resource-

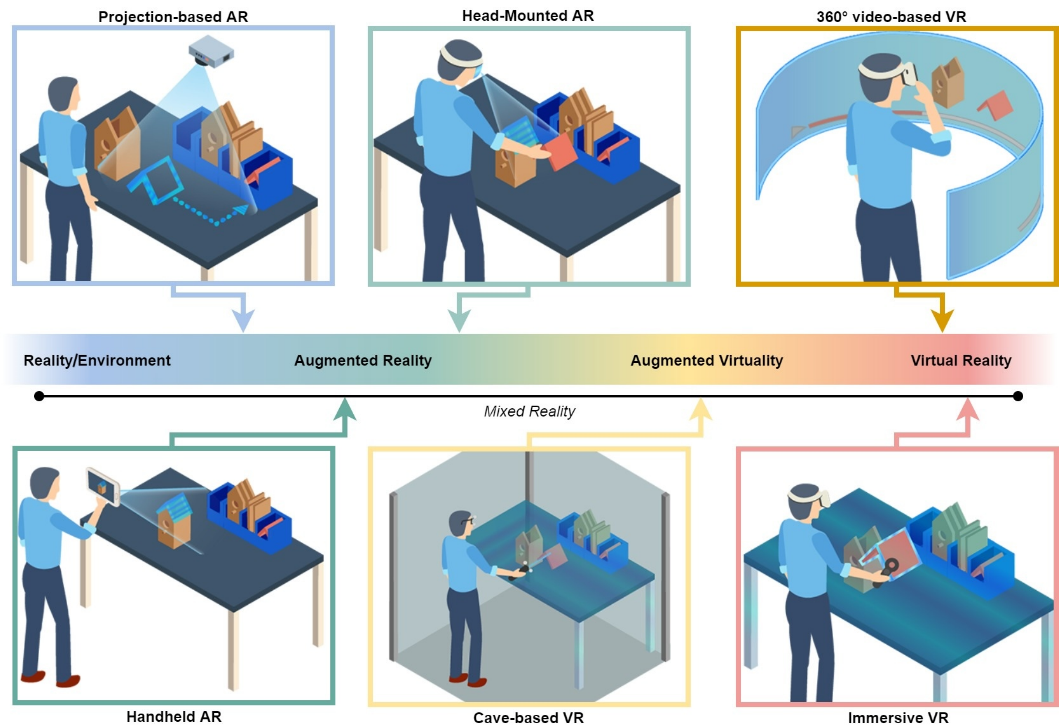


Figure 1.1: Exemplary Mixed Reality technologies like projection-based, handheld, or head-mounted AR, cave-based VR systems, 360-degree videos, and immersive VR, contextualized on the Virtuality Continuum by Milgram [321]. Moving toward the right on the continuum representation, incorporation of real-world elements decreases and incorporation of computer-generated content increases. Figure by Carolin Hainke & Jonas Blattgerste, licensed under CC-BY 4.0

consuming, and ultimately do not scale well, as in this, software developers have to always be involved in the creation of the content. There might be specific application domains, where this granularity of implementation decisions by software developers is necessary, but for the majority of use cases where AR content is needed, there is a need for more efficient approaches [58].

In the end, content needs to be available at scale for the technology to succeed, as the possibilities arising through this potential paradigm change of how we access and interact with digital content and information is already well recognized by researchers and industry alike. According to Masood et al. [312], the compound annual growth of the industrial AR market alone is projected to be around 74% between 2018 and 2025, with a projected aggregated market of 76 billion dollars in 2025 and only increasing market interest beyond that. During the final stages of writing this thesis, Apple just announced the release of its AR headset: "Apple Vision Pro". In the past, Apple releasing products in a technology family was often a driving factor for technology adoption in consumers afterward. Therefore, while AR hardware in the form of head-mounted displays, projection-based approaches, and handheld smartphone-based approaches improved drastically throughout recent years, with increased display sizes and resolution but also improved tracking capabilities, this projected growth alone produces the problem of generating content for these AR hardware solutions, that satisfies this exponentially growing need. And precisely because it is a paradigm change of digital content presentation, the challenge of transforming existing content toward the usage in the AR context is, in my opinion, at least somewhat overlooked. I think it will not be possible to satisfy the upcoming need of AR content without incorporating non-programmers into the process of creating AR content. Therefore, we need to move from exemplary implementations toward modular concepts and tools that non-programmers can use to create their AR content. This process is generally referred to as "Authoring" AR content and in this, programmers create the AR authoring tools, which can then be used by non-programmers/domain-experts to create AR constructs (therefore, things created with the AR authoring tool) independently, that the users can then utilize in the end (see Figure 1.2).

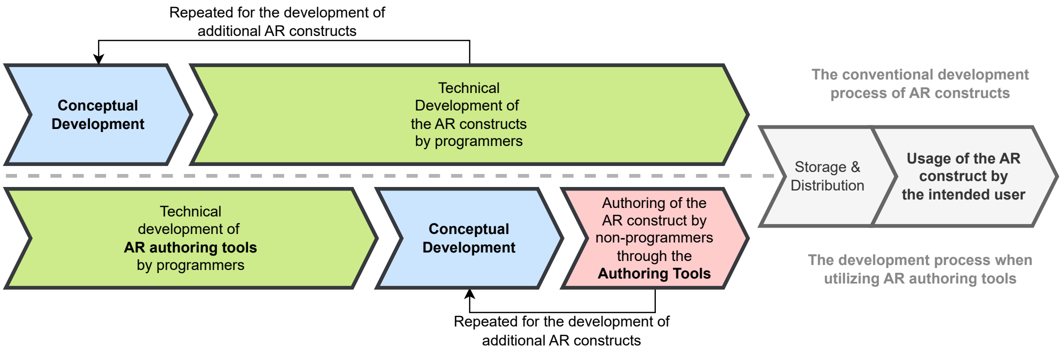


Figure 1.2: The conventional development process (top) and the development process through AR authoring tools (bottom) to distribute AR content, here described as AR constructs, to the intended user of the AR content.

As will be addressed in detail throughout this thesis, this challenge is also already widely recognized by leading researchers, often detailed as one of the main challenges to overcome for widespread adoption of AR, and there are increasing endeavors to address it in the scientific literature. What is missing is a holistic view on this challenge from two orthogonal perspectives: From the theoretical perspective, what AR authoring tools were already developed and evaluated in the literature, and what can be learned from reviewing, combining and structuring existing efforts? From the practical perspective, how can the development of an AR authoring tool be successfully conducted in accordance to actual contextual challenges and opportunities?

1.1 Research Inquiries

While these are broad questions and inquiries which will likely take decades and considerable efforts to fully address, with this thesis, I try to contribute toward creating a first structured understanding of the design space of AR authoring tools by not only contributing a first construction of the design space itself but also exploring it and creating an AR authoring tool for AR procedural task training based on the design space in combination with domain knowledge and practical learnings from first developing AR trainings the conventional way in this exact context the design space is then explored for.

1.1.1 Research Objective & Questions

To accomplish this overall inquiry, this thesis consists of three major components with their own, more granular objectives and research questions.

The first component is a systematic scoping review and mapping study, which reviews and maps the landscape of the literature on AR authoring tools of the last 20 years comprehensively. Based on this, the "Design Space of Augmented Reality Authoring Tools" is constructed as the major objective of the first component. It is a catalog of potential design decisions for AR authoring tools for other researchers or developers to utilize to inform future developments of AR authoring tools based on learnings from previous efforts.

The second component is an exemplary conventional development of an AR construct, the Heb@AR App, in the context of academic midwifery education. This part not only has the objective to create a novel AR training app for procedural task trainings for midwifes, but also contributes 2 novel, scalable AR interaction concepts, a practical transfer-procedure to convert procedural task trainings toward procedural AR trainings, and discusses practical aspects like scalability of handheld AR as a central concept in the vision of AR-based trainings.

Finally, the third component of this thesis has the objective to contribute the open-source TrainAR authoring tool for procedural AR trainings, based on the exploration of the theoretical design space of AR authoring tools with the practical learnings, the vision of ARBTs, and domain knowledge gained during the conventional development of the Heb@AR App. While the constructed Design Space of AR authoring tools goes beyond the scope of procedural task training, this domain is arguably one of the most challenging to explore because of the complexity of the AR constructs and therefore is appropriate for a first exploration of the constructed design space.

The exploratory research questions which will be addressed beyond the holistic inquiry throughout this thesis, can be found in their respective chapters, in Section 2.2 for the first component, Section 4.2.2 and Section 4.6 for the second component, and finally at the start of Chapter 7 for the third component.

1.2 Structure & Content of this Thesis

From the perspective of the chapters of this thesis, the components are structured as follows. After the introduction of this thesis in Chapter 1, a systematic scoping review and subsequent literature mapping study is presented on the overarching topic of AR Authoring Tools in Chapter 2. Then, in Chapter 3, the design space of AR authoring tools is constructed based on the reviewed literature. Chapter 4 shortly diverges from the authoring topic and reports on learnings from the research project Heb@AR, where scalable procedural handheld AR trainings were developed and evaluated for the academic midwifery education context. Here, the vision of scalable AR-based Trainings is discussed, an actual conventional development process of the AR trainings is explained from the HCI perspective, and current evaluation efforts are used to discuss the Heb@AR App as a first exploration of the usefulness of realistically scalable handheld AR trainings. Afterward, in Chapter 5, these practical learnings are used as the variables to explore the design space of AR authoring tools and draft design recommendations for an AR authoring tool in this context. Chapter 6 then contributes the TrainAR framework, an AR authoring tool with a didactic consideration framework and comprehensive documentation. TrainAR is then evaluated from the perspective of the usefulness of the AR authoring tool but also its created AR trainings in Chapter 7, before Chapter 8 concludes this thesis.

As visualized in Figure 1.3, colored by the component, in this, the thesis not only contributes a purely theoretical design space based on a comprehensive systematic scoping review, but also uses practical learnings from an actual conventional development of AR trainings to explore this design space for the context of procedural AR trainings afterward. Based on the combination of the theoretical and practical perspectives, then an open-source AR authoring tool is contributed.

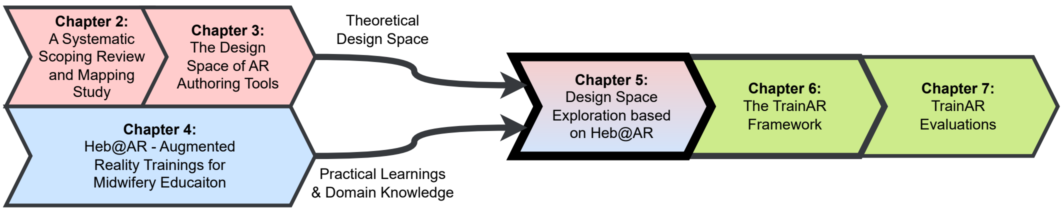


Figure 1.3: A schematic structure of this thesis, colored by component: Practical learnings & domain knowledge of Heb@AR (Chapter 4) are used as variables to explore (Chapter 5) the design space of AR authoring tools (Chapter 2 & 3), before the AR authoring tool TrainAR (Chapter 6 & 7) is developed and evaluated based on this exploration, combining theory and practice.

1.2.1 The Reader's Guide to this PhD Thesis

While this holistic inquiry in itself has value beyond the sum of its parts, the thesis is comparatively long, and it includes several interdisciplinary perspectives, which might not all be of interest for readers with different backgrounds. Therefore, the thesis is deliberately structured in a way, so it can be partially read.

If the reader is interested in how the design space of AR authoring tools was constructed and explored, e.g., to enhance it or develop a design space for a different design challenge, they can read Chapter 2, Chapter 3, and Chapter 5 for an exemplary construction and exploration. If the reader is only interested in exploring the design space of AR authoring tools to develop an AR authoring tool informed by it for their context, they should read Chapter 3, and Chapter 5.

If they are interested in the vision of scalable handheld AR trainings and how the Heb@AR App was successfully developed and evaluated in the context of academic midwifery education, they can read Chapter 4 independently. If, on the other hand, the reader is interested in learning how the practical learnings from the conventional development process of procedural AR trainings in the context of the Heb@AR Project, influenced the development of the TrainAR framework, they can read Chapter 4, Chapter 5, and Chapter 6.

Finally, if the reader is interested in using TrainAR to develop procedural AR trainings for their context and wants to inquire about the framework's usefulness, they can read Chapter 6 & 7.

Because of this approach, some introductions, transitions, and explanation of concepts might be shortly repeated between chapters to ensure that each reading flow is comprehensible.

1.2.2 Previously Published Research Articles

Parts of this thesis are directly based on the previously published research articles. Therefore, texts, figures, ideas, and concepts from these publications are transferred or adopted throughout the thesis without explicit citation:

1. J. Blattgerste, K. Luksch, C. Lewa, M. Kunzendorf, N. H. Bauer, A. Bernloehr, M. Joswig, T. Schäfer, and T. Pfeiffer. "Project Heb@AR: Exploring handheld Augmented Reality training to supplement academic midwifery education". In: *DELFI 2020 – Die 18. Fachtagung Bildungstechnologien der Gesellschaft für Informatik e.V.*. Ed. by R. Zender, D. Ifenthaler, T. Leonhardt, and C. Schumacher. Gesellschaft für Informatik e.V., Bonn, 2020, pp. 103–108. ISBN: 978-3-88579-702-9
2. J. Blattgerste, K. Luksch, C. Lewa, and T. Pfeiffer. "TrainAR: A Scalable Interaction Concept and Didactic Framework for Procedural Trainings Using Handheld Augmented Reality". *Multimodal Technologies and Interaction* 5:7, 2021. ISSN: 2414-4088. DOI: 10.3390/mti5070030
3. J. Blattgerste, J. Behrends, and T. Pfeiffer. "A Web-Based Analysis Toolkit for the System Usability Scale". In: *Proceedings of the 15th International Conference on PErvasive Technologies Related to Assistive Environments*. PETRA '22. Association for Computing Machinery, Corfu, Greece, 2022, pp. 237–246. ISBN: 9781450396318. DOI: 10.1145/3529190.3529216
4. J. Blattgerste, C. Lewa, K. Vogel, T. Willmeroth, S. Janßen, J. Franssen, J. Behrends, M. Joswig, T. Schäfer, N. H. Bauer, A. Bernloehr, and T. Pfeiffer. "Die Heb@AR App - Eine Android & iOS App mit Augmented Reality Trainings für selbstbestimmtes und curriculares Lernen in der hochschulischen Hebammenausbildung". In: *Wettbewerbsband AVRiL 2022*. Ed. by H. Söbke and R. Zender. Gesellschaft für Informatik e.V., Bonn, 2022, pp. 4–9. DOI: 10.18420/avril2022_01
5. J. Blattgerste, K. Vogel, C. Lewa, T. Willmeroth, M. Joswig, T. Schäfer, N. H. Bauer, A. Bernloehr, and T. Pfeiffer. "The Heb@AR App – Five Augmented Reality Trainings for Self-Directed Learning in Academic Midwifery Education". In: *20. Fachtagung Bildungstechnologien (DELFI)*. ed. by P. A. Henning, M. Striewe, and M. Wölfel. Gesellschaft für Informatik e.V., Bonn, 2022, pp. 245–246. ISBN: 978-3-88579-716-6. DOI: 10.18420/delfi2022-052
6. J. Blattgerste and T. Pfeiffer. "TrainAR: Ein Augmented Reality Training Autorensystem". In: *Wettbewerbsband AVRiL 2022*. Ed. by H. Söbke and R. Zender. Gesellschaft für Informatik e.V., Bonn, 2022, pp. 40–45. DOI: 10.18420/avril2022_06

7. J. Blattgerste, J. Franssen, M. Arztmann, and T. Pfeiffer. "Motivational benefits and usability of a handheld Augmented Reality game for anatomy learning". In: *2022 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. 2022, pp. 266–274. DOI: 10.1109/AIVR56993.2022.00056
8. J. Blattgerste, J. Behrends, and T. Pfeiffer. "TrainAR: An Open-Source Visual Scripting-Based Authoring Tool for Procedural Mobile Augmented Reality Trainings". *Information* 14:4, 2023. ISSN: 2078-2489. DOI: 10.3390/info14040219

Furthermore, this thesis reports results from publications, which I was involved in during my PhD, but not the primary author of: Lewa et al. 2021 [274], Luksch et al. 2021 [293], Lewa et al. 2022 [275], Vogel et al. 2022 [481], Dominguez et al. 2022 [114], and Arztmann et al. 2022 [18]. These results, figures, or text passages are always explicitly cited and not directly transferred.

1.2.3 Open-Source Publications

Besides the research publication, several of the technical components of this thesis were already published in open-source repositories on GitHub. Firstly, the TrainAR authoring tool, that was developed for the purpose of this thesis (see Chapter 6 & 7), is accessible under `https://github.com/jblattgerste/TrainAR` using the MIT license, including a comprehensive documentation (`https://jblattgerste.github.io/TrainAR/`). Furthermore, the SUS Analysis Toolkit, a usability benchmarking toolkit that was developed as a side contribution, is accessible under `https://github.com/jblattgerste/sus-analysis-toolkit` through the MIT license. Finally, the supporting Material of the Heb@AR App (Chapter 4) is accessible under the CC BY 4.0 license at `https://github.com/Mixality/HebAR`.

1.2.4 Open Educational Resource Publications

While not currently open-source, the Heb@AR App described in Chapter 4, is published as a free Open Educational Resource in the Android (`https://play.google.com/store/apps/details?id=de.Mixality.HebAR`) and iOS (`https://apps.apple.com/app/heb-ar/id1621822317`) app stores.

1.2.5 Data Publications

In the Appendix 3 of this thesis, tables are included for all dimensions of the systematic scoping review and mapping study reported in Chapter 2, specifying which paper was mapped onto which expression of each dimension. Additionally, for convenience and sustainability, the results of the mapping study are also published in a CC-BY 4.0 licensed multivariate dataset [49].

1.2.6 Acknowledgement of External Contributions

Luckily, I did not have to embark on the journey of scientific exploration and discovery reported in this thesis alone. Several people have contributed to ideas, concepts or technical implementations reported in this thesis or at least influenced them.

Most importantly, many of the perspectives and the vision of scalable AR-based Trainings are a shared vision that arose from countless discussions with Prof. Thies Pfeiffer over the years. Additionally, as the Heb@AR App was developed as part of the interdisciplinary research project Heb@AR, the project partners (Kristina Vogel, Tabea Willmeroth, Carmen Lewa, Dr. Mathias Joswig, Prof. Thorsten Schäfer, Prof. Nicola H. Bauer, Prof. Annette Bernloehr) not only all contributed to the evaluation, ideas and the conceptual development of the app, but also the development of the transfer process itself. As TrainAR is an abstraction and continuation of many of the learnings from the development of the Heb@AR App, in this, they also indirectly contributed to its development. Additionally, Michaela Arztmann supported the design and interpretation of the evaluation efforts grounded in self-determination theory that are reported in this thesis.

In terms of technical contributions, Sven Janßen, Jannik Franssen, Jan Behrends, and Nils Münke all contributed to the technical development of the Heb@AR App as student workers. Benita Stärke contributed to the app as an external contractor for 3D modelling and 2D/3D animations. The AR-Markers were designed with the help of Carolin Hainke. Beyond the development of the Heb@AR App, Sven Janßen and Jan Behrends also contributed to the technical development of the TrainAR Authoring Tool and Jan Behrends contributed to the evaluation efforts as part of his Master's thesis [37]. The technical development of the SUS Analysis Toolkit was primarily performed by Jan Behrends, based on conceptual specifications, as part of his Bachelor's thesis [38] and as a student worker beyond. Other students also contributed towards the project efforts as part of their thesis, which influenced reported ideas and technical concepts. Sven Janßen explored the transformation of the TrainAR interaction concept toward head-mounted AR devices in his Bachelor's thesis [209] and distribution aspects of authored AR trainings in his Master's thesis [210]. Jörg Eggeling explored the usage of the TrainAR authoring tool to develop an AR training in the context of offshore radio relays in his Master's thesis [123]. Jannik Franssen explored the use of analytics functionality in the Heb@AR App in his Bachelor's Thesis [142] and Lars Pastoor explored cloud-based AR training delivery solutions in his Bachelor's Thesis [371].

To acknowledge the contributions of coauthors of papers, conceptual ideas which were developed together with partners, but also the technical contributors to the software components reported in this thesis, the scientific "we" is used in the chapters reporting on the Heb@AR App (Chapter 4), the AR Authoring Design Space Exploration (Chapter 5), and the two TrainAR Chapters 6 & 7. The theoretical Chapters 2 and 3 are written in the passive voice.

1.2.7 Acknowledgement of Funding

This thesis was partially written based on results from the research project Heb@AR (01.11.2019 - 31.12.2022) and was therefore supported by grant 16DHB3021, project "HebAR - AR-Based-Training-Technology", by the German Ministry for Education and Research (BMBF).

Additionally, the thesis was supported by funding from the Mixality Lab at University of Applied Sciences Emden/Leer through a "qualification position" (01.01.2023 - 30.06.2023).

2 A Systematic Scoping Review and Mapping Study of AR Authoring Tools

"The plural of anecdote is data." — Raymond Wolfinger
"The plural of anecdote is not data." — Kenneth Kernaghan

In recent statements, the research and consulting company Gartner, well known for the "Gartner hype cycle", predicts that by 2026, more than 80% of technology will be produced by people outside of IT professions and departments with little to no programming expertise [399]. While those specific numbers might be a bold claim, there is an apparent need and trend toward tools enabling non-programmers to create their own software and products with the ever-increasing technological transformation of almost all parts of our lives.

Ultimately, this is no different for Augmented Reality (AR), in itself an emerging field. For AR as a technology to succeed, useful AR content is required, and to create the content in sufficient quality and quantity at scale, non-programmers have to be enabled to create this AR content. AR technology additionally brings novel challenges in this regard, as it requires three-dimensional content that is contextualized and interacted with through novel context-aware interaction metaphors. That the creation, adaptation, and customization (the authoring) of AR content is, in fact, a need, and the lack of authoring capabilities is holding back widespread adoption of AR, was already discussed by several researchers. For example, Kurkovsky et al. [244] discussed authoring and content management capabilities as one of the three big challenges that is holding back the adoption of Handheld Augmented Reality. In line with these considerations, Schmalstieg et al. [418] discussed "easy to use authoring tools for creating AR content" as one of the 5 big challenges that have to be addressed before AR 2.0 is reached "at massive scale", a comparative metaphor retrospectively analyzing the advancements from Web 1.0 to Web 2.0.

While these hurdles seem to be generally recognized by the field, a comparatively low percentage of the effort of AR research is spent towards AR authoring tools. To provide anecdotal evidence as a means of substantiation of this claim, when Zhou et al. [526] analyzed the topics of ISMAR (IEEE International Symposium on Mixed and Augmented Reality), one of the premier conferences for AR research, from the very first conference in 1998 to the tenth conference in 2008, they found that most publications focused on Tracking, Interaction and Calibration. Only 3.8% of the publications focused on AR authoring and discussed that "some peripheral topics, like [...] authoring are underrated" and efforts only started to gather interest in the later half of the timeframe. When Kim et al. [233] replicated the systematic review of ISMAR publications for the timeframe of 2008 to 2017, they found that there was no apparent trend of this interest continuing. The opposite was the case, the topic of AR authoring had one of the lowest percentages of publications addressing the topic, with now only 2.3%. In later years, there were sometimes no publications on the topic at all. Additionally, when reviewing secondary literature that performs patent analysis to extract industrial interests and trends of AR topics, it is apparent that software and concepts beside tracking algorithms are generally peripheral, but the topic of AR content creation or authoring is entirely absent [85, 102, 129].

This is not to imply that there are no continuous efforts on the topic. In fact, there are several scientific efforts, even entire projects surrounding the authoring and content delivery challenge, like the Studierstube project by Schmalstieg et al. [417], the longstanding open-source effort of the ARToolKit project by Kato et al. [224], or the AMIRE project by Grimm et al. [173]. Beyond that, researchers tried to gather requirements for the authoring process of AR applications [20, 243], explored distribution aspects [310] and collaborative role organization during authoring [205, 218]. There is also already some research on comparing different AR authoring approaches, not only in secondary literature (see the following Section 2.1) but also empirically. For example, Lee et al. [268] compared immersive AR authoring with traditional 2D desktop authoring in terms of usability and efficiency, Madeira et al. [301] compared the usability between desktop and handheld AR interaction techniques for AR authoring, and Yang et al. [513] compared desktop and Head-Mounted Display (HMD)-based immersive authoring approaches.

As can be seen and will be further underlined throughout this chapter, the field of AR authoring tools is not new and there are continuously growing efforts. They are just not growing as fast as the field of AR at large, and they are often hard to identify, e.g. because they are reported as side contributions, or they are particularly interdisciplinary. What is currently missing is an overview: a comprehensive effort to understand the "state of the art" of AR authoring tools. This could help to align efforts towards advancing the field beyond its current hurdle. To contribute towards filling this gap, this chapter reports a systematic scoping review in combination with a systematic literature mapping study, with the rationale of understanding and documenting what AR authoring tools currently exist, how they are designed, evaluated and used, and to analyze trends and challenges as a starting point to address in future work. The methodology of the systematic scoping review was chosen as, in practice, the difference between a systematic literature review and a systematic scoping review is, simply, that the scoping review allows for broader inclusion criteria and more exploratively formulated research questions with which the literature is reviewed [337]. The mapping study methodology was chosen to provide a structured understanding and comprehensible presentation of the comparatively large dataset of publications found during the review.

The remainder of this chapter is structured as follows: Section 2.1 provides an overview of previous reviews and classification efforts of AR authoring tools, Section 2.2 states the goals and questions of the review, Section 2.3 describes the methodology behind the systematic review process, Section 2.4 assesses the review process and literature found, and Section 2.5 describes the methodology of the mapping study. Then, the literature map of AR authoring tools is described in Section 2.6 and the scope and gaps of the field of AR authoring tools is discussed in Section 2.7. Section 2.8 discusses the limitations of the mapping study, before Section 2.9 provides explicit pointers for current and future work, and Section 2.10 summarizes the chapter.

2.1 Previous Classification Efforts

While comprehensive efforts to scope, review, or even classify the scientific literature on AR authoring tools are missing, there are several attempts to classify or even map AR authoring tools, that are not comprehensive in scope or sufficiently sound in methodology to draw the current scope of the field from, but do give a good first impression of how the field developed. Some are secondary literature solely focused on understanding the field and understanding the scope of AR authoring tools; others are side contributions, where a classification of previous efforts is presented before introducing the author's own authoring tool. It should be noted that, while these efforts do exist, they are not particularly visible. Most of the efforts presented in the following were identified during the systematic review itself and not known before conducting it.

2.1.1 Systematic Classification Efforts

The earliest secondary work that tried to create an understanding of differences in AR authoring tool is likely by Roberto et al. [400] from 2016. They reviewed 24 AR authoring tools (commercial and academic) and classified the AR authoring tools on two dimensions: The AR authoring paradigm (Stand-Alone, or AR Plug-in), AR deployment strategy (platform-specific, platform-independent, e.g., through the usage of description languages). Subsequently, they proposed a "general model" to classify authoring tools on, which differentiates 4 expressions: A standalone authoring tool, that uses a platform-specific distribution, a standalone authoring tool that uses platform-independent distribution, and plug-in-solutions with either platform-specific or independent solutions for distribution of the content.

Apaza et al. [12] conducted a systematic literature review in 2018, which included 74 scientific publications on AR authoring tools, where they extracted data to answer 8 research questions. To answer the research questions, they extracted 13 dimensions to analyze: Title, name, year of publication, venue type, venue name, AR programming framework, AR tracking method, development platform, interface projection platform, authoring user interface, general model (inspired by [400]), domain, and validation method.

In 2020, Freitas et al. [143] systematically reviewed 38 AR "authoring", or rather as they describe it: rapid prototyping tools, from both scientific but also gray literature. They reviewed, which artifacts can be used (tools, frameworks, and software), which of them are used most, and what are commonly faced challenges. Furthermore, they mapped the source and publication year and the fidelity of the proposed prototypes.

Dengel et al. [104] conducted a systematic review of AR authoring tools in the educational context in 2022, where they reviewed 26 scientific and gray literature publications on AR authoring tools and mapped them according to the hardware used, evaluation efforts, programming skills required, level of interactivity, affordability/licensing, and toolkit type.

In line with these efforts, Ez-Zaouia et al. [521] reviewed 21 publications in the context of AR authoring tools for the educational context, to answer the questions of what AR features and modalities the tools offer and how emerging tools support the teachers' needs. Finally, they propose a design space with 4 dimensions: Authoring workflow, AR modality, AR use, and content and user management, in which each of the dimensions again has sub-categories, e.g., the Authoring Workflow dimension then classifies tools based on if they support collaborative authoring, which platform they use, and based on the contents source. Altogether, they map the 21 publications included in the review onto 17 dimensions.

While not primarily focused on AR authoring tools, Palmarini et al. [362] conducted a systematic literature review on AR in maintenance, which reviewed 30 publications to answer 2 research questions: What is the "state of the art" of AR applications in the maintenance context and what are potential future developments. Here, besides mapping the hardware, tracking methods, development platform, and visualization approaches used, they also mapped if the AR applications included had authoring functionality. Also in the context of AR for maintenance, Del Amo et al. [138] conducted a systematic review of content-related techniques for knowledge transfer, which were not primarily based on AR authoring but categorized dimensions like: Assets used, operation, task to be created, knowledge (procedural or declarative), and the level of automation during authoring. Limbu et al. [285] systematically reviewed AR authoring approaches that use sensors and Augmented Reality to record expert performances as instructions and specifically differentiate, how the tasks are captured (e.g., through demonstration, or modelling with task analysis descriptions) and which instructional design methods are deployed.

2.1.2 Non-Systematic Classification Efforts & Proposals

Non-systematically, Terenzi et al. [458] discussed the differentiation of the level of integration of AR authoring tools into other software frameworks in line with the differentiation of Plug-in and Standalone AR authoring tools by Roberto et al. [400]. Beyond this simple differentiation, they furthermore also introduce the concepts of partially integrated Plugin-solutions, and argue the case for why Plug-in solutions might lower the required effort for professionals to acquire sufficient skill proficiency to utilize the AR authoring tool effectively. This is done as a framing for the proposal of a plugin solution called AR-media, which they propose to be used to connect 3D content to AR markers directly in Autodesk 3Ds Max.

Before introducing a learning environment for AR mobile learning, which also includes authoring functionality, Cabillo et al. [98] reviewed 22, mostly commercially available, AR authoring tools and frameworks based on how they are licensed and how much programming skills are required to utilize them. Meccawy [316] reviewed the general context and process of creating a XR learning experience in an effort to provide a "roadmap" for educators to follow if they want to create their own XR trainings. In this, they also include AR authoring tools and roughly group them by the required programming skills and target usage context.

Ahead of introducing the AR authoring tool "AuthAR", Whitelock et al. [500] propose the "design space of AR assembly task tutorials", where they non-systematically establish presented content, authoring location, content creation automation level, content editing, and interaction techniques used as dimensions of interest. They subsequently use this design space to explain the rationale behind the development of their authoring tool. In similar fashion, Begout et al. [36]

non-systematically grouped 23 commercially available AR authoring tools and frameworks based on the "main features" necessary, from their perspective, to perform task authoring of AR instructions in the manual assembly context. Furthermore, Geng et al. [163] reviewed and categorized commercially available AR authoring tools based on the authoring process, intended users, and structure of the authoring process. They then used this as a starting point to differentiate their contribution of a proposed AR authoring tool for industrial maintenance assistance.

Hampshire et al. [184] proposed the first effort towards understanding the AR authoring design space in a taxonomic sense, and proposed the differentiation of AR authoring tools into "low-level programming", "high-level programming", "low-lovel content design" and "high-level content design", based on how much programming skills and media competencies are required to use them. Additionally, they non-systematically classify some already developed AR authoring tools to demonstrate the taxonomies expressions and discuss how they expect an inherent tradeoff between concept abstraction and application interface abstraction. Yilmaz et al. [516] subsequently classified more commercially available and early academic AR authoring tools on this proposed taxonomy, expanded it towards differentiating between Mobile and Desktop platforms and discussed which ones would realistically be usable for content creation by teachers.

2.2 Review Goal & Research Questions

As can be seen, there are initial endeavors, which emerged throughout recent years, and they tried to classify, map, or structure the understanding of AR authoring tools. Some as a main contribution and some as a side contribution, before proposing their own AR authoring tool. Especially the later suggest that there is a need for clarification and a common conceptual understanding of AR authoring tools. At least, to distinguish novel contributions. While these endeavors are important first steps, they lack comprehensiveness, systematicness, and depth of dimensional classification. The selection of dimensions of interest to review themselves vary as well, indicating there is not an established set of dimensions of interest yet.

To fill this gap, the goal of this review is to create an overarching understanding of the field of AR authoring tools, by scoping the research field and mapping AR authoring tools proposed over the years, to create a foundation for directed future work in the field. Therefore, the research questions for the scoping review and literature mapping study are the following:

1. **What is the current scope of the research field of AR authoring tools?** For example, how many, when and where were papers published on the topic? Which research fields (e.g., Education, Medicine, Industry) are covering the topic? What are those AR authoring tools used for? How are the usage of tools or created instructions evaluated?
2. **How can differences and similarities of proposed authoring tools be mapped?** For example, what are the capabilities of proposed AR authoring tools? How and by whom can those tools be used? What are the capabilities and modalities of the applications created through those authoring tools?

2.3 Review Methodology & Procedure

As the methodology of the systematic literature search and review, a combination of database searches in primary publisher databases and research aggregators was chosen, with a subsequent, extensive, so-called 'parallel snowballing' [327] search strategy. In this, one or several database searches, using predefined search strings, are combined with forward snowballing, where all citations of included publications are scanned and reviewed, and backward snowballing, where all references of publications are scanned and reviewed, to identify additional publications to include. This approach is referred to as a "full-fledged hybrid search strategy" [506]. The combination of the database search and snowballing process was based primarily on the "Guidelines for snowballing in systematic literature studies and a replication in software engineering" by Wohlin [504] but the database searches were carried out and reported according to PRISMA guidelines and the checklist for systematic literature reviews [361]. The PRISMA guidelines do not provide concrete, realistic guidelines for the snowballing component of the search strategy (see Section 2.4). The hybrid search strategy of combining database searches with subsequent snowballing, is shown to outperform standalone approaches in terms of comprehensibility of found literature [24, 207, 327, 506], as forward snowballing improves precision and backward snowballing recall [327] in a "full-fledged hybrid search strategy". In particular, this strategy is the most effective hybrid search strategy, which is the least dependent on the quality of the initial database, at the cost of increased required effort [506] and therefore decreased efficiency and precision overall. Non-hybrid approaches, while similarly effective compared to each other, can lead to barely overlapping sets of publications as a result, which could even lead to contradictory impressions of the reviewing researchers [62]. Therefore, to satisfy all the criteria of a high-quality systematic literature review, namely being systematic, comprehensive, and transparent [172], the hybrid approach of combining six database searches with subsequent parallel forward- and backward snowballing was chosen, despite the increased required efforts and the limitations of not being fully PRISMA conform. For readability, the "full-fledged hybrid search strategy" will be referred to as "hybrid search strategy" and the "parallel snowballing" as just "snowballing" from here on out.

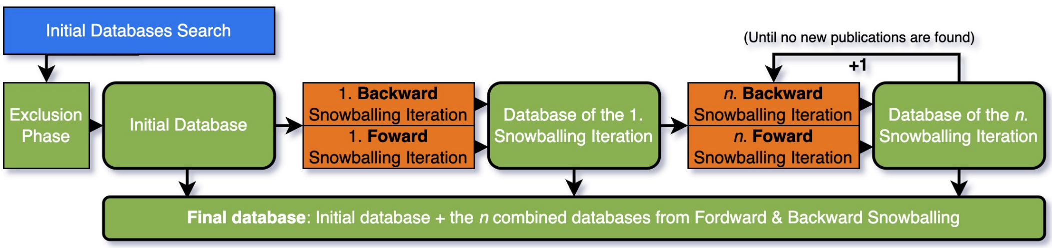


Figure 2.1: The "full-fledged hybrid search strategy" [506] with "parallel snowballing" [327] used for the systematic review. An initial database search with subsequent exclusion phase based on predetermined criteria is followed up by n iterations of combined forward- and backward snowballing, until no new publications to include are found. This figure is recreated based on the combination of ideas in concept figures described by Wohlin et al. [504, 506].

2.3.1 Database Selection

For the initial database search, four publisher databases were used: the ACM Digital Library, IEEE Xplore, Elsevier ScienceDirect, and SpringerLink. Furthermore, two literature aggregators: Google Scholar and Web of Science, were selected to supplement the primary databases.

The primary publisher databases were identified through a non-representative exploration using Google Scholar. Google Scholar is the most comprehensive literature aggregator [131, 176, 487] and is shown to provide relevant literature more efficiently, while remaining as effective, as comparative approaches [426]. After some terminology exploration, the search string:

"intitle:"augmented reality" AND (intitle:"content creation" OR intitle:"authoring")

was used and the first 5 pages with 10 publications each, were analyzed for their publisher to identify the most promising publishers. While it is often criticized as non-transparent and constantly changing [167], the ranking of publications in Google Scholar correlates with citation counts [308] and seems to favor established publishers, making it a good starting point to identify influential publications and their publishers to find the most promising databases for the actual search.

For the literature aggregators, Google Scholar and Web of Science were selected, as they are the biggest and most inclusive active aggregators [309] and provide powerful search instruments. Google Scholar is often critiqued for its inclusive policy, ranking and lacking algorithmic transparency, and it is generally not advised to use it exclusively [167]. Especially as search strings have to be sufficiently concrete to get a number of results, that can realistically be retrieved. Nonetheless, it also includes gray literature, which could have important insights and pointers, even if not ultimately included in the review. It also prevents publication bias [504]. Scopus was deliberately not chosen as one of the aggregators, as it is a subscription-based service, that is not freely accessible. It should be noted that there is literature recommending Scopus as even the only database to use for the database search during a hybrid search strategy [327]. This adds efficiency at the cost of reproducibility and transparency because of inaccessibility.

2.3.2 Search String Construction

After the database selection, the search string was constructed. Because of the hybrid approach, which includes subsequent snowballing after the initial database search, the search had to be as inclusive as possible, while still yielding realistically retrievable publication numbers, but did not have to be comprehensive. Several search strings were explored in Google Scholar. Strings containing "creating", "editing", "annotating" or "editing" were tried, but even in combination with "Augmented Reality" yielded too many results to be retrievable or too many "false positives", therefore publications that at first glance clearly were not in scope of the review. Search strings that yielded good results with realistically retrievable numbers of publications were, e.g., combinations of "Augmented Reality" and either "authoring" or "content creation", which are close to what was already used in the database selection process. Therefore, search strings for each database were constructed, that were as close as possible to this concept with the given set of tools (see Table 2.1). As the capabilities of the search functionalities of the databases differ greatly, this is an unfortunate but common practice in systematic literature reviews, e.g., see similar recent systematic literature reviews in the field of Augmented Reality, by Palmarini et al. in 2018 [362], Khowaja et al. in 2020 [228], or Gattullo et al. in 2020 [156]. This limitation is an ongoing discussion in the literature review community, see e.g., Gusenbauer et al. [177]. All constructed search strings for each database, the type of search used, and number of records retrieved are visualized in Table 2.1.

Database	Type	Search String	Records
ACM Digital Library	Title	*[PublicationTitle: "augmented reality"] AND [[Publication Title: "authoring"] OR [Publication Title:"content creation"]]*	28
	Abstract	*[Abstract: "augmented reality"] AND [[Abstract: "authoring"] OR [Abstract: "content creation"]]*	67
	Keywords	*[Keywords: "augmentedreality"] AND [[Keywords: "authoring"] OR [Keywords: "content creation"]]*	31
IEEE Xplore	Title + Abstract + Keywords	*(("Author Keywords":"augmented reality" OR "Abstract":"augmented reality" OR "Publication Title":"augmented reality") AND (("Author Keywords":"content creation" OR "Abstract":"content creation" OR "Publication Title":"content creation") OR ("AuthorKeywords":"authoring" OR "Abstract":"authoring" OR "Publication Title":"authoring")))*	125
Elsevier ScienceDirect	Title + Abstract + Keyword	*"augmented reality" AND ("authoring" OR "content creation")*	76
SpringerLink	Title + Text	Title contains: *"authoring"* + Text contains: *"augmented reality"*	58
	Title + Text	Title contains: *"content creation"* + Text contains: *"augmented reality"*	9
Google Scholar	Title	*intitle:"augmented reality" AND (intitle:"authoring" OR intitle:"content creation")*	129
Web of Science	Title + Abstract + Keywords	*(TI="augmented reality" OR AB="augmented reality" OR AK="augmented reality") AND ((TI="authoring" OR AB="authoring" OR AK="authoring") OR (TI="content creation" OR AB="content creation" OR AK="content creation"))*	101
		Combined Records Retrieved:	**624**

Table 2.1: The six databases used to retrieve articles, the types of searches used (title, abstract, keyword, text or combination searches) and the search strings that were used to retrieve publications. A combined 624 publications (including duplicates) were retrieved from this initial database search.

2.3.3 Database Results

On the 07/07/2020, all publications from the six databases were retrieved using the search strings shown in Table 2.1. Publications were retrieved manually, without any crawlers or automations. If databases offered functionality to export all selected articles as citations (this was the case for IEEE Xplore and ACM Digital Library), the functionality was used, manually reviewed and the PDFs of the articles themselves were manually supplemented. All publications were initially stored in a Microsoft Excel file and as PDFs and version controlled through GitHub. While this approach is not particularly efficient, it ensured that no publications were lost, for example because they were indexed but not accessible, and the version controlling would later ensure that the screening/review process could be properly documented.

As shown in Table 2.1 in the "Records" column and visualized in Figure 2.2 on the left side, this resulted in 624 publications overall, including duplicates. Out of those, 129 publications were found through Google Scholar, 126 in the ACM Digital Library, 125 in IEEE Xplore, 101 through Web of Science, 76 in Elsevier ScienceDirect, and the remaining 67 in SpringerLink.

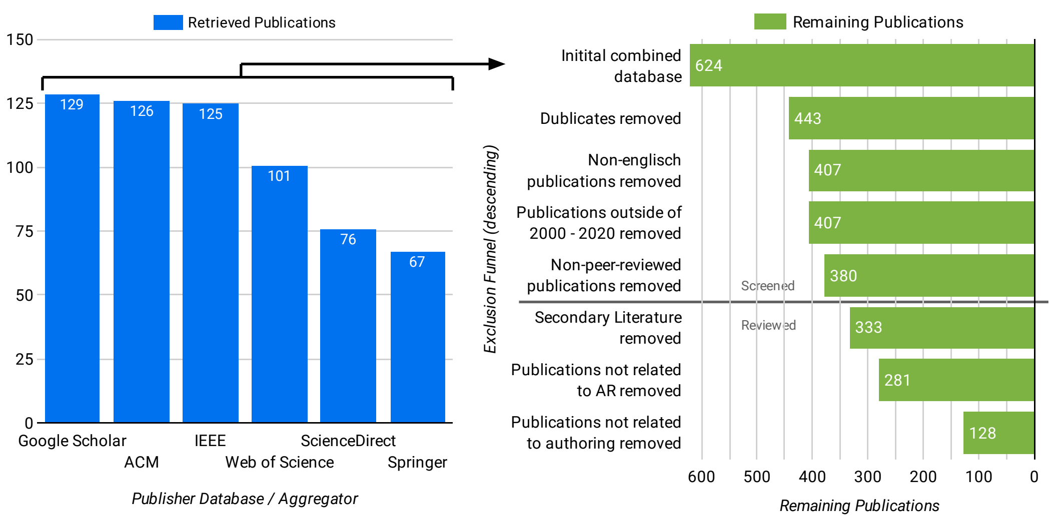


Figure 2.2: The 624 publications and their source database (left). The inclusion/exclusion funnel of publications from the retrieved database to the final 128 publications that were included in the initial database (right).

2.3.4 Inclusion & Exclusion Criteria

To determine which publications are included in the review, six hierarchical inclusion criteria were defined. Here, in line with review methodologies in many fields referred to as "scoping reviews" or "systematic scoping reviews" [511], no criteria assessing the publication's quality were used. In this, the review is as comprehensive as possible, providing a "snapshot of the field" [511]. To be included, publications must:

1. be written in English.
2. be published in the timespan of 2000 and 2020.
3. be in the form of a poster-, short-, long-, or journal publication from a peer reviewed source.
4. be a primary source contributing a tool and neither review articles, discussions, study testing existing applications, tutorials, nor proposals.
5. fit the definition of Augmented Reality proposed by Azuma [22] for the created construct, as this is the most known definition and more inclusive compared to, e.g., the definition by Milgram et al. [322] that only includes head-mounted display (HMD) based Augmented

Reality. In his survey [22], Azuma defines AR systems with the three characteristics: Visually combining real and virtual environments, interactive in real-time and registered in three dimensions.

6. include a novel higher-level authoring, or “content creation” according to Hampshire et al. [184], tool as a significant part of its contribution. Hereby, authoring and content creation tools are defined as an application or a part of an application, that allows human users to manually create, contextualize, extend or edit static, descriptive or procedural Augmented Reality content, that can then be persistently stored, retrieved and utilized by another or the same user of the AR application e.g., for learning, assistance, or entertainment purposes. Specifically, higher-level authoring tools are defined as authoring tools that do not require programming knowledge, though tools based on markup languages and visual scripting are included, which are classified as “low-level content design” tools [184].

To further elaborate on these criteria, in the initial removal of duplicates, only the exact same publications were removed. If papers were not the same publications, but described the exact same AR authoring tool, without apparent changes/updates in their form or function, they were excluded during Inclusion Rule 6. Usually, when several publications described the same authoring tool, either the longer, or alternatively the newer version was chosen. Also removed through Inclusion Rule 6 were publications, that did not author persistent content, like remote assistance applications. Inclusion Rule 6 also excluded automated authoring approaches that had no manual authoring component. Based on Inclusion Rule 3, no PhD thesis, project reports, or other gray literature sources were included. Furthermore, Inclusion Rule 5 relates to the created construct, not the authoring tool creating it. This means, e.g., publications describing desktop-based authoring tools that created AR apps for Smartphones, were included. On the other hand, not included were AR authoring approaches that, e.g., author movement paths for robots, animations to be used in movies, or other non-AR media constructs as a result of the authoring in AR.

2.3.5 Screening & Review Process

After all 624 publications were retrieved, first, 181 duplicates were removed. The remaining 443 publications were subsequently screened. While automated approaches that use Machine Learning can increase the efficiency of screening in systematic reviews, they are shown to come with increased risks of missing relevant literature [155], and therefore the screening was performed manually. 36 publications were removed because they were not written in English, 0 publications were removed because they were outside the established timeframe, and 27 publications were removed because they were not peer-reviewed. After the screening process, 380 reports were sought for retrieval. As all publications could be retrieved, they were subsequently manually reviewed. Out of the 380, publications, 47 were removed because they were secondary literature, 52 were removed because the created AR construct did not fit the definition of AR and 153 were removed because they did not fit the definition of a novel AR authoring tool. This resulted in 128 publications after the initial database search, which means the precision of the search, calculated as the proportion of papers included in the review out of the total number of papers screened and reviewed, was

28.44%. The screening and review process was performed by a single researcher, over the course of about 7 months. The screening and review process is visualized in Figure 2.2 on the right side and in the PRISMA flowchart in Figure 2.5 on the left side.

2.3.6 Snowballing Procedure

The 128 publications remaining in the initial database after the review were subsequently used as the starting point for the snowballing (see Figure 2.1). The snowballing process was performed in February and March 2021.

The references of the publications were screened and reviewed through the PDF file for the backward snowballing, using a browser plugin which could link to an article in Google Scholar based on the title or DOI. Google Scholar's "citation" functionality was used for the forward snowballing, as it is the most comprehensive literature aggregator [131, 176, 487] and references were extracted from the PDF and searched on Google Scholar first. If Google Scholar did not index the publication at the time, they were searched on Google. Publications were "ad hoc" screened in Google Scholar and publications that were interesting based on title and abstract were reviewed, using the inclusion criteria specified in Section 2.3.4. Therefore, publications were not in any case manually extracted before screening/reviewing, and it was not recorded for which reason publications were excluded. This extraction process would have been substantially more time-consuming with the number of publications screened and would have only contributed marginally. In this, the snowballing process is in line with Wohlins' hybrid search strategy [207] but is not PRISMA conform (see Figure 2.5 on the right side in the "Identification of studies via other methods"). Browser plugins were used to clearly highlight already visited links and titles of publications already included in the review, to recognize them in Google Scholar and on PDFs.

As the initial database search was performed on the 07/07/2020, but the timeframe of included papers for the review ranges from 2000 to 2020, some publications expected to be identified during forward snowballing were not available during the initial database search, which should result in lower precision for the forward snowballing.

2.3.7 Snowballing Results

In the first iteration of backward snowballing, the 3,107 references of the 128 publications of the initial database were screened and reviewed. 37 publications were identified in this step (precision 1.19%). Afterward, 4052 citations of the 128 publications were screened and reviewed through Google Scholar, where 62 publications were identified (precision 1.53%). Subsequently, this will be referred to as "citations", or "citations at retrieval", though more accurately they are the citations on the day forward snowballing was performed on this specific publication during the two-month timeframe. After both snowballing directions were completed, the $37 + 62 = 99$ publications were combined into the database of the first snowballing iteration, which had a combined $686 + 1696 = 2382$ references and $1216 + 678 = 1894$ citations. The second backward snowballing iteration identified 14 new included publications from the 2382 publications screened and reviewed (precision 0.59%), the second forward snowballing iteration identified 27 out of 1894 publications to include (precision 1.43%). The combined database of $14 + 27 = 41$ publications had $346 + 710 = 1056$ references and at retrieval was cited $910 + 591 = 1501$ times. The third iteration identified 5 out of 1056 publications (precision 0.47%) during backward-, and 19 out of 1056 publications (precision 1.27%) during forward, snowballing. The fourth iteration identified no publications to be included out of 615 (precision 0%) during the backward-, and 1 out of 427 publications (precision 0.25%) during the forward snowballing. When the 2 citations at retrieval and 32 references of the last identified publication were screened and reviewed in the fifth snowballing iteration, no new publications were identified and the snowballing process concluded.

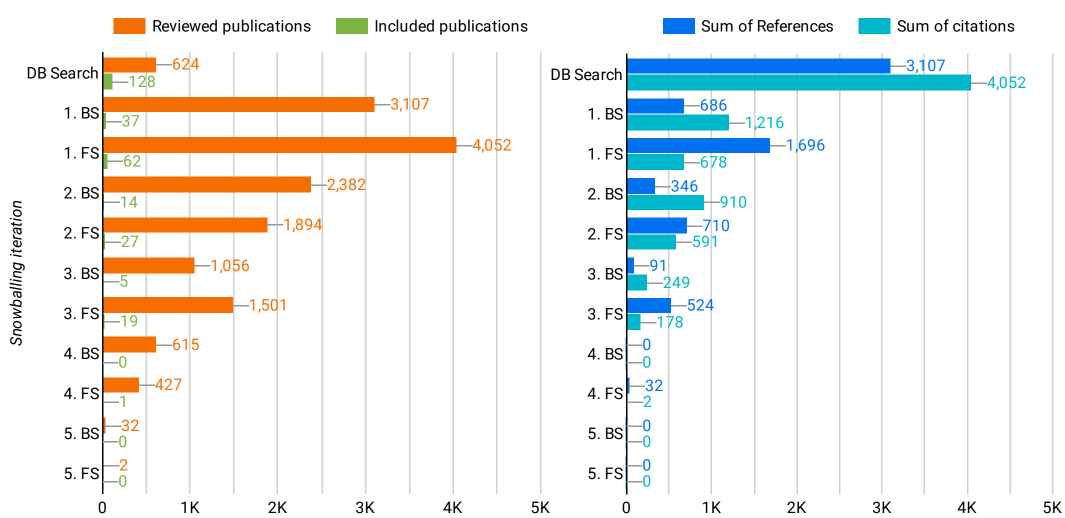


Figure 2.3: The snowballing process of the hybrid search strategy with the 128 publications of the initial database search (DB search) as a starting point, which identified 165 new publications after screening/reviewing 15068 publications (including duplicate references/citations) after 5 iterations of Backward- (BS) and Forward snowballing (FS).

The snowballing process is visualized in Figure 2.3 and included under "Identification of studies via other methods" section of the PRISMA flowchart in Figure 2.5, on the right side. Overall, $7192 + 7876 = 15068$ publications were reviewed and screened, including duplicates. Out of those, 165 were included in the review, with 109 publications coming from the forward-, and 56 coming from the backward snowballing iterations. This makes the finale size of the database of included papers 293 (see Figure 2.4). The overall precision of the snowballing process was 1.1%, with a precision of 0.78% for the combined backward, and 1.38% for the combined forward directions. Examining the backward and forward snowballing directions independently, the number of screened/reviewed publications and the number of included publications are both strictly monotonically decreasing throughout the iterations. This indicates that the initial search strings performed well, and no keywords were missing, which would have resulted in new clusters of publications only using these descriptions. That the snowballing results in as many or more publications as the initial database search is not uncommon [24, 207, 327], as database searches alone have recall percentages of about 43% to 80% [327]. In this review, the relative recall of the database search was 47.58%, likely because of the interdisciplinary perspectives involved, but the snowballing process decreased the overall precision from the initial 28.44% to 1.86%.

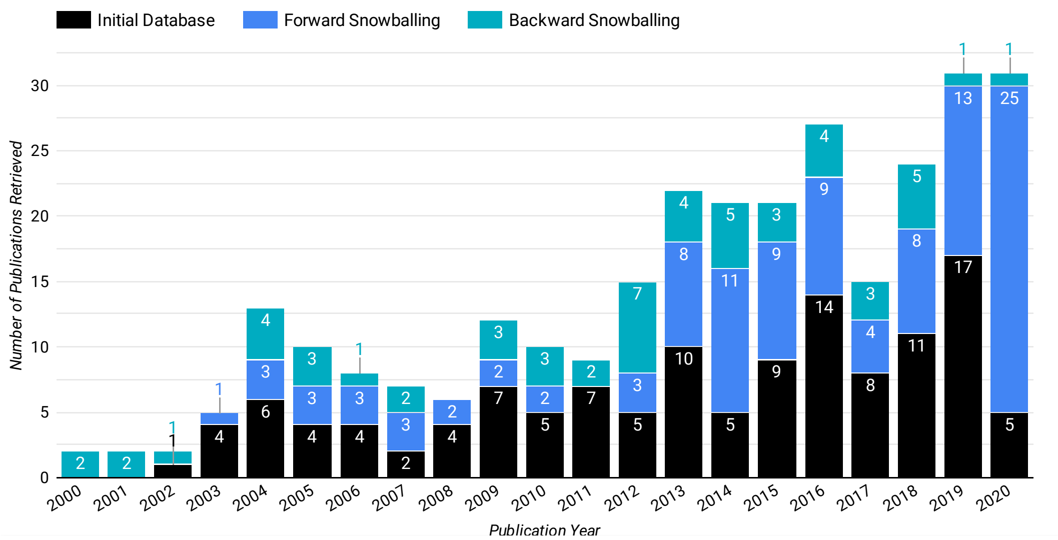


Figure 2.4: The 293 publications included in this review by publication year, split into the search procedure, they were retrieved with. A table with all publications, grouped by the search strategy they were found with, can be found in Appendix 3.

2.4 Search & Review Assessment

Before starting to map the results of this systematic review, the satisfaction of the criteria of a high-quality systematic literature review (systematic, comprehensive, and transparent) [172] are assessed. The quality of the review is assessed by discussing its conformity with PRISMA [361] and other recent guidance to conducting systematic reviews. Additionally, its threats to validity are stated and finally, the publications included are compared to publications included in five similar reviews on AR authoring tools, to assess the comprehensiveness of the review.

2.4.1 PRISMA & Other SLR Guidance Conformity

PRISMA [361], short for "Preferred Reporting Items for Systematic Reviews and Meta-Analyses", is an evidence-based set of rules on the "reporting of reviews evaluating the effects of interventions" [361]. While primarily focused on medical interventions, it is often used for other systematic reviews, as it proposes strict methodological rules and comes with a standardized checklist and flow diagram. It therefore provides guidance for quality systematic reviews beyond the medical field. As the 2020 edition of the PRISMA guidelines [361] explicitly includes single-reviewer study assessments, described in an elaboration paper with "Assessment of each record by one reviewer — Single screening is an efficient use of time and resources, but there is a higher risk of missing relevant studies" [360], the database search (Section 2.3.3) is conducted with complete PRISMA conformity. The results from the snowballing (Section 2.3.7), are not fully conform, as not all 15068 results were extracted before screening, and it was therefore not documented for which of the criteria publications were excluded. Both searches are included in the PRISMA flow diagram in Figure 2.5. The size of the snowballing would not have been manageable without this "ad hoc" screening approach, and it would have added little to no further insights. Notably, while not in line with PRISMA guidelines, this is a common practice in systematic reviews that deploy snowballing or hybrid strategies [207], and increases comprehensiveness at the potential cost of transparency. As this review is not in the context of evidence-based medicine and comprehensiveness is the most important factor in scoping reviews, this was a deliberate decision. Furthermore, for the same reason, no official PRISMA review protocol was prepared and registered.

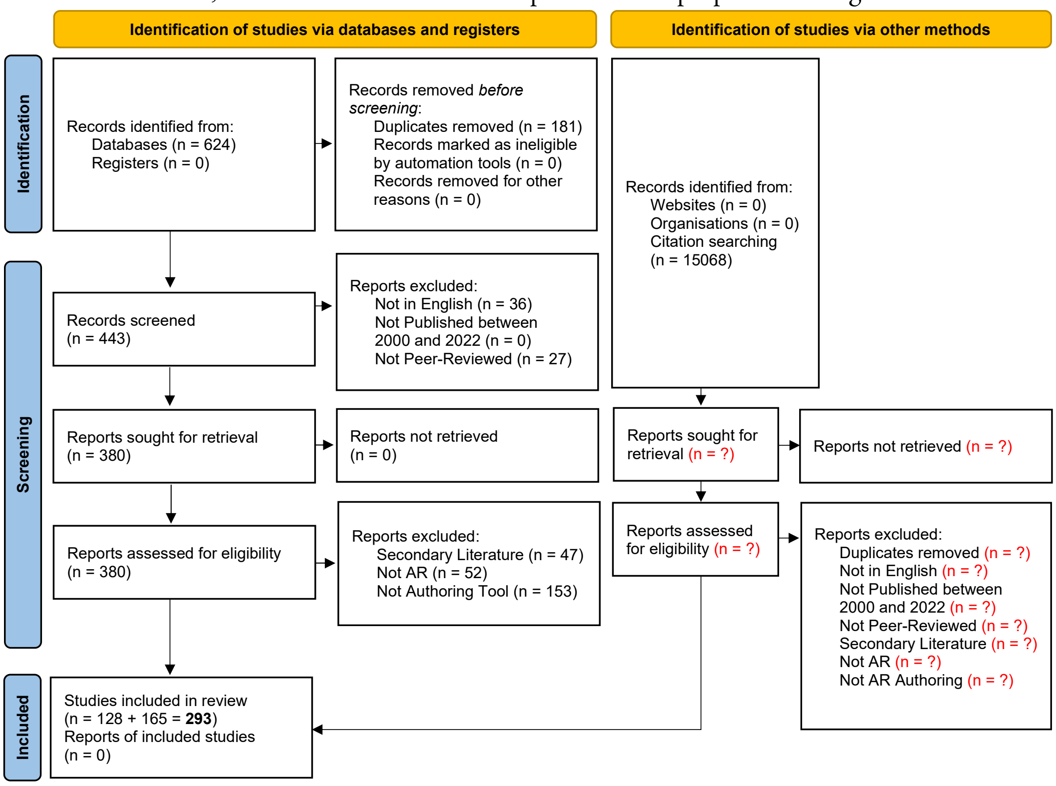


Figure 2.5: The PRISMA 2020 flow diagram [361] with the combined results from the initial database search (with 128 publications, which are fully PRISMA conform) and the subsequent snowballing (165 publications), which cannot be properly reported through PRISMA [361].

It should be noted that there is an extension to the PRISMA guidelines specifically for scoping reviews, the PRISMA-ScR guidelines [461] that are less strict in the search and inclusion process. While the search, screening and review was based on the stricter PRISMA guidelines, relevant items of the PRISMA-ScR guidelines were followed in the mapping stage. The PRISMA and the PRISMA-ScR checklists are reported in Appendix 1 and 2 respectively.

Furthermore, beside the conformity with PRISMA and Wohlins guidelines [504], the review is also in line with the "Guidance on Conducting a Systematic Literature Review" by Xiao et al. [511], a recent, highly influential strategy guide for systematic literature work. While no "quality assessment" of included papers was performed, this is stated as an explicit strategy when conducting a scoping review to "identify the conceptual boundaries of a field, the size of the pool of research, types of available evidence, and any research gaps" [511]. The only design decision not in line with their guidelines is the single-reviewer decision. In contrast to the 2020 PRISMA guidelines, they explicitly suggest performing screening, reviewing and data extraction with "two or more researchers in parallel".

2.4.2 Threats to Validity

When analyzing the map of threats to the validity of systematic literature reviews proposed by Zhou et al. [527], most of the 23 potential threats are well covered through the extensive hybrid search methodology. There should be no publication or culture bias, no incorrect search method deployed, no incorrect usage of terminology, and no primary study duplication. There were no problems in accessing publications or databases, all strategies and results were reported in detail, and the initial search terms performed well (see Section 2.3.7). The non-conformity with PRISMA guidelines of the snowballing is a transparency limitation and explicitly no threat to validity. The remaining threats to the validity of this review are the primary study generalizability, restricted time span, the potential biases in study selection, and potential missclassifications of publications caused by the single reviewer design.

As this is a scoping review with a mapping study, no "quality assessments" (Reporting, Rigor, Credibility, or Relevance) [528] for exclusion of publications were used during the inclusion phase. While this helps to get an overview of the research field, it could distort the perception of the field based on low-quality publications included, that lack generalizability. Another threat to validity is the restricted time span, as only publications from 2000 to 2020 were included in the review. At the time of publication of this thesis, it misses over 2 years of the fast-growing field of Augmented Reality. This could lead to a distorted perception of the current state of the field. Approaches to resolve this threat are discussed in detail after the mapping, in the Future Work Section 2.9.1.

Furthermore, there might be potential biases that threaten the validity of this review, as a single reviewer conducted the screening, review, and mapping. In line with this threat to validity, the single reviewer could also have made human errors during the procedures, especially as the review was conducted over the timespan of several months and 15692 publications were at least screened. Studies indicate that single reviewers have an overall error rate of about 10% [152, 493] in systematic reviews, which is significantly higher than dual-reviewer approaches [152]. But other researchers also found that error rates can be significantly lower than that, if conducted by experienced reviewers and therefore "could still represent an appropriate methodological shortcut in rapid reviews, as long as it is conducted by an experienced reviewer" [482]. Overall, these threats to validity are common in systematic reviews in the field of Augmented Reality and in line with validity threats in recent, influential systematic reviews of the field [105, 156, 362].

2.4.3 Comparison to Overlapping Systematic Reviews

To assess the comprehensibility of this review and partially defuse the threat to validity of conducting this systematic review with only one reviewer, the literature found is compared with literature found in other systematic reviews on the topic of AR authoring tools. To accomplish this, the following five systematic reviews were selected for comparison:

- *"Authoring Tools for Augmented Reality: An Analysis and Classification of Content Design Tools"* by Roberto et al. [400] in 2016
- *"Systematic mapping study on High-level Content Design Frameworks for Augmented Reality"* by Apaza et al. [12] in 2018
- *"A Systematic Review of Rapid Prototyping Tools for Augmented Reality"* by Freitas et al. [143] in 2020
- *"A Design Space of Educational Authoring Tools for Augmented Reality"* by Ez-zaouia et al. [521] in 2022
- *"A Review on Augmented Reality Authoring Toolkits for Education"* by Dengel et al. [104] in 2022

The reviews were published between 2016 and 2022 (see Figure 2.6). Therefore, three of the reviews [104, 143, 521] were conducted and published in parallel and were not present before starting early explorations for this review in 2019. Two reviews were present before conducting this review [12, 400] but one of them was only identified during the screening process of this review itself [12] because of a low citation count and unusual terminology usage.

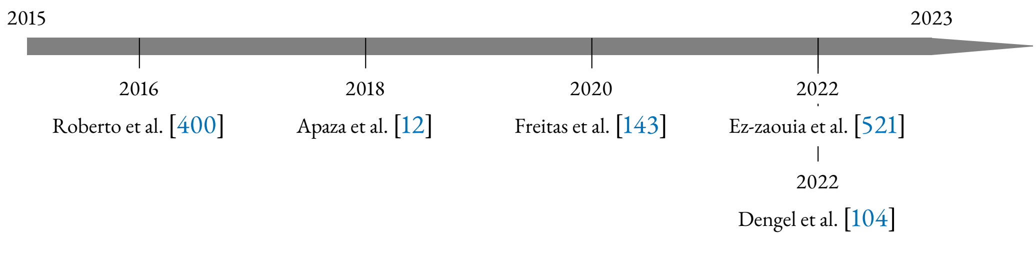


Figure 2.6: The chronological order of the 5 systematic reviews previously published on or closest to the topic of Augmented Reality authoring tools, that this review's results are compared against.

Compared to this review's 293 included publications, all reviews were considerably smaller, with 74 [12], 38 [143], 26 [104], 24 [400], and 21 [521] publications (see Table 2.2). For some reviews, this is partially explained by a smaller scope, e.g., focusing on AR authoring tools in education[104, 521], but generally, this is also likely caused by selected search strategies.

A comparison of the literature included in each review and the overlap with all other reviews is visualized in Table 2.2. Generally, there is very little overlap between the reviews by other researchers, which is likely because of the sizes of the reviews and the inclusion of gray literature. Comparing the overlap between this review and all other reviews, the biggest overlap can be found with the review by Apaza et al.[12]. Here, 56 out of the 74, or 65 out of 74 publications when counting different publications on the same authoring tool as overlap, are also included in this review, resulting in an asymmetrical overlap of 75.68% to 87.84% of their review. Comparing the overlap of this review with the review of Roberto et al. [400], there is an asymmetrical overlap of 54.17% (13 out of 24 publications) of their review. In this specific case, the non-overlapping "publications" are primarily gray literature and references to commercially available tools.

Overlap with	**Blattgerste**	**Apaza et al. [12]**	**Dengel et al. [104]**	**Ez-Zaouia et al. [521]**	**Freitas et al. [143]**	**Roberto et al [400]**
Blattgerste	293					
Apaza et al. [12]	56 (65)*	74				
Dengel et al. [104]	7 (11)*	2	26			
Ez-Zaouia et al. [521]	1	0	0	21		
Freitas et al. [143]	4 (5)*	2	1	0	38	
Roberto et al. [400]	13	11	3	0	1	24
None	**229 (215)***	**17**	**19**	**20**	**34**	**10**

Table 2.2: The overlap of the retrieved literature from this systematic review compared to the 5 most similar reviews previously published. Numbers in brackets, followed by an asterisk (*), represent the overlap, when different publications on the same AR authoring tool are counted as overlap.

When re-evaluating all 183 included publications in the 5 reviews with this review's inclusion criteria stated in Section 2.3.4, 14 AR tools were covered by another publication in this review (inclusion criteria 6). This was the case for 1 publication in Freitas et al.[143], 9 publications in Apaza et al.[12], and 4 in Dengel et al.[104]. 3 Publications by Ez-zaouia et al.[521] were outside this review's specified timeframe (inclusion criteria 2). 49 publications would have violated this review's Inclusion Criteria 3, as they were gray literature, weblinks, or references to commercially available tools. This was the case for 8 publications in [104], 14 in [521], 17 in [143] and 10 in Roberto et al.[400]. Furthermore, 13 publications did not meet this review's definition of AR (inclusion criteria 5), with this being the case for 3 in [12], 1 in [521], 8 in [143], 1 in [400] and 1 in [104]. Finally, 18 publications did not meet this review's stated definition of an AR authoring tool (criteria 6). This was the case for 3 in [12], 6 in [104], 8 in [143], and 1 in [521]. Finally, this review contains 215 authoring tools that satisfied our criteria but were not reviewed by previous efforts.

During this assessment, 4 publications were identified that retrospectively could have been included in this review. 3 publications were part of the review by Apaza et al.[12]: [124, 389, 454], and one was part of the review of Ez-zaouia et al.[282]. 2 of the publications were reviewed, and they were both edge-cases (or "borderline articles" according to Wohlin et al. [506]), but were ultimately excluded [124, 282] through inclusion criteria 6. The other two publications [389, 454] were neither found during the database searches, nor through the snowballing. As only 2 debatable edge-cases and 2 publications that were not officially found through this review's methodology were identified in this assessment, no further publications are included in the review as a result of the re-review.

2.5 Scoping & Mapping Study Methodology

To answer the research questions specified in Section 2.2, a combined methodology of a scoping review and a literature mapping study is employed. While, in the discipline of human-computer interaction, this combination of systematic methodologies is sometimes simply referred to as a systematic literature review, and the methodologies overlap in procedure and goal, they do differ. Most importantly, while goals and research questions were stated, the review does not provide evidence for a specific research question, but rather broadly asks these questions with the aim to scope, synthesize, and summarize the field. This distinction is, e.g., discussed by Soaita et al. [439].

In this chapter, the scoping review methodology component provides a broad overview of the existing literature, synthesizes key concepts and theories, and scopes the landscape of the literature, to identify gaps for potential future work. This part of the methodology is in conformity with the PRISMA-ScR guidelines[461] (see Appendix 2) and serves as the basis for the literature mapping.

With the methodological mapping study component, key themes, differences, expressions of tools, and relationships between different tools and research areas are summarized and visualized. This part of the methodology is thematically based on Peterson et al.[374], where they describe it with: "the main focus here is on classification, conducting thematic analysis and identifying publication fora". While inspired by their methodology, the proposed systematic review process is not utilized, as the process of this review is already based on more established work. The visualization approach of using categorical "bubble" charts is also disregarded, as there are more precise visualization approaches, that are easier to interpret[90]. This is especially true with the multivariate categorical dataset of this review, as the number of dimensions/categories and the differences in their depth, leads to a variety of different expressions, where different visualizations are most appropriate. Furthermore, while some systematic mapping studies exclusively focus on bibliographical mapping, this part is only peripherally addressed. The main focus is on the expressions of the proposed AR authoring tools of the literature, utilizing visualization approaches and exemplary textual descriptions of expressions of the dimensions.

2.5.1 Mapping Procedure

As the review procedures for exclusion and inclusion of publications during the database search (see Section 2.3.3) and snowballing (see Section 2.3.7), were performed by a single reviewer, at the time of completing the review stage of the publications, all 293 publications included were already read at least once. Combining non-representative emerging impressions, notes, and observations during this review procedure with dimensions that overlapping reviews already established as potentially interesting dimensions to map (see Section 2.1), 26 dimensions to map the 293 publications on were defined. As shown in Table 2.3, these dimensions primarily aim to create an understanding of the technical expressions of reported AR authoring tools but also try to map, e.g., how tools are evaluated, and with which purpose and in which context they are used. Furthermore, bibliographic expressions are mapped, e.g., when, where, in which format, and with which publisher they were published. All publications were re-read at least once for the mapping.

Inductive, Deductive, and Descriptive Coding

The mapping of the 26 dimensions across the 293 publications was performed with a coding scheme inspired by the combination of qualitative content analysis by Phillip Mayring [314], the ideation of deploying structure-content analysis methodologies in systematic reviews to understand young research fields by Seuring et al. [425], and coding schemes deployed in some of the most influential recent systematic literature reviews in the research field of AR [233, 319].

As visualized in Table 2.3, in this, the expressions of the dimensions are created through the usage of three methodologies: descriptive, deductive, and inductive coding of schemes. Expressions for 5 dimensions were descriptively mapped; therefore, factual information is transferred from the publication into the multivariate categorical dataset. Furthermore, Expressions of 10 dimensions were deductively mapped, meaning a top-down (or "theory-driven") approach is used, where publications are mapped onto pre-defined levels of the category/dimension. These deductively mapped dimensions were mostly inspired by dimensions of interest brought forward by previous reviews (see Section 2.1) or boolean categories. Finally, the remaining 11 dimensions were inductively mapped. With this bottom-up (or "data-driven") approach, the levels of the categories emerge during the coding process, where first literal descriptions are interpreted and transferred, then the expressions are gradually refined, and then the coding system is iteratively simplified. The simplified theme is then used in the map. This simplification of expressions of the dimensions inherently faces the dilemma of having to decide between precision and conciseness. In this mapping study, schemes were generally simplified until they can be practicably visualized, but particular care was given to not oversimplify expressions, which would lead to loss of information of potential expressions.

All expressions of dimensions were mapped for all publications. When specific expressions were not explicitly stated, they were inferred implicitly from the context, if reasonable. When they could not be inferred from the context with sufficient confidence, they were mapped as "not specified".

Mapping Study Timeframe

The first mapping of the 7618 expressions (293 publications with 26 dimensions to map each) was performed in the timeframe from April 2021 to January 2022. In the subsequent timeframe between January 2022 and March 2023, the dataset was cleaned, and especially the inductively coded dimensions were refined and simplified. In this timeframe, the focus was furthermore on explorative plotting and understanding of the expressions of the multivariate categorical dataset.

Dimension	Coding	Levels	Description
App Relationship	Deductive	External, Internal	Is the authoring tool included in the same application as the AR viewer, or are they separate applications?
Authoring Hardware	Inductive		What hardware is utilized by the authoring tool (e.g., desktop, handheld or web-based authoring)?
Authoring Interactions	Inductive		What interaction concept is used for the authoring?
Authors	Descriptive		Who published the Paper?
Availability	Deductive	Open Source, Available, Not Available	Is the proposed authoring tool actually available, maybe even as an open-source project?
Citations	Descriptive		How often was the paper cited? (At retrieval and 2 years later)?
Construct Author	Inductive		Who is the envisioned author (creator) of the AR construct, and therefore user of the AR authoring tool?
Construct Distribution	Deductive	Local, Server	Are the produced AR constructs distributed locally or through a server?
Construct User	Inductive		Who is the targeted user of the viewer tool for the authored AR construct?
Content Sequentiality	Deductive	Sequential, Static	Is the content static content or a procedural sequence of different AR content?
Content Type	Inductive		Which modality (e.g., 3D Models, Text, Audio) are used for the AR content?
Contribution	Deductive	Main, Secondary	Is the main contribution of the paper the authoring tool itself, or is the authoring tool a supplement to the main contribution?
Deployment Context	Inductive		In what context (e.g., medicine, industrial assembly, maintenance) is the tool used?
Deployment Purpose	Inductive		What purpose is the AR content authored for? E.g., to assist people, for learning or entertainment?
Format/Type	Deductive	Journal, Paper, Short, Poster	Is the paper a journal article, conference proceeding (long & short) or a poster?
In-Situ Authoring	Deductive	Yes, No	Is the content authored "in-situ", therefore directly in the augmented context it is later used in?
Markup Notation	Inductive		Does the authoring tool utilize or propose a scene description, or markup language (e.g., XML, JSON)?
Modularity	Deductive	Standalone, Plugin	Is the proposed authoring tool a plugin for a host software as defined by Roberto et al.[400] or a standalone application?
Publication Year	Descriptive		In which year was the paper published?
Publisher	Descriptive		Which publisher was used for the publication?
References	Descriptive		How many references are included in the paper?
Scene Preview	Deductive	Yes, No	Does the authoring tool include a 3D preview of the created AR content?
Tracking Type	Inductive		What tracking technologies (e.g., marker, QR, feature points) are used by the AR hardware?
Usability Evaluation	Deductive	Authoring, Usage, Both, None	Is the Usability of the authoring process or the usage of the authored AR constructs evaluated in the paper?
Usage Hardware	Inductive		What hardware is used to display the authored AR construct (e.g., desktop, handheld, HMD)?
User Interactions	Inductive		What interaction concept is used for viewer application?

Table 2.3: The 26 primary dimensions, the literature, is mapped on. Each is shown with the coding scheme used (descriptive, deductive, or inductive), the levels for the deductively coded themes, and descriptions of the questions this dimension tries to answer. The table is sorted alphabetically and colored by the coding scheme used.

2.6 The Literature Map of AR Authoring Tools

The following 26 subsections describe the dimensions of the literature map of AR authoring tools on an abstract level to provide a general overview of the field of AR authoring tools. During this, dimensions are gradually introduced, all expressions of each dimension are visualized and for some dimensions, exemplary implementations are textually described with references to the included literature for clarification. This procedure is in line with comparatively comprehensive systematic reviews in the field of AR research [105]. After dimensions are introduced, they are occasionally reused as breakdown dimensions in the introduction of subsequent dimensions to anecdotally visualize key trends. Additionally, exploratory association analyses are performed at times to analyze these trends. It's important to note that these analyses should neither be interpreted as claims about causation, nor as claims about associations at the population level, such as predicting features of future tools; they are purely exploratory, descriptive, and specific to the multivariate data set at hand. The Appendix 3 of this thesis includes the complete set of tables, specifying which publications were mapped to which expression of each of the dimensions. Additionally, the results of the mapping study are published in a CC-BY 4.0 licensed multivariate dataset [49].

2.6.1 Publication Year

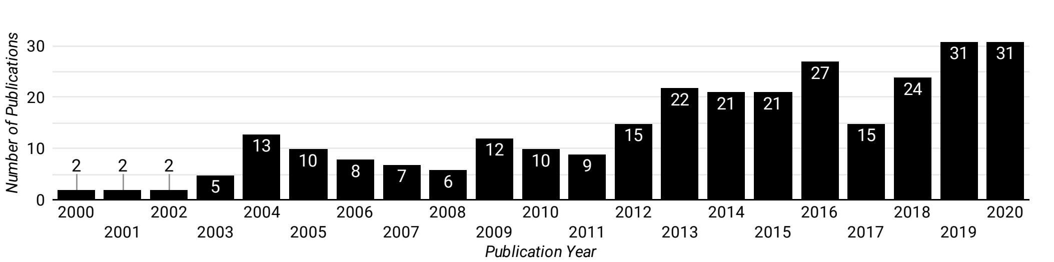


Figure 2.7: The 293 publications contributing an AR authoring tools, grouped by publication year.

As can be seen in Figure 2.7, and is referenced to each publication in Appendix Table 4, there is an increasing number of published papers on AR authoring tools throughout the reviewed 20 year timespan. Moverover, goodness of fit (R^2) can be used to assess this growth trend of published papers. As the R^2 value represents the proportion of the variance that is accounted for by the model, it indicates how well models explain the observed trends in the data. If the goodness of fit for the linear and exponential case is tested, they yield results of $R^2 = 0.82$ and, $R^2 = 0.83$ respectively, indicating there to be either linear or exponential growth in efforts. Then, if *intitle:"Augmented Reality"* as a search string is used in Google Scholar for the same period, a non-representative, order-of-magnitude estimation can be calculated for the overall observed growth trend of AR efforts in the literature [1]. Here, the linear model yields results of $R^2 = 0.85$, while the exponential model yields $R^2 = 0.99$, strongly indicating exponential growth of overall efforts of the field. In this, it appears that the overall efforts in the field of AR are likely outgrowing the efforts regarding AR authoring tools.

2.6.2 Authors

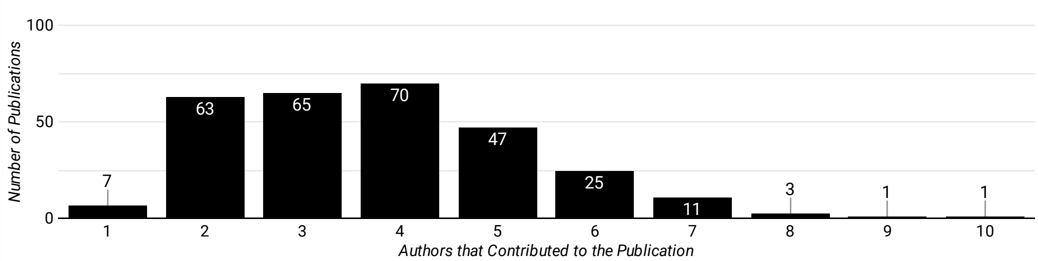


Figure 2.8: The 293 publications by number of authors who contributed to the publication.

Visualized in Figure 2.8 and referenced in Appendix Table 5 are the number of publications mapped onto how many authors contributed to the manuscript. On average, 3.8 (SD = 1.57, Mdn = 4) authors contributed to the 293 publications. Overall, 1112 author names are listed on the publications, of which 828 were individual researchers. The 6 researches who contributed to the most publications on AR authoring tools covered in the review timespan were Prof. Mark Billinghurst (Universities of South Australia & Auckland), who contributed to 12 publications [42, 182, 192, 204, 266, 267, 338, 379, 380, 422, 472, 491], Prof. Woontack Woo (Korea Advanced Institute of Science and Technology), who contributed to 8 publications [180, 181, 231, 234, 365, 367, 432, 519], Prof. Dieter Schmalstieg (Graz University of Technology), who contributed to 8 publications [29, 250, 251, 252, 253, 254, 262, 265], Prof. Tobias Langlotz (University of Otago), who contributed to 7 [250, 251, 252, 253, 254, 255, 491], Prof. Blair MacIntyre (Georgia Institute of Technology) with 6 publications [195, 249, 296, 297, 298, 388], and Prof. ONG Soh Khim (National University of Singapore) with 6 publications [190, 213, 355, 530, 531, 532].

2.6.3 Publication Format/Type

Figure 2.9 (referenced in Appendix Table 6) shows the publications mapped onto the format they were published in on the left. 73.4% of publications were published in conference proceedings, of which the majority were published as full papers (42% of overall publications), followed by short papers (21.5% of overall publications), and finally posters (9.9% of overall publications). Only roughly a quarter of publications (26.6% of publications) were published in journals.

[1]Therefore restricting the searches to one year each, resulting in 233 (2000), 325 (01), 395 (02), 469 (03), 558 (04), 599 (05), 585 (06), 697 (07), 743 (08), 863 (09), 1240 (10), 1620 (11), 1900 (12), 2140 (13), 2280 (14), 2470 (15), 2970 (16), 3680 (17), 4610 (18), 5380 (19), and 5440 (2020)) results found.

The prominence of conference proceedings as a publication format is not unusual in the field of HCI. Premier conferences in HCI not only match top journals in the number of published papers, but are also on par with them in terms of their theoretical impact factor [317]. Nonetheless, as can be seen on the right side of Figure 2.9, there is an increasing trend of publishing papers through academic journals.

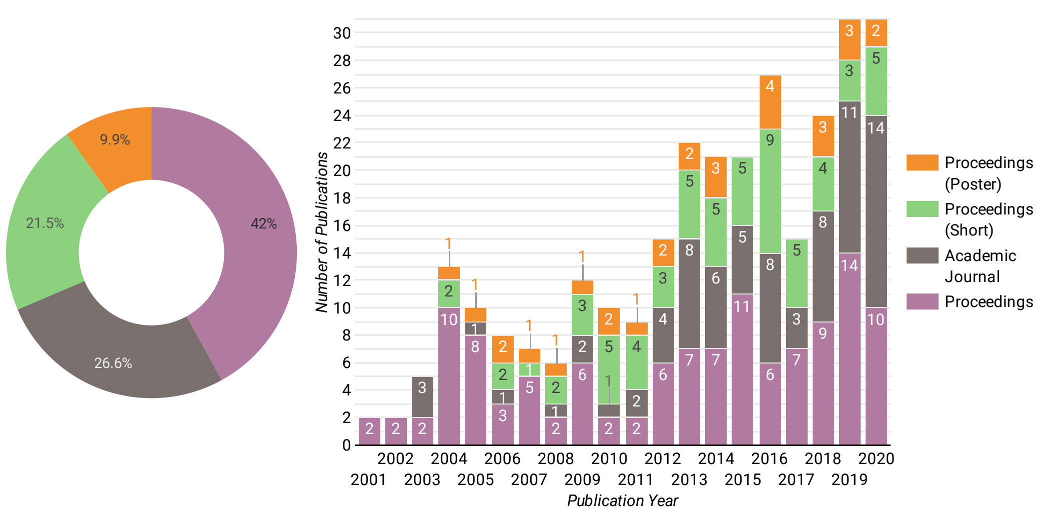


Figure 2.9: The 293 publications, split into their publication format as a donut chart (left) and used as a breakdown dimension when plotting the number of publications for each year (right)

2.6.4 Publisher

As can be seen in Figure 2.10 and is referenced to each publication in Appendix Table 7, publications are generally distributed across a wide range of publishers. As can be expected, the majority of publications are published in IEEE or ACM, generally regarded as the premier publishers for the field of AR/VR research, as they publish the premier conferences and journals [317]. When only considering publishers with more than 5 publications and grouping the remaining publications as "Others", it is furthermore apparent that publications that are neither published in ACM, IEEE, Springer, or Taylor & Francis, are generally spread across different publishers and subsequently also a diverse set of conferences and journals. This is likely caused by the inherent interdisciplinary perspectives involved in authoring tools, as results can be either reported from the technical perspective or from the perspective of the context, making the publications also eligible to be published in the journals of other fields. While this is generally welcomed from the perspective of involving the scientific communities of those fields into the development process, this makes publications challenging to identify.

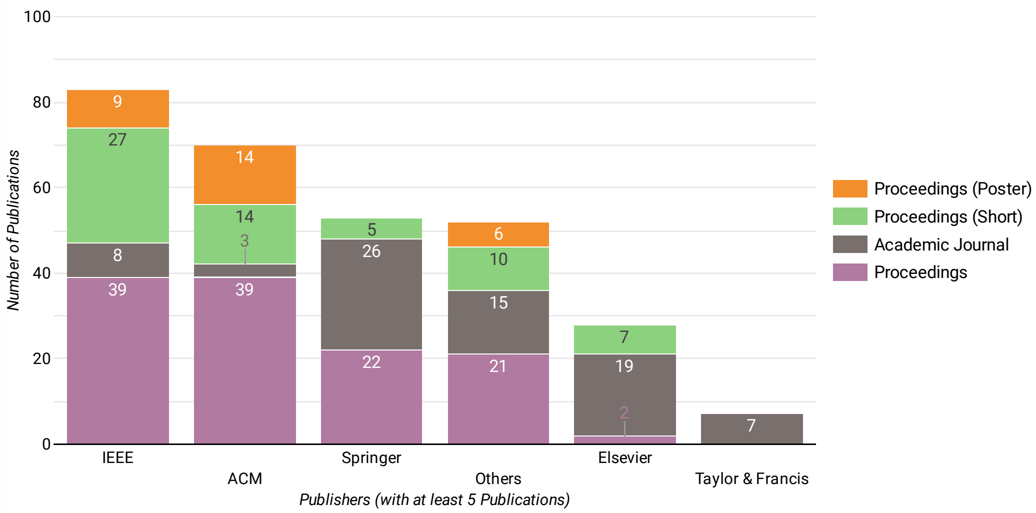


Figure 2.10: The 293 publications mapped onto the 5 publishers with the most publications. Publications that were published through other publishers were grouped as “Others”.

2.6.5 Contribution

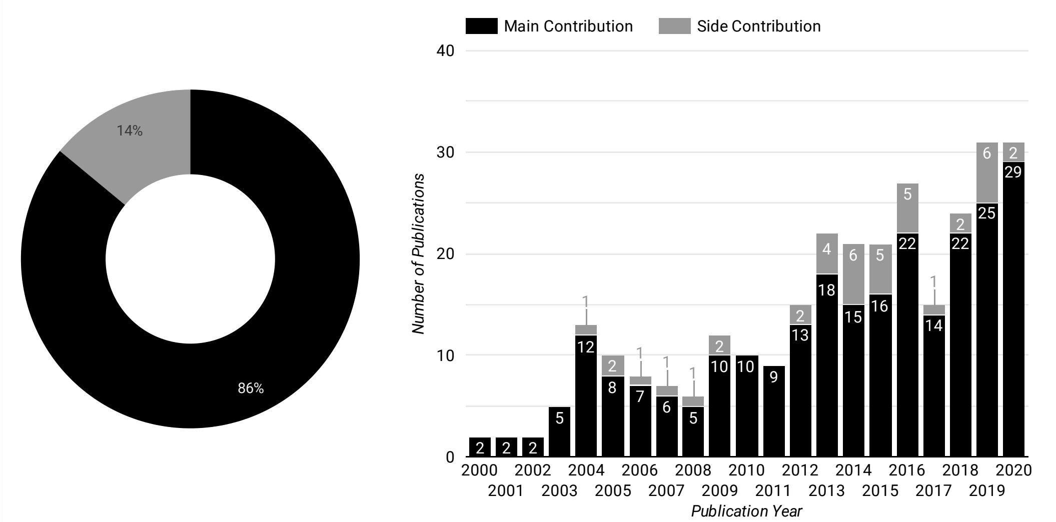


Figure 2.11: The 293 publications, split into whether the AR authoring tool is the main or a side contribution of the publication (left) and the trend of this distinction over the years (right).

Likely the case for similar reasons, an additional challenge in identifying publications concerned with AR authoring tools is the occurrence of publications which report on AR authoring tools, but only do so as a side contribution and not as the primary contribution of their work. While, as is apparent in Figure 2.11 and references in Appendix Table 8, this is not common in the mapped literature, as it is only the case for 14% of the publications, these publications, and their contributions are especially challenging to identify. Examples of these publications are researchers proposing a usability questionnaire and developing an AR authoring tool for annotation purposes to have a complex system to evaluate their questionnaire with [414] or researchers proposing an AR tutor system for educational purposes, where the authoring component is only one contribution [294]. Notably, more than half (51.22%) of the publications reporting AR authoring tools as side contributions were published in academic journals.

This dimension was primarily mapped to inquire if this was an increasing trend throughout the years, which could have indicated maturity of the field, which was not the case.

2.6.6 Citations

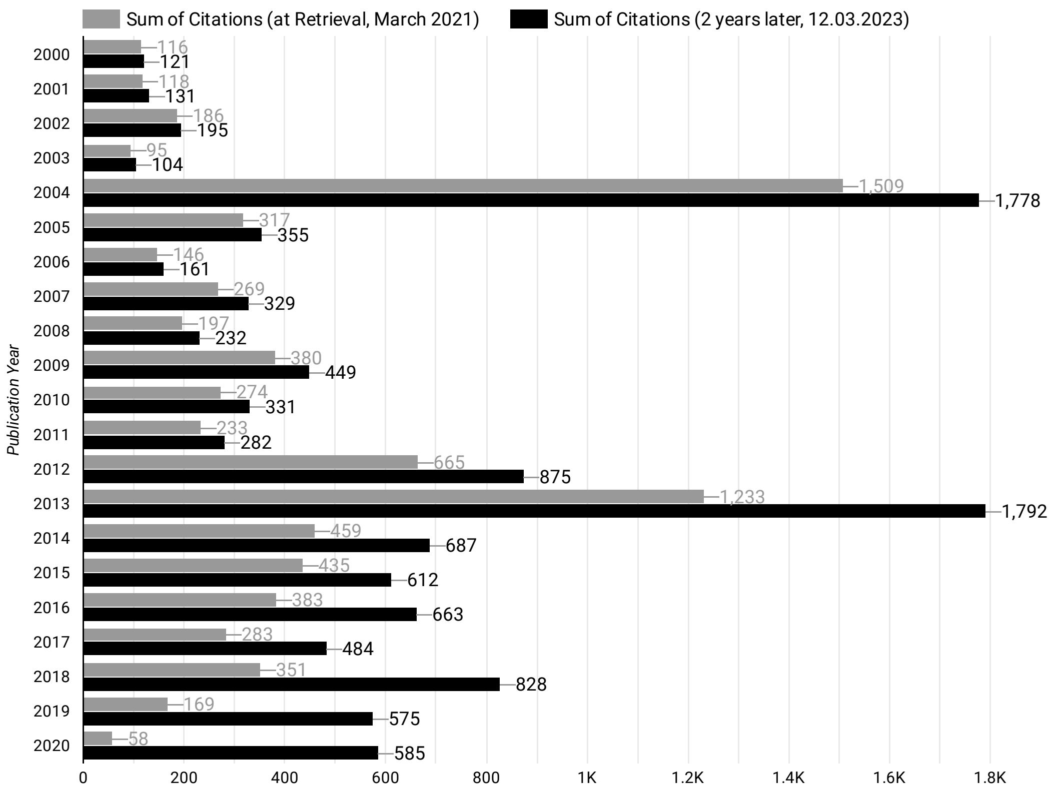


Figure 2.12: The sum of citations at retrieval in March 2021 and 2 years later in March 2023, grouped by publication year of the literature.

In March 2021, the 293 publications were cited 26.88 (SD = 53.54, Mdn = 10) times on average. Two years later, in March 2023, the average citation count was 39.48 (SD = 71.98, Mdn = 19). Figure 2.12 shows the sum of citations at retrieval in March 2021 and the sum of citations 2 years later, in March 2023, grouped by publication year of the publications. Two insights are apparent. Firstly, there seem to be two spikes in the sum of publications in 2004 and 2013. This is caused by 5 highly influential publications in 2004 [266, 283, 296, 508, 522] and 2 highly influential publications in 2013 [494, 507]. Secondly, though not surprising as there were also more publications published, recent publications received the majority of the recorded citation growth over the 2 year timespan.

The 5 most cited publications in March 2023 were cited 746 [507], 494 [508], 433 [494], 414 [296], and 287 [283] times. Notably, the 3 most cited publications all reported AR authoring tools as a side contribution [494, 507, 508], and are also the publications which caused the spikes in citations when grouping them by publication year. The 5 most cited authors were Dr. Rafal Wojciechowski with 1276 citations across 4 publications, Prof. Wojciech Cellary (Poznan University of Economics and Business) with 1243 publications across 3 publications, Prof. Mark Billinghurst (Universities of South Australia & Auckland) with 832 citations across 12 publications, Prof. Martin White (University of Sussex) with 805 citations across 5 publications, and Prof. Blair MacIntyre (Georgia Institute of Technology) with 706 citations across 6 publications.

2.6.7 References

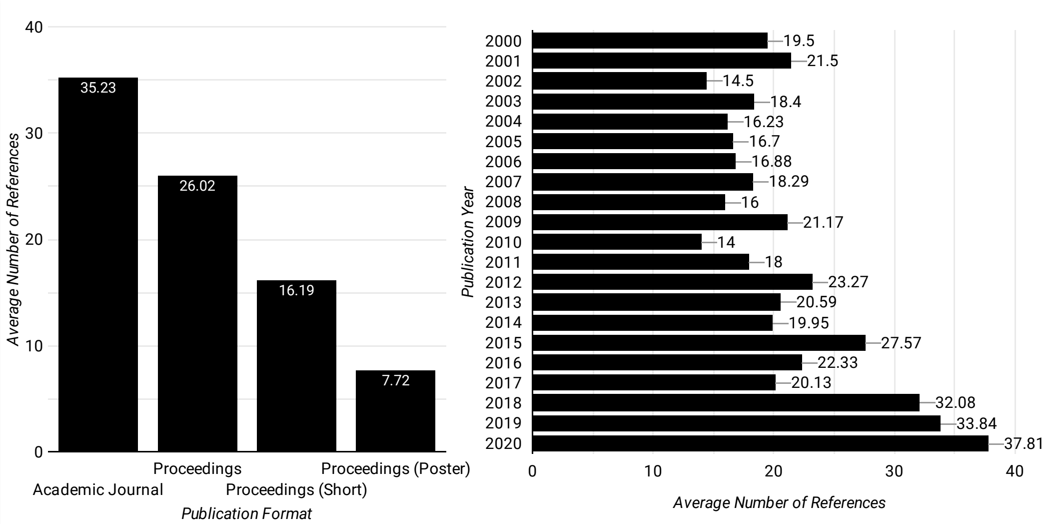


Figure 2.13: The average number of references in a publication, grouped by publication format (left) and the publication year (right)

On average, publications published in academic journals had 35.23 (SD = 19.7, Mdn = 23) references, publications in conference proceedings had 26.02 (SD = 13.78, Mdn = 23) references, and short papers had an average of 16.19 (SD = 9.83, Mdn = 14) references, while poster papers only averaged 7.72 (SD = 3.8, Mdn = 8) references per publication (see Figure 2.13, left). A significant medium correlation (r(291) = 0.34, p =< 0.001) is apparent, where through the years, more references are incorporated in the publications (see Figure 2.13, right). But this trend is also associated with the increased occurrence of journal publications on the topic (see Figure 2.9), which is apparent through association analysis ($\eta = 0.51, \eta^2 = 0.26$), which indicates that approximately 26% of the variance in the number of references can be explained by the publication format.

2.6.8 Deployment Purpose

When mapping the inductively coded dimension of the deployment purpose of the AR constructs created with the AR authoring tools, five general purposes can be found: Assistance, Entertainment, Prototyping, Learning, and Multipurpose or tools that do not specify a specific purpose. As can be seen in Figure 2.14 (top left) and is references in Appendix Table 9, the 293 publications are generally evenly distributed among the 5 deployment purpose groups, ranging from 17.4% of publications creating AR constructs for the purpose of learning to 23.5% of publications which create constructs for the purpose of Assistance.

An example for the assistance purpose would, e.g., be an AR authoring tool to create AR instructions for furniture assembly [522]. Deployed AR constructs for the purpose of learning are, for example, used to teach maintenance procedures to new workers [494], and authoring tools that create AR constructs for entertainment purposes, e.g., create AR games [182], or create interactive museum exhibits [239]. The deployment purpose of prototyping could for example be an AR authoring tool that is used for factory layout planning [372].

Visually inspecting Figure 2.14 (top right), there seems to be only slight associations between the deployment purpose and where, when only considering the top 3 publishers, publications are published. But, the deployment purpose, for example, is most prominently published in ACM, while the multipurpose and assistance purpose is most prominently published in IEEE. As can be seen in Figure 2.14 (bottom), all deployment purposes seem to roughly grow at a comparable rate. The only exception is the “multipurpose” expression, likely has with increasing efforts, more specific purposes are explored.

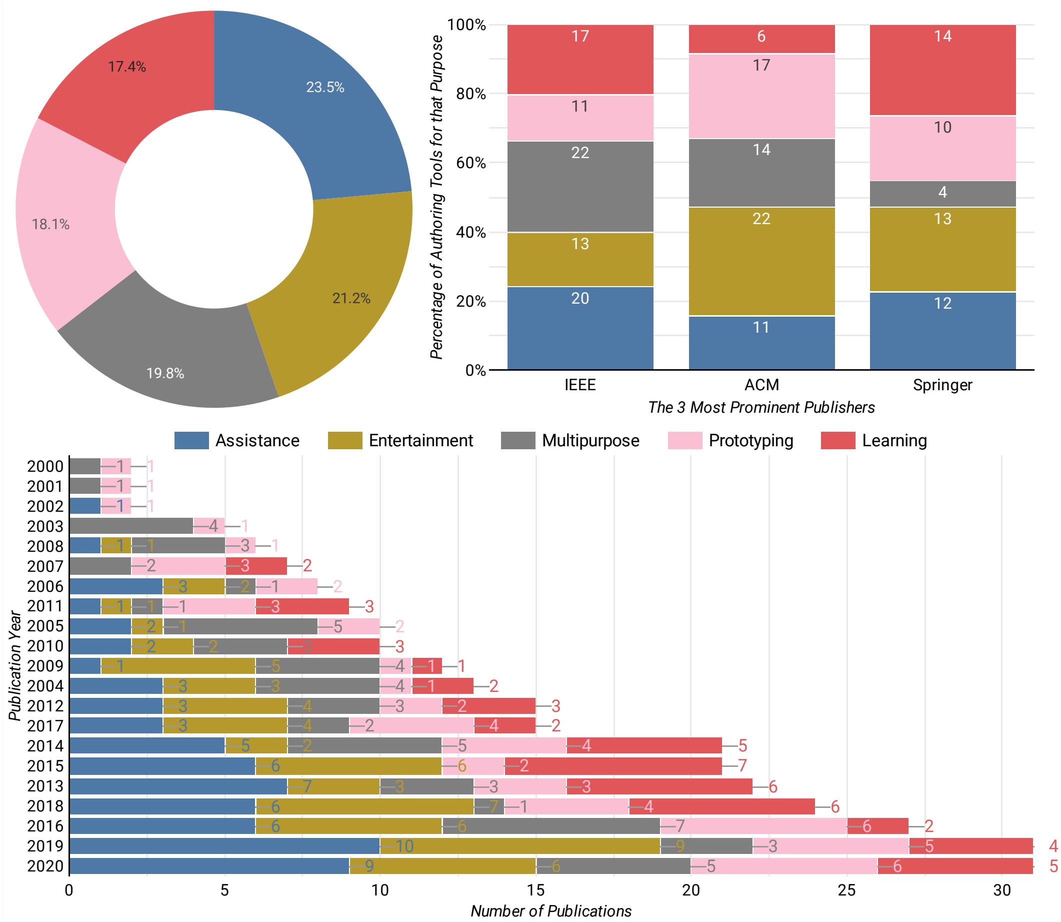


Figure 2.14: The 293 publications split into the deployment purpose the AR constructs are authored for (top right), and the publisher they are published with (top left). The bottom graph shows the trend of deployment purposes over the years.

2.6.9 Deployment Context

Visualized in Figure 2.15 and referenced in Appendix Table 10, are the grouped and summarized contexts which were visible in the literature. Without fully describing each context, this ranges from AR authoring tools to create industrial assembly instructions or learning content, over the usage of AR authoring tools to create AR games for entertainment purposes, to finally using AR authoring tools to create AR-based military training material. In Figure 2.15, the grouped deployment contexts are furthermore broken down into the deployment purpose. As expected after visual inspection, there is a significant strong association between the deployment purpose and the deployment context of authored AR constructs, when analyzing the relationship using Cramér's V, based on the Pearson's chi-squared test ($\chi^2(60) = 638.13, p < 0.001$, Cramér's V

= 0.74) [2]. This strong association could have been induced through a bias in the simplification of the deployment context, but is likely an actual, inherent relationship between the deployment context and purpose dimensions.

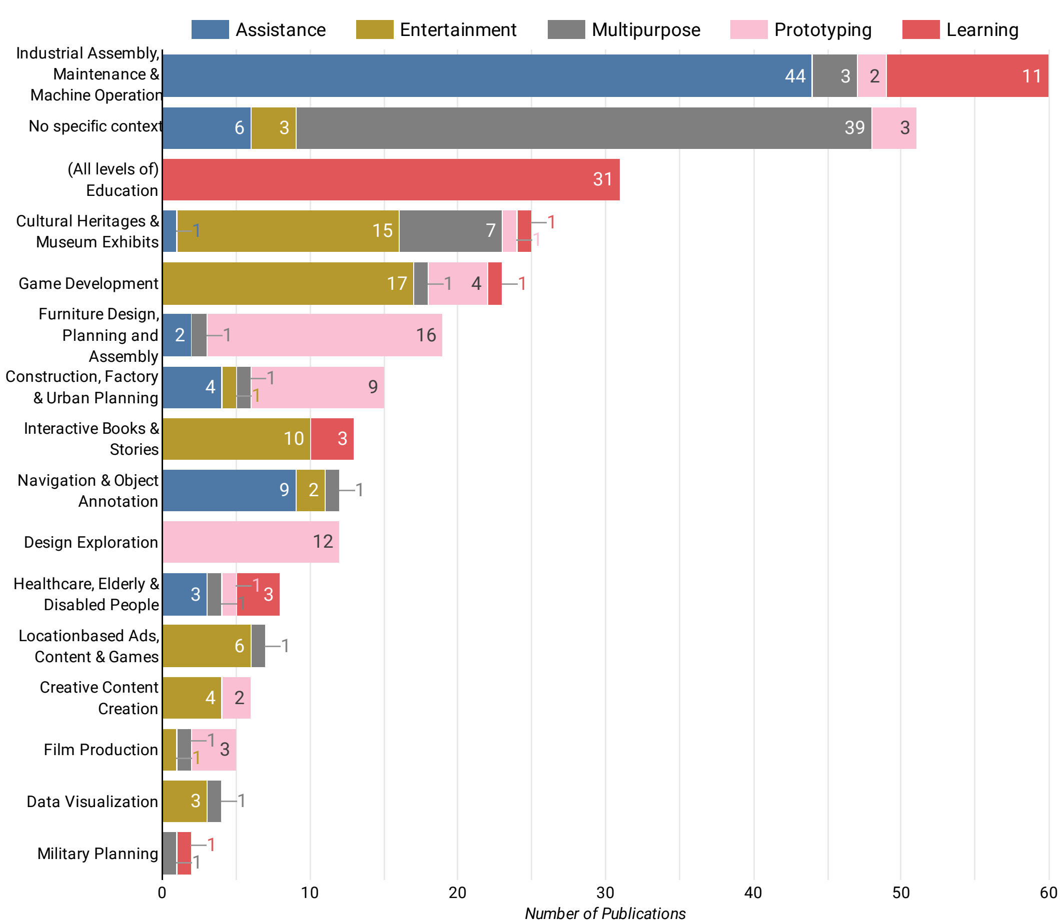


Figure 2.15: The 293 publications, grouped by deployment contexts the AR constructs created with the AR authoring tool are intended to be used in, broken down into the deployment purposes in that context.

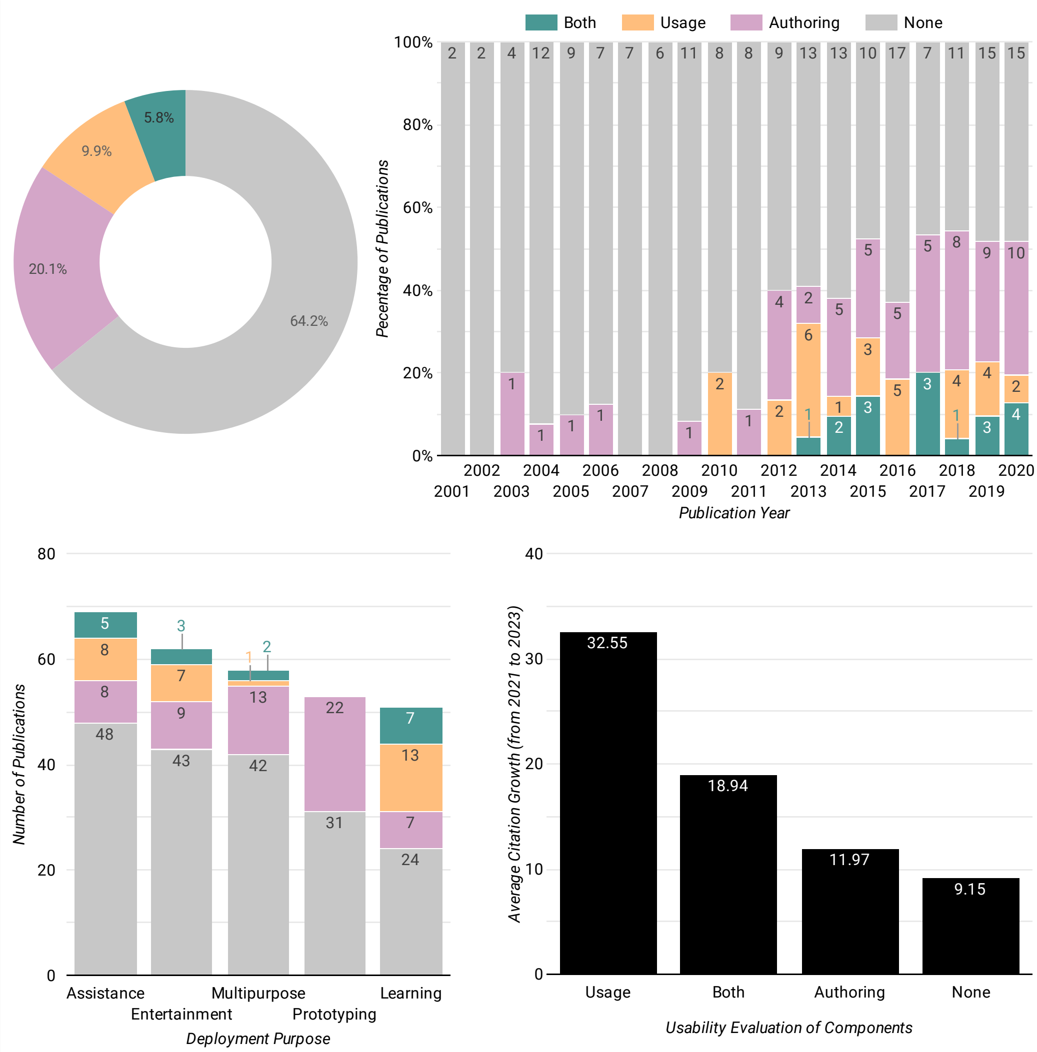


Figure 2.16: The 293 publications split into whether they evaluate no aspect, the authoring perspective, the usage of authored AR construct perspective or both (top left). The graph on the top right visualizes this trend over the publication years. The bottom left graph shows the evaluation efforts grouped by the deployment purpose, and the bottom right graph shows the association between the average citation growth over two years and the evaluation efforts.

2.6.10 Usability Evaluation

As can be seen in Figure 2.16 and is referenced for each publication in Appendix Table 11, the majority of publications on AR authoring tools neither evaluate the usability of the authoring tool, nor the usability of the authored construct (64.2%). Only 20.1% of publications evaluated the authoring component, 9.9% the usage of the authored construct and 5.8% both perspectives (top left). This result is in line with findings by Dengel et al. [104], who found that "regarding the toolkits' evaluation, more than half of the papers (17) do not report any evaluation methods" in the educational AR authoring tool context in a previous review and is likely at least partially because of the complexity of the multi-layered perspectives to evaluate for AR authoring tools [166].

As can be seen on the top right of the Figure 2.16, over the years, there is an increasing trend of evaluating tools, but there still seems to be primarily technical focus. On the bottom left, differences in evaluation efforts between deployment purposes are visualized. As can be expected, authoring tools for the purpose of prototyping are evaluated in terms of the authoring component more often, and tools for the purpose of learning are generally evaluated the most, with most evaluation efforts focusing on the usage of authored AR constructs.

Besides the trend of increasing endeavours over the years of reporting evaluation efforts as a contribution of the publication of the AR authoring tool, there also appears to be a tendency for increased citation growth for publications, which reported evaluations. As can be seen on the bottom right of Figure 2.16, the citations of publications which reported usability evaluations on the Usage of authored AR constructs grew by 32.55 (SD = 56, Mdn = 8), while publications reporting evaluations on both the authoring and the usage grew by 18.94 (SD = 20.19, Mdn = 7), and publications reporting evaluations of only the authoring perspective grew by 11.97 (SD = 11.94, SD = 8). Over the two year timespan, publications, which reported no evaluation efforts, only had an average citation growth of 9.15 (SD = 14.13, Mdn = 4).

2.6.11 Availability

While there are many proposals of AR authoring tools in the 293 publications, 272 (92.83% of the mapped literature) are not actually available, even when searching for them through external search engines like Google. Only 8 AR authoring tools are available as binaries (2.73%) and 13 are available as Open Source (4.44%) (see Appendix Table 12). The 13 AR authoring tools, which are available as open-source, are detailed and linked in Table 2.4. While many of the publications likely evaluated the feasibility of specific implementation aspects, contexts, or human factors, when combining this with the also relatively low percentage of evaluations reported in the literature (see Figure 2.16), this observation inevitably raises questions about this discrepancy between the number of proposals and actual availability of AR authoring tools and implies a gap

[2] Technically, in the following, requirements for the χ^2 test for Cramér's V are often not fully met, as the contingency tables have values below 5, which can influence the test's validity and expedite Type I errors. Cramér's V is used in the following to explore associations regardless because of its practicality and descriptive purposes and with it, the χ^2 test statistics are reported for transparency, despite their limited validity.

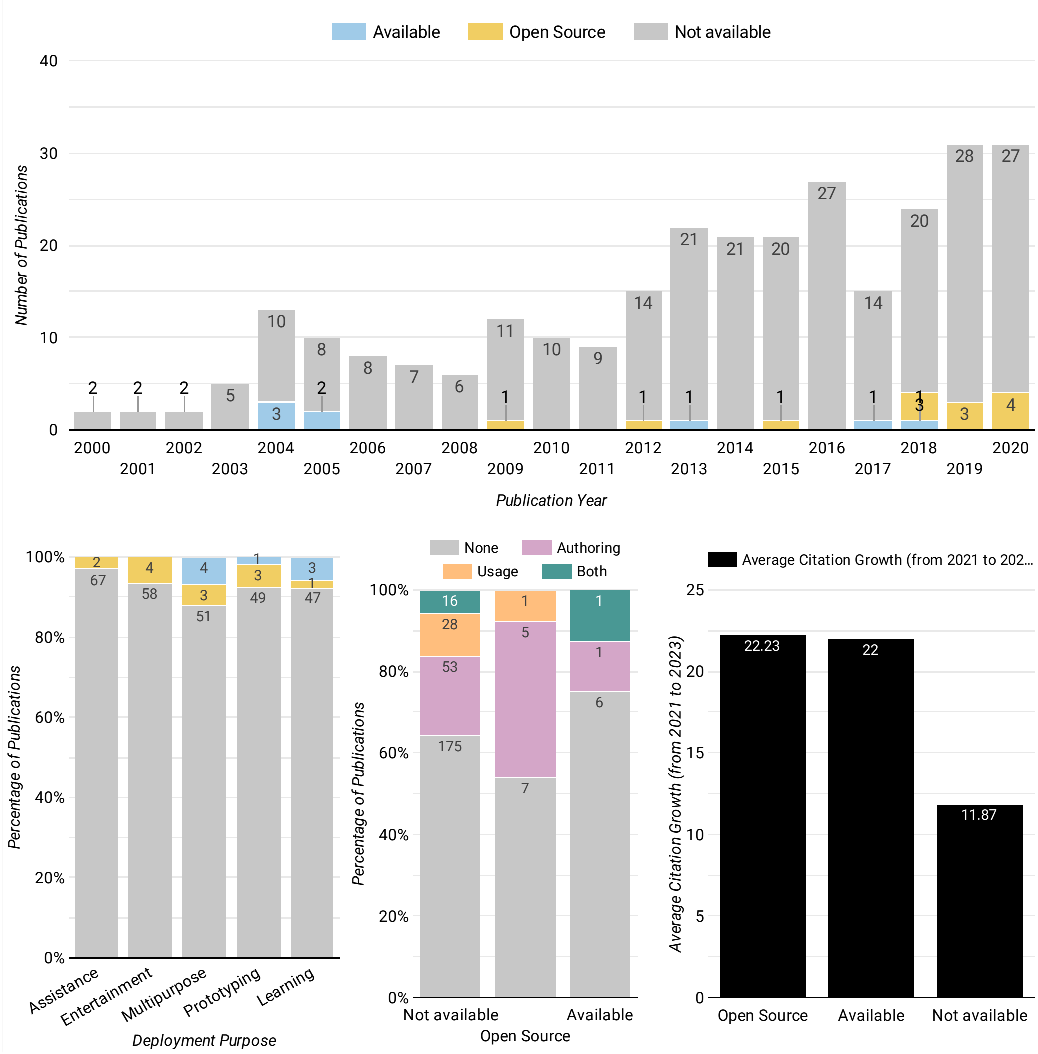


Figure 2.17: The 293 publications, grouped by the publication year and broken down into whether they are actually available as binaries or open source (top). The bottom left graph shows the publications grouped by the deployment purpose, broken down by the availability. The bottom middle plot shows the publications, grouped by availability, broken down by which aspect of the AR authoring tool was evaluated. The bottom right plot shows the publications, grouped by availability, with their average citation grows over the two year timespan.

in the literature, or rather necessitates of shifting of the focus of current efforts. This percentage of actually available AR authoring tools is also not in line with findings in previews reviews. For example, Dengel et al. [104] found that "of this total of 80 AR authoring toolkits (shown in Supplementary Appendix S1B) only a bit over a third (36.8%) were openly accessible".

Publication	**Year**	**Last Commit**	**Link**
Radu et al. [388]	2009	03.2012	https://github.blairmacintyre.me/site-archive/ael-2015/research/authoring/arspot/
Hardy et al. [186]	2012	19.11.2013	https://code.google.com/archive/p/ubidisplays/
Feuerstack et al. [139]	2015	05.07.2020	https://github.com/sebastiangtts/ar-product-configurator
Kelly et al. [225]	2018	05.11.2019	https://github.com/LaboratoryForPlayfulComputation/arcadia
Sicat et al. [436]	2018	12.04.2023	https://github.com/ronellsicat/DxR
Haynes et al. [189]	2018	03.06.2020	https://github.com/landscapear/wsiarapp
Apaza et al. [13]	2019	19.07.2019	https://github.com/yg-apaza/simplear-editor, https://github.com/yg-apaza/simplear-viewer
Wang et al. [492]	2019	31.05.2021	https://github.com/zachzeyuwang/AniCode
Hodaie et al. [196]	2019	27.08.2020	https://github.com/zardosht/isar
Nguyen et al. [346]	2020	14.05.2020	https://github.com/Alex-Nguyen/BlocklyAR
Reipschläger et al. [395]	2020	18.11.2020	https://github.com/imldresden/u2vis
Chen et al. [82]	2020	30.09.2020	https://github.com/PapARVis
Gottschalk et al. [171]	2020	03.07.2020	https://github.com/SebastianGTTS/ar-product-configurator, https://github.com/SebastianGTTS/feature-modeler

Table 2.4: The 13 AR authoring tools which are available as open source, with their publication year, date of last commit (accessed on the 22.06.2023) and link to their repositories.

As visualized in Figure 2.17 on the top half of the plot, there is at best a very slight trend of increasing open-source efforts in the timespan of 2018 to 2020, but as visualized on the bottom left graph, this trend is generally evenly split between the deployment purposes. As visualized in the middle bottom graph, open-source tools are also evaluated the most, with a focus on the authoring perspective. Furthermore, almost all tools, which were evaluated for the usability of the created AR construct, are neither available nor open-source. But, in line with the citation growth for reporting evaluation efforts in the publication, publishing the AR authoring tool does increase the average citation growth. Tools that were open-source or available as binaries had an average citation grows of 22.23 (SD = 24.34, Mdn = 11) and 22 (SD = 25.53, Mdn = 9.5) over the two years respectively, while tools that were not available only had an average citation growth of 11.87 (SD = 22.79, Mdn = 6). Tools that reported neither an evaluation nor are available in any form received the lowest citation growth of any combination of the two dimensions, with an average of 8.23 (SD = 12.17, Mdn = 4). There was no tool which evaluated both the authoring and the usage perspective and was available as open-source.

2.6.12 Construct Author

The envisioned construct author was one of the more challenging dimensions to identify but, as can be seen in Figure 2.18 and is referenced for each publication in Appendix Table 13, there are some insights. Firstly, about 20.82% of publications neither explicitly specify who the envisioned author is, nor can it reasonably be deducted from the context or descriptions in the publication proposing the AR authoring tool. As can be expected, this was the case for most tools which did not specify a specific deployment purpose of the AR constructs, and were more focused on the technical exploration of AR authoring tools. Besides these, 21.16% of the AR authoring tools are intended to be used by end users, 17.75% by designers, 13.31% by teachers and trainers, 6.14% by domain or task experts, 5.12% by engineers, and the remaining tools by other, more specific, potential user groups. Visually inspecting the figure reveals that the author of constructs for the purpose of learning are primarily created by teachers or trainers, while AR constructs for the purpose of prototyping are predominantly envisioned to be authored by end users or designers. But this mapping is likely substantially influenced by the fact that, for example, "teachers" being the author of an AR construct for the purpose of learning is more self-evident than creating a detailed vision of whom a prospective user for an AR authoring tool is, that is used to assist workers in the maintenance of complex machinery. Another aspect, which was not anticipated in the methodology for the mapping and therefore is hidden in the expressions of the dimension, is the incorporation of multiple stakeholders which are actively involved in using the AR authoring tool.

Ultimately, while this dimension was one of the more challenging dimensions to identify expressions for during the mapping process, this challenge of mapping who the AR constructs author and therefore the user of the proposed AR authoring tool is actually supposed to be, is likely the more important insight gained from trying to inductively map the literature based on this distinction, compared to identifying which AR construct author groups do exist in the literature. As will be discussed in Section 2.7, researchers contributing AR authoring tools, should clarify this human factor consideration in their publication.

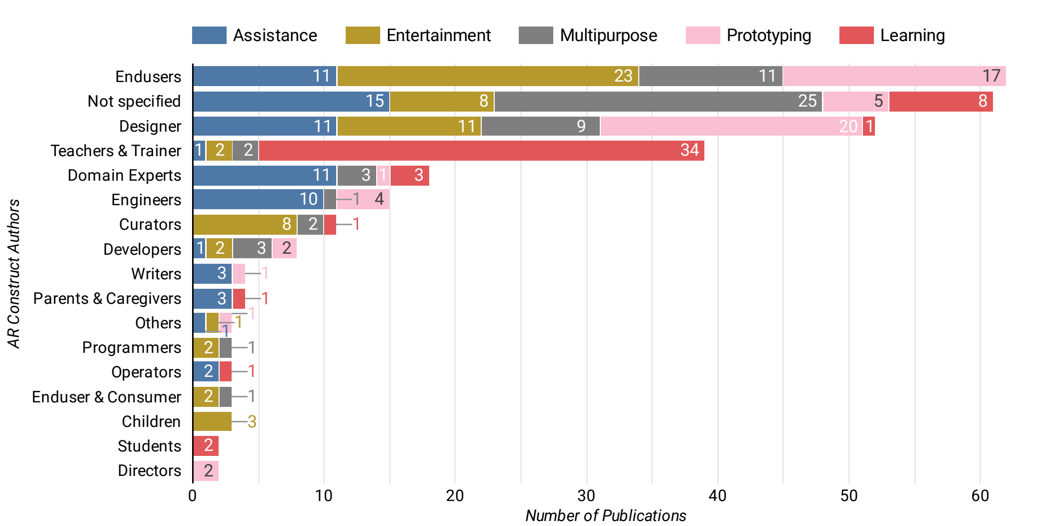


Figure 2.18: The 293 publications, grouped by the envisioned author, who uses the AR authoring tool to create an AR construct, broken down by the deployment purpose of the construct.

2.6.13 Authoring Hardware

As some proposed AR authoring tools work on multiple hardware choices (5.12% of publications), Figure 2.19 visualizes both, all supported AR hardware choices (top), and the one-to-one mapping to the hardware choice with "Multiple Device Types" as one of the hardware choices (bottom, also see Appendix Table 14). Generally, as can be seen in the figure, the most commonly used hardware platform used for the authoring of the AR constructs are Desktop PCs, with over half of all tools supporting this hardware platform (50.85% of proposed AR authoring tools), followed by handheld devices, which are supported by 29.35%, HMD-based approaches (12.63%), web-based approaches which should also be platform-independent (9.56%), VR-authoring of AR content (2.05%), and projection-based hardware for AR authoring (1.37%). When visually inspecting the bottom graph of Figure 2.19, there appear to be only slight tendencies of an association of a specific hardware choice based and the deployment purpose of authored constructs.

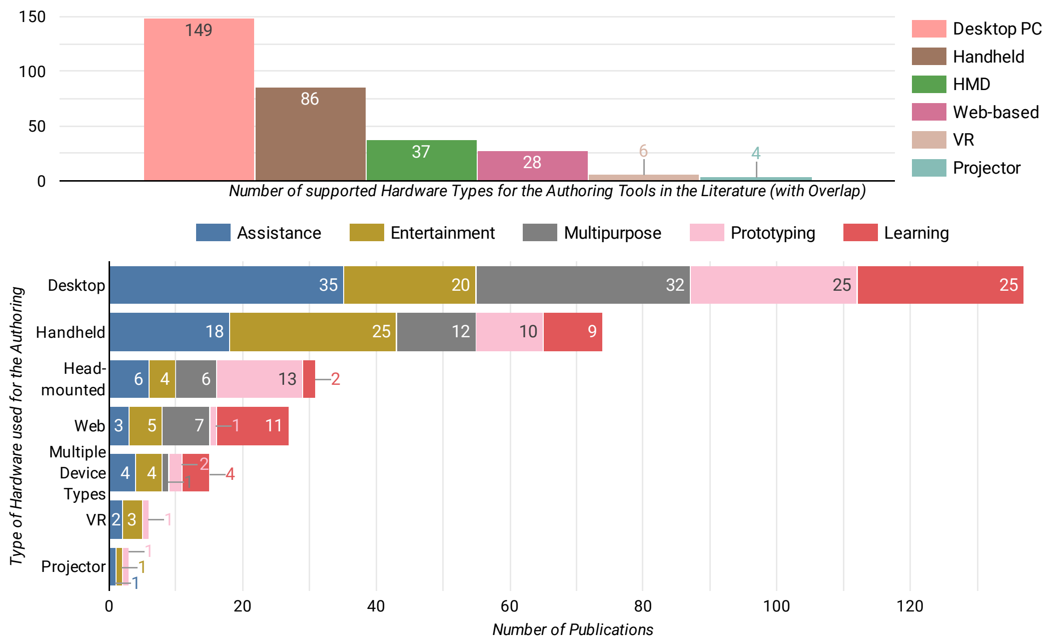


Figure 2.19: The number of proposed AR authoring tools in the 293 publications which support this hardware platform for the authoring of constructs, including overlaps (top) and the one-to-one mapped choice of hardware, including "Multiple Device Types" (bottom).

2.6.14 Markup Notation

While the majority of publications either did not explicitly state or used a markup language (62.46%), 30.72% of the publications did use a human-readable XML-based markup language. These publications include, for example, the usage of common markup languages like ARML, HTML, X3D, or VRML. But, out of the 90 publications using an XML-based markup language, 63 proposed and used a proprietary XML-based markup notations for their specific use-cases. Other markup notations formats were only sparsely used. 4.1% of publications used JSON to serialize the AR constructs, 2.39% proprietary markup notations, and 1 publication used a CSV file [67] (see Appendix Table 15).

As visualized in Figure 2.20 on the left, there are only slight tendencies for the usage of markup notations based on the deployment purpose of the AR construct (While significant, this tendency only shows a small association, $\chi^2(12) = 21.64, p = 0.042$, Cramér's V $= 0.16$). The lower percentage of publications using markup notations for entertainment or prototyping purposes is likely associated with the fact that content in these domains is often authored "impromptu", authored by the user of the construct themselves, and is lower in complexity. When visually inspecting the right plot of Figure 2.20, this trend also seems to be declining over the years and relative to

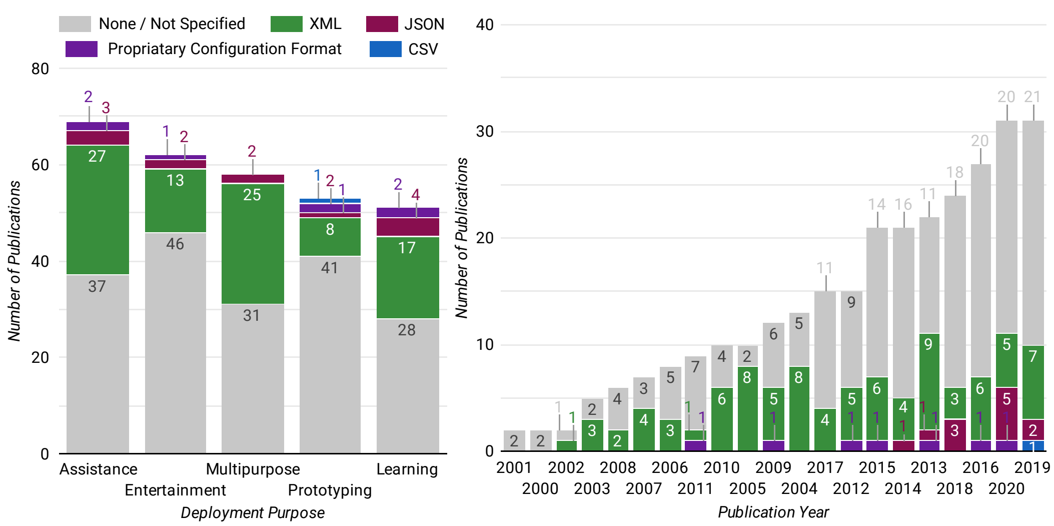


Figure 2.20: The 293 publications grouped bey the deployment purpose (left) and the publication year (right), broken down into the markup notation used to serialize the AR construct.

the overall new publications on AR authoring tools, fewer new tools utilize a markup language. While visually apparent, this association was not significant ($\chi^2(60) = 72.14, p = 0.135$).

2.6.15 Modularity

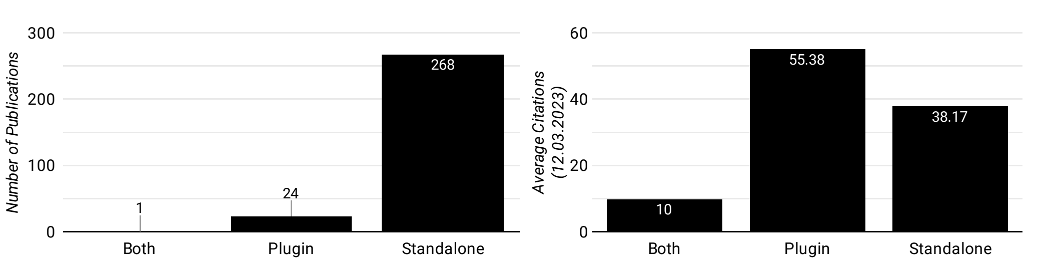


Figure 2.21: The 293 publications grouped by whether they are a standalone application intended to be used in a host software, or both (left) and the average citation count in March 2023 based on the implementation decision (right).

While Roberto et al. [400] made the modularity, therefore whether authoring tools are standalone applications or plugins used in host-software like Macromedia Director [470], Unity [436], or Microsoft PowerPoint [187], one of the four factors determining their classification of AR authoring tools, in the systematically reviewed 293 publications of this chapter's review, 91.47% of publications were described as standalone approaches (see Figure 2.21, left and Appendix Table 16). Interestingly, as can be seen on the right of the figure, plugin approaches have a higher average citation count in 2023, with 55.38 (SD = 90.94, Mdn = 10.5) citations, while standalone applications have 38.17 (SD = 70.2, Mdn = 19.5) citations on average. As visible through the medians already, this is caused by some highly influential publications on AR authoring tools, which proposed their tools as plugin solutions [187, 296, 326, 372, 436].

2.6.16 Scene Preview

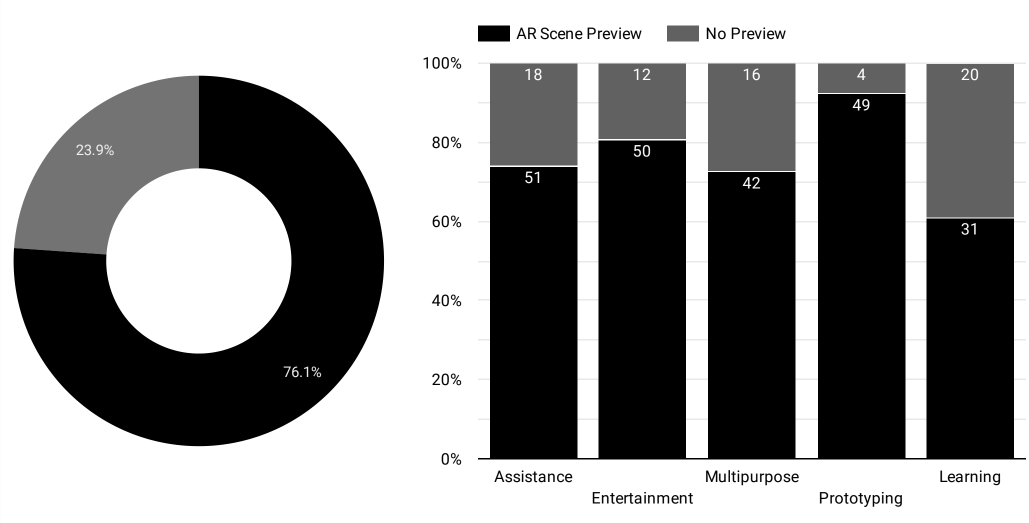


Figure 2.22: The 293 publications split into whether they provide a 3D (not necessarily in-situ AR) scene preview or not (left) and grouped by the deployment purpose of the authored AR construct, broken down by this differentiation.

As visualized in Figure 2.22 on the left and referenced for each publication in Appendix Table 17, 23.9% of authoring tools did not provide a 3D scene preview. Therefore, the AR construct was neither previewed in-situ as AR content, nor decontextualized, for the author, before it was distributed to the user. As can be seen on the right side of the figure, this was most often the case for AR constructs deployed for the purpose of learning (39.22% of tools). All these tools were web- or desktop-based AR authoring tools, that were mostly creating AR constructs to be used on mobile platforms, e.g., an AR authoring tool where the teacher can author a location-based AR game level through a web-based authoring tool, which students then use on their handheld devices [335]. As can be expected, for the purpose of prototyping, only 7.55% of AR authoring tools had to no 3D preview functionality.

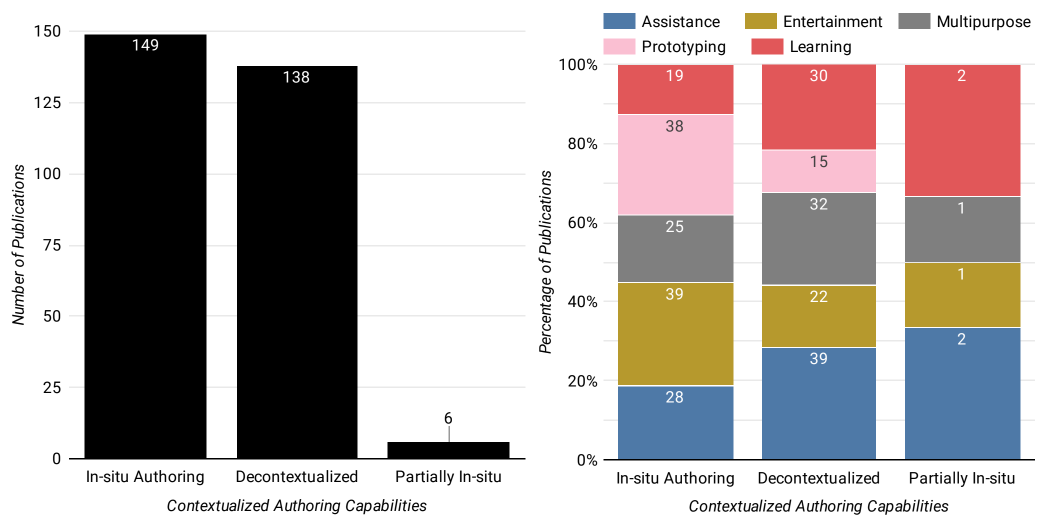


Figure 2.23: The 293 publications, grouped by the contextualization of the authoring process (left) and additionally as a stacked Barchart, broken down by the deployment purpose (right).

2.6.17 In-Situ Authoring

Figure 2.23 on the left side, groups the 293 publications based on whether AR content is authored in-situ, decontextualized or partially authored decontextualized and later positioned in-situ. This differentiation is referenced for all publications in Appendix Table 18. Generally, the design decisions are evenly split between in-situ authoring (50.85% of publications) and decontextualized authoring (47.1%), with only 2.05% of publications implementing a partial in-situ authoring method. As can be seen on the right side, AR constructs authored for the purpose of prototyping or entertainment are more often authored in-situ, while AR constructs authored for learning, assistance, general purposes are authored decontextualized. Though significant, the association between the dimensions appears to be small ($\chi^2(8) = 21.9, p = 0.005$, Cramér's V $= 0.19$).

While this is to be expected, whether content is authored "in-situ" or decontextualized is significantly and strongly associated with the availability of 3D preview functionality ($\chi^2(2) = 87.26$, $p < 0.001$, Cramér's $V = 0.55$), as content which is authored in-situ provides 3D preview functionality in most circumstances.

2.6.18 Authoring Interactions

Interestingly, the majority of proposed interaction concepts (55.63%) simply rely on traditional interaction metaphors of the chosen hardware type, e.g., traditional UI-interactions and on-screen touches on handheld devices, and the other 44.37% of tools utilize non-traditional interaction metaphors as the main interaction form to author the AR constructs. But, the diversity of these interaction concepts stems from their incorporation of numerous aspects, creating innovative combined interaction metaphors, making them challenging to abstractly map. Contrary to the envisioned AR author (see Figure 2.18), the problem is not that concepts are not described or distinguished clearly in the publications, but the opposite; there appear to be many novel combination approaches. And the choice of the interaction concept for the AR authoring tool is neither associated with the publication year ($\chi^2(540) = 432.81, p = 1$), nor the deployment purpose of the AR construct ($\chi^2(108) = 133.03, p = 0.051$). Therefore, while all non-traditional combinations are referenced in Appendix Table 19, it is only anecdotally described here, which interaction metaphors emerged beside the traditional interaction concepts. Notably, the appendix table maps the combination of approaches, which are not described here in detail but might be of interest.

Described broadly, the first interesting group of such non-traditional interaction concepts encompasses Markup Languages and Visual Scripting, which were used for 28 (9.56%) and 18 (6.14%) of reviewed AR authoring tools, respectively. These mechanisms afford a certain degree of abstraction, allowing the user to author complex AR constructs with programmatic or symbolic language. The markup languages hereby allow the author to manually adjust XML-based descriptions, and the visual scripting allows for the implementation of complex, potentially non-linear, logic without the need for actual programming.

Also used as the authoring metaphor in several proposed AR authoring tools were gestures & hand tracking, external controllers, and tangible marker & objects & interfaces which enable construct authors to manipulate the AR content in-situ. Furthermore, creativity-focused authoring interaction approaches like the utilization of drawing or sketching as interaction metaphors were identified. All these offer a more "impromptu" authoring of AR content, providing authors the freedom to create and manipulate constructs freely within the physical space. Moreover, voice interaction is used, providing an auditory method of control, and textual authoring offers the ability to create AR constructs purely through textual description. The concept of gaze as the interaction concept presents an immersive approach, with the interaction driven by the direction of the author's eye- or head movement. Uniquely, Robots and Drones have also been enlisted as interaction concepts, offering a remote, large-scale interaction capability. Lastly, video and picture annotation of scenes has been used as an authoring interaction concept to create AR constructs for the recorded scenes, e.g., locations or buildings.

2.6.19 App Relationship

As visualized in Figure 2.24 on the top left and is referenced in Appendix Table 20, 49.5% of publications are external authoring tools, therefore the application which creates the AR construct is not the same as the application which uses the AR construct. Besides these, 45.7% of tools are internal authoring capabilities of the application which also uses the AR construct, 3.8% have split authoring capabilities, e.g., where content is prepared in a web environment and then refined and contextualized in AR on the usage device [437], and 1.02% of publications offer both internal an external authoring solutions interchangeably. Visual inspection of the top right graph of Figure 2.24 indicates there to be a slight trend for internal tools to be used for the purpose of en-

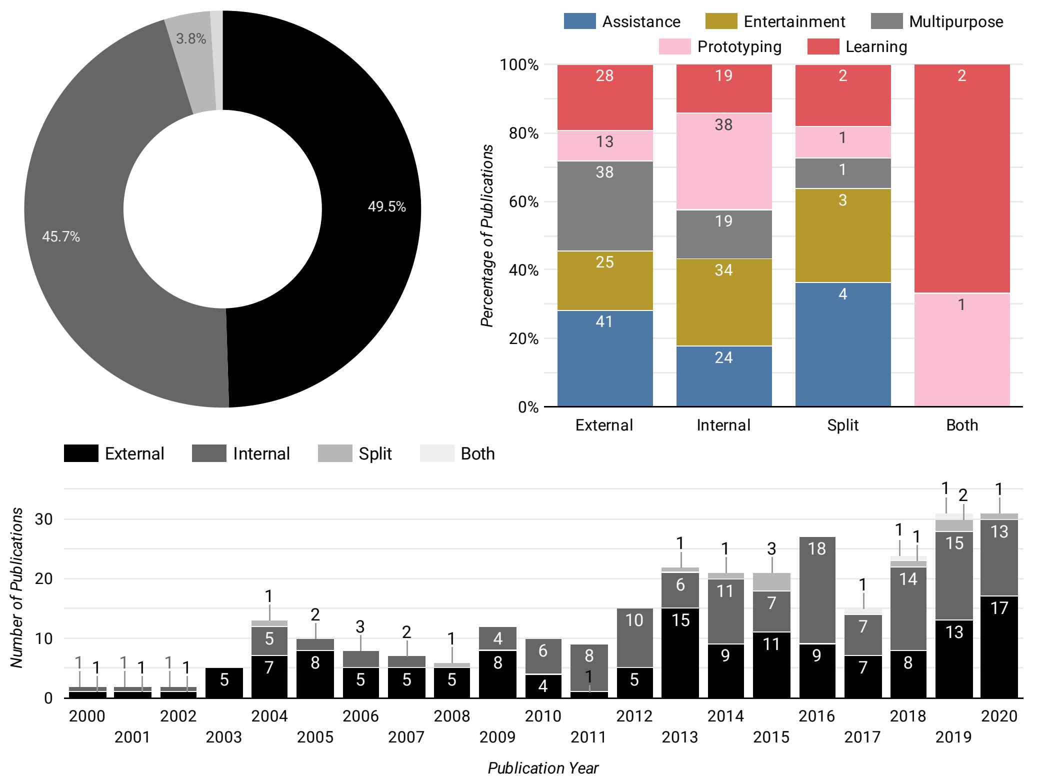


Figure 2.24: The 293 split into external and internal AR authoring tools (top left) and broken down into the deployment purpose of the created AR construct (top right). The bottom graph shows the publications grouped by year, broken down into the relationship between authoring and AR construct usage tool.

tertainment and prototyping, while external tools seem to slightly trend toward external tools. In line with the usage of markup languages, this slight trend is likely associated with the fact that content in these domains is often authored "impromptu" and lower in complexity. There also appears to be a general trend toward increased representation of internal AR authoring tools throughout the years (see Figure 2.24, bottom).

2.6.20 Construct Distribution

As visualized in Figure 2.25 on the top left and referenced in Appendix Table 21, 60.1% of reported AR authoring tools distribute or store the AR constructs locally. Only 39.99% of publications distribute their content through external servers. As visualized in the top right, the distribution decision is significantly associated with the hardware choice for the authoring component ($\chi^2(6) = 42.31, p < 0.001$, Cramér's V $= 0.38$). As can be seen in the bottom left graph of

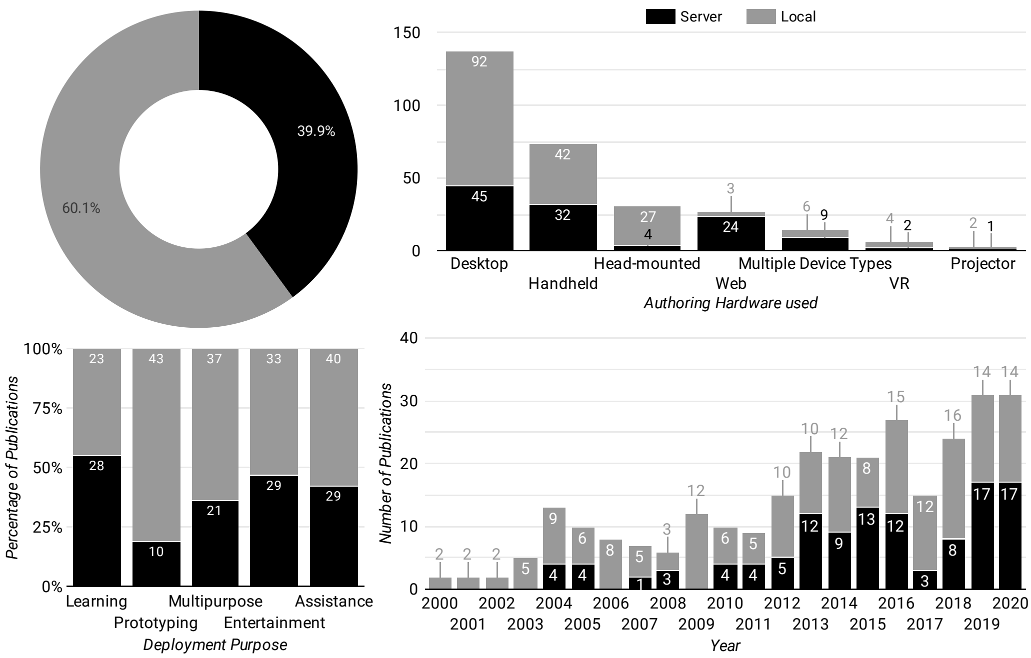


Figure 2.25: The 293 publication split into whether they distribute their AR constructs locally or through an external server (top left), this differentiation used as the breakdown dimension, when grouping the publications by the authoring tools' hardware choices (top right), deployment purpose (bottom left), and publication year (bottom right).

Figure 2.25, the majority (54.9%) of AR constructs for the purpose of learning are distributed through external servers, while this is only the case for 18.87% of tools where the purpose of the AR construct is prototyping. Through visual inspection, there appears to be an increasing trend of distributing content through external servers over the years (see Figure 2.25, bottom right).

2.6.21 Construct User

In line with the inductive coding of the envisioned AR constructs author, the AR constructs user was also particularly challenging to not only map in the first place but also subsequently simplify in a way that it provides an overview, without losing interesting insights from the distraction between expressions. Furthermore, even when trying to reasonably infer the intended AR construct user from the text of the publication (and e.g., the deployment purpose and context), for 11.26% of publications it was not possible to map them, and they are therefore grouped as "Not specified" (see Figure 2.26 and Appendix Table 22). Overall, 40.96% of reported AR authoring tools explicitly or implicitly target end users or consumers with the AR construct that is authored, e.g., for assistance, prototyping or entertainment purposes. 15.36% of publications target students, primarily for the stated purpose of learning, and the remaining tools (< 10%) target a variety of different construct user groups like employees, maintenance & assembly workers, designers, technicians, or children. This differentiation is seemingly quite specifically described at times, and in other circumstances describes quite general groups, or self-evident relationships like students using AR constructs for the purpose of learning. Ultimately, in line with the AR constructs author, this challenging of identifying for whom the AR construct is actually authored, is likely the most important insight gained from trying to inductively map the literature based on this dimension. As will be discussed in Section 2.7, researchers contributing AR authoring tools, should make this human factor consideration particularly clear.

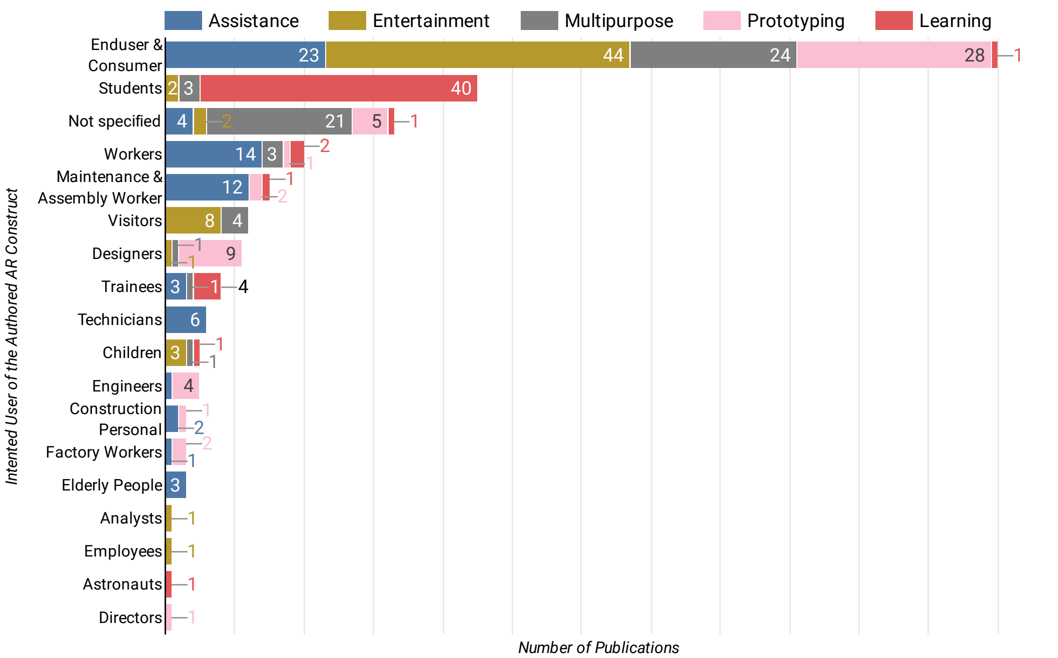


Figure 2.26: The 293 publications grouped by who is the envisioned user of the authored AR construct, broken down by the deployment purpose.

2.6.22 Usage Hardware

While Desktop PCs are the most common hardware choice for authoring, as can be seen in Figure 2.27 and is references for each publication in Appendix Table 23, the majority of hardware choices envisioned for the usage of authored AR constructs are handheld devices (55.63%) or HMD-based hardware choices (30.72%). This is followed by desktop PCs (18.43%), projectors (2.39%), web-based approaches (0.68%) and finally not specified or inferrable from the context for the remaining 2.05%.

The choice of the hardware for the usage of the authored AR constructs has a significant medium association with the choice of the hardware for the authoring of them ($\chi^2(36) =$

273.64, $p < 0.001$, Cramér's $V = 0.39$). This is largely explained by many of the proposed tools utilizing the same hardware (55.63%), or even the same application (45.7%, see Figure 2.24), for both, the authoring and the usage of the AR constructs.

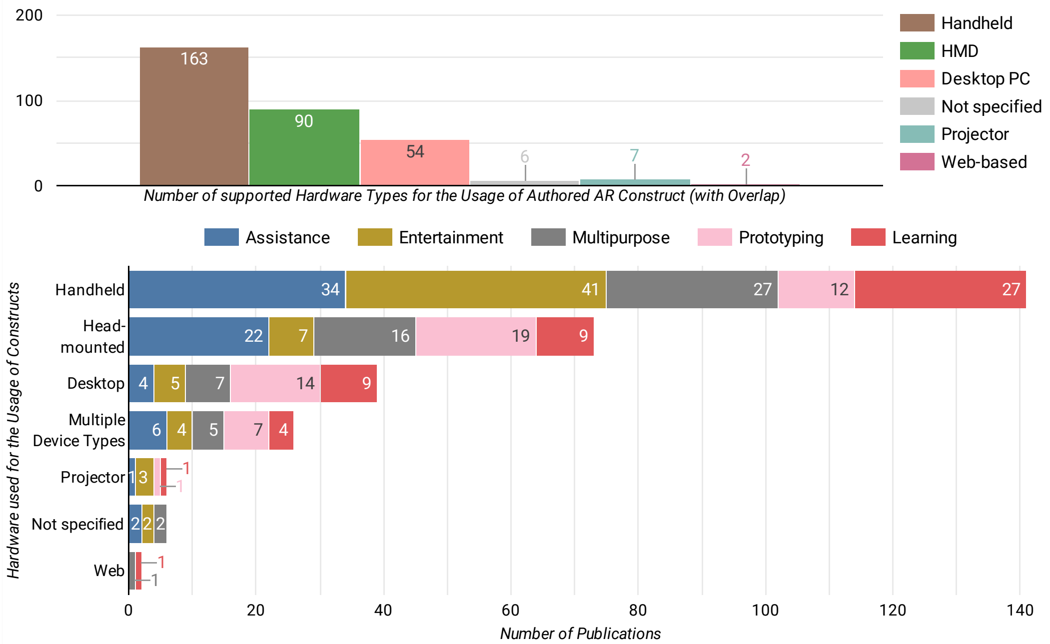


Figure 2.27: The number of supported hardware types in the 293 publications, which can be used to utilize the authored AR constructs, including overlaps (top) and the one-to-one mapped choice of the usage hardware, including "Multiple Device Types" as one expression (bottom).

2.6.23 User Interactions

While it was challenging to map the interaction concepts for the authoring stage, the expressions for the interaction concepts of the usage stage of the authored AR constructs were more clearly identifiable and could be mapped, as is visualized in Figure 2.28 and referenced in Appendix Table 24. While there were some proposals combining multiple interaction techniques (6.48%), the main interaction techniques for the usage of authored AR constructs proposed in the 293 publications were simply viewing the authored AR content (42.66%), traditional interaction techniques like button presses, generally UI element, and touch gestures on handheld devices (32.08%), tangible markers (10.58%), gesture/handtracking (3.07%), external controllers (2.73%), voice interaction (1.71%), and head/eye-gaze interaction (0.34%). Comparing the simplicity of interaction metaphors described for the usage of AR constructs with the complexity of techniques described for the authoring of them and combining this with the fact that the authoring perspective is also evaluated more often for its usability (see Figure 2.16), it appears that the authoring process is more often the focus of current efforts, compared to the construct that is created through the authoring process. While this appears logically consistent retrospectively, it is noteworthy that 42.66% of authoring tools envision no interaction beyond "viewing" the AR construct, which inherently must also have implications for the authoring process of them itself.

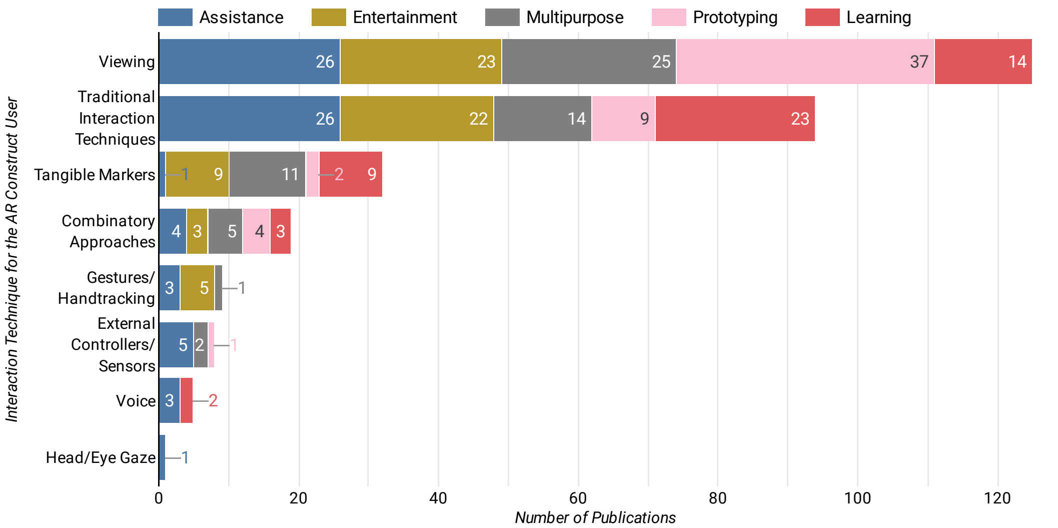


Figure 2.28: The 293 publications, grouped by the interaction technique used for the tool that uses the authored AR construct, broken down by the deployment purpose of the AR construct.

2.6.24 Content Type

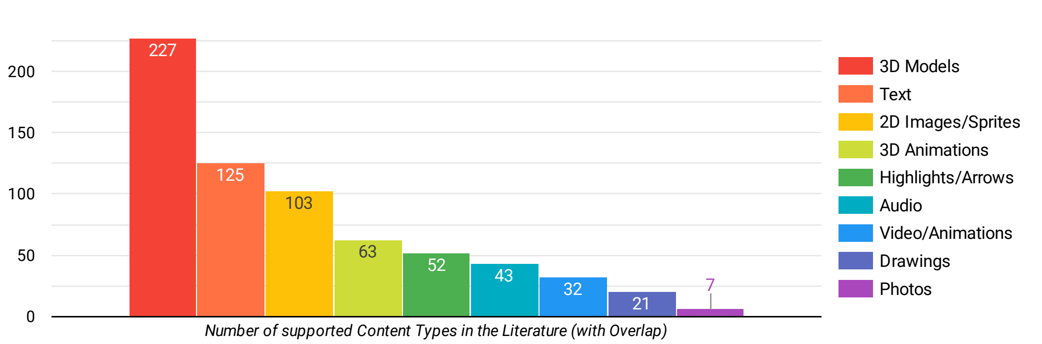


Figure 2.29: The amount of AR authoring tools reported in the literature that support the 9 identified content types. This graph includes overlap between content types, as AR authoring tools usually support multiple content types.

Visualized in Figure 2.29 and referenced in Appendix Table 25, are the 9 content types which were identified during the mapping, including considerable overlap, as tools often supported multiple content types. As is to be expected, the most common supported content type is static 3D models, with 77.47% of AR authoring tools supporting this type. This is followed by textual content (42.66%), 2D images/sprites (35.15%), 3D animations (21.5%), highlights and arrows (17.75%), audio (14.68%), video (10.92%), drawings (7.17%), and finally photos, which are only supported by 2.39% of reported authoring tools. There are some significant, medium associations between the support of the content types 3D models and 2D images/sprites ($\chi^2(1) = 24.21, p < 0.001$, Cramér's $V = 0.29$), 3D models and 3D animations ($\chi^2(1) = 23.33, p < 0.001$, Cramér's $V = 0.28$), text and 2D images/sprites ($\chi^2(1) = 27.14, p < 0.001$, Cramér's $V = 0.30$), and Video/Animations and auditory content ($\chi^2(1) = 29.74, p < 0.001$, Cramér's $V = 0.32$).

As other researchers already classified AR content type utilization, e.g., in the industrial AR context [156], or proposed taxonomic understandings of AR content [150], these endeavors are in line with the types of content identified in this review, and the classification of content of the AR constructs itself is not authoring-tool specific, the occurrence of specific content types is provided only descriptively here and not further classified or simplified.

2.6.25 Content Sequentiality

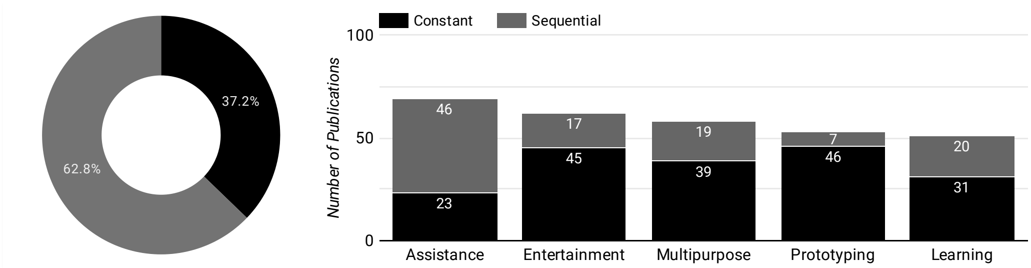


Figure 2.30: The 293 publications, split into whether AR constructs encompass a user-controlled sequence of states (sequential) or a single constant state/visualization (left) and the publications grouped by the deployment purpose broken down by this differentiation (right).

As visualized in Figure 2.30 on the left and is referenced for each publication in Appendix Table 26, the majority of publications (62.8%) report an AR authoring tool that authors sequential AR constructs; therefore AR constructs that encompass a user-controlled sequence of states, instead of just a constant visualization (37.2% of publications). Visual inspection of the right side of the figure indicates that AR constructions for assistance purposes have the largest percentage of sequential AR constructs (66.67%), while AR constructs for the purpose of prototyping have the lowest percentage (13.21%).

2.6.26 Tracking Type

Finally, in terms of tracking techniques utilized, as can be seen in Figure 2.31 and is references for each publication in Appendix Table 27, the majority of reported authored AR constructs (158, 53.92% of publications) used simple marker tracking. This was followed by markerless tracking (57 publications, 19.45%), e.g., through ARCore, ARKit, or the markerless tracking functionalities build into common HMD-based approaches like the HoloLens. Other tracking capabilities included object recognition and tracking, GPS-based tracking, no tracking, external sensors for localization, QR-codes, RFID chips, internal sensors and, as visualized in Figure 2.31, several combination approaches of them.

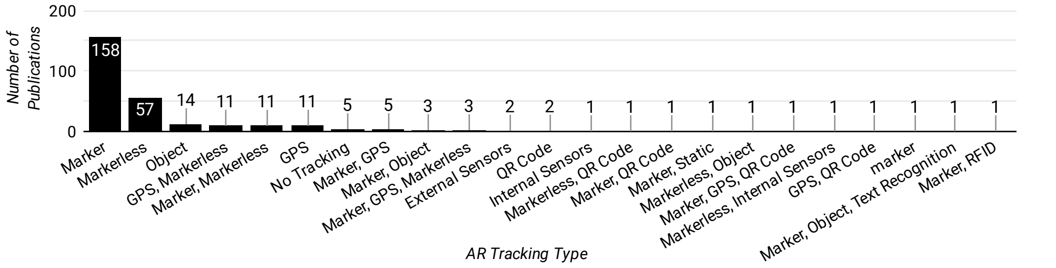


Figure 2.31: The 293 reviewed publications grouped by which tracking techniques were utilized for the usage of the authored AR construct.

2.7 Overall Scope & Gaps of the Field

Having now analyzed, mapped, and reported the 26 dimensions for the 293 publication, in adherence to the PRISMA ScR guidelines [461], the overall scope and gaps of the field can be discussed. Hereby, the main results are summarized, charted as they relate to the research questions, and a general interpretation is provided. Due to the complexity of the map itself, but also the field of AR authoring tools, it is likely not possible to holistically and concisely describe the whole scope with the field's gaps, and therefore only broad trends are scoped. Other researchers are encouraged to use the multivariate dataset [49] to explore more potential gaps overall or specific to their contexts, questions, or combination of dimensions of interest.

Scoping the bibliographic trends from the review, while seemingly not growing as fast as the overall field, there appears to be a growing body of research on AR authoring tools over the years, which is visible in the number of publications, the publications themselves referencing back to more publications, and the growth in citations. Additionally, efforts are increasingly published in academic journals and usually through the prominent HCI publishers (IEEE, ACM, and Springer), though there are a substantial number of publications which were published through a diverse set of publishers and subsequently journals and conferences. The efforts were made by various authors, though there are some prominent authors, who made significant contributions, as can be inquired from Section 2.6.2. Interestingly, in the review phase it became apparent, that

AR authoring tools were sometimes reported as side contributions in a publication, which makes them harder to identify, though ultimately, this was only the case for 13% of publications covered in this scoping review. The alluvial diagram in Figure 2.32 visualizes the flow of publications between the publishers, the publications format, and whether the AR authoring tools were a main or side contribution of the publications.

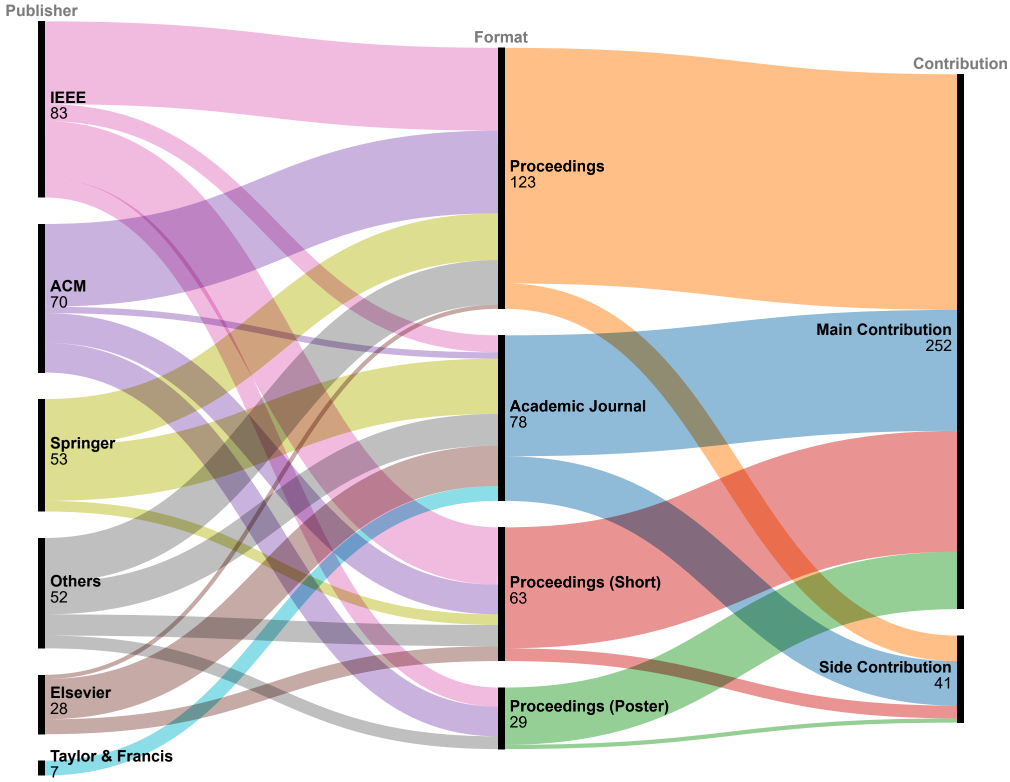


Figure 2.32: An alluvial diagram representing the relationships between three bibliographic dimensions of the publisher of each paper, the paper's format, and whether the AR authoring tools were the main or a side contribution of the publication.

In terms of the scope of reported AR authoring tools themselves, there are several interesting insights. Four distinct deployment purposes of authored AR constructs could be identified: Constructs for learning, for assistance, for prototyping or for entertainment. These were used across a diverse set of deployment contexts, which were highly associated with their deployment purpose. But, as visualized in Figure 2.33 when inquiring who the authors of the AR constructs or the users of them are envisioned to be, this was often not explicitly stated. In terms of reported technical implementations, like the hardware choice for the author and user, the interaction concepts and other decisions like preview functionality, distribution aspects, and markup notation usage, a variety of combinations became apparent with little association between them (e.g., see

Figure 2.34). Besides some almost self-evident relationships, which were analyzed and reported in the literature map, it appears that the scope of overall efforts is still preliminary and largely exploratory. This reinforces the assumption, that an overview was likely missing to understand the current scope of AR authoring tools in the scientific literature. Additionally, it also appears that there are no "distinct types" of tools which could be classified, at best there are some preliminary tendencies, e.g., design decisions based on the deployment purpose. Furthermore, arguably tools could be grouped into tools which create "ad-hoc" content, and tools which create more complex AR constructs, which seems to inform some design decisions. Maybe tools could simply be classified as internal authoring tools or external authoring tools as overarching types, as suggested by previous work [400]. But in the end, analyzed associations with the literature map at hand, would not fully support this as an overarching "type of AR authoring tools". Furthermore, while Hampshire et al. [184] proposed a taxonomic understanding of AR authoring tools which is not in contradiction to the findings of this scoping review, it only considers the AR constructs complexity and the authoring tools user interface as dimensions to inform who could be a potential author of AR constructs and, as can be seen in this scoping review, there are several additional dimensions to consider. Overall, it appears that guidelines or at least some guidance is currently missing in the field of AR authoring tools to work towards a more structured understanding.

Gaps of the Field of AR Authoring Tools

While the diversity of the scope of the field of AR authoring tools in itself is not a gap, but likely rather the result of interdisciplinary perspectives, the necessity for complex technical & human-factor considerations, and the relative novelty of the field itself making efforts exploratory at this point, there are certainly some clear gaps.

The three most challenging dimensions to inductively map and then meaningfully summarize were the interaction concept for the AR authoring tool, the envisioned author of the AR construct, and the envisioned user of it. While the dimension of the interaction concept of the AR authoring tool was challenging to map because of the variety of combination approaches of the expressions, for the AR constructs author and user it was a challenge because of many publications not specifically stating who the intended authors and users of their AR authoring tool are supposed to be. Therefore, one gap of current efforts is to more explicitly involve the human factors into the currently mostly technical considerations and explicitly report on these decisions.

Another gap in the literature stems from the intended interaction concepts for the AR constructs usage. Many publications simply envision "viewing" the authored AR content as the intended interaction concept for the usage of the AR construct. While this is likely stemming from a focus on the technical feasibility evaluation of the authoring perspective, this is the lowest complexity of potential interactivity of the construct to be authored. It is to be expected that the complexity of the authored AR construct substantially influences design decisions, like the interaction concept, for the AR authoring tool. For this, the prevalence of "viewing" as the intended interaction concept of the usage of authored AR constructs is arguably a gap in current efforts.

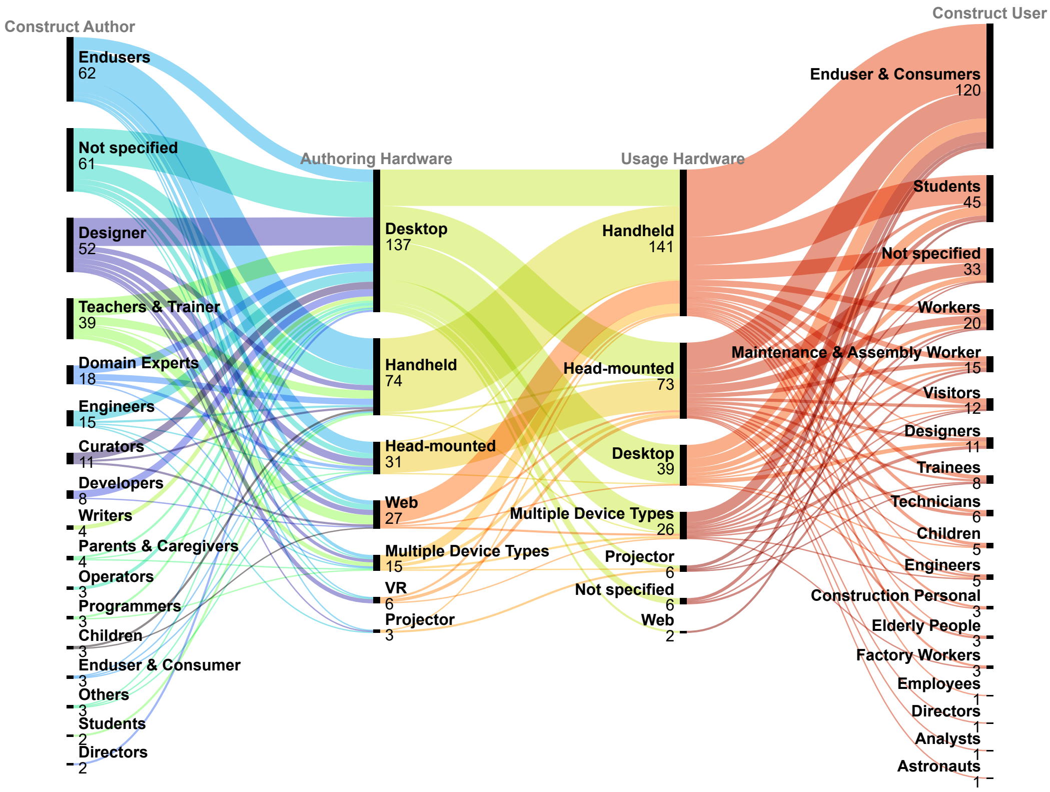


Figure 2.33: An alluvial diagram representing the relationships between four dimensions of the literature map: Construct Author, Authoring Hardware, Usage Hardware, and Construct User. The diagram highlights how the dimensions are interconnected in the mapped 293 reported AR authoring tools, with the width of the flow lines indicating the extent of each relationship.

Analyzing the usability evaluation efforts of AR authoring tools, there also appear to be research gaps in this aspect. Though evaluation efforts generally increased throughout recent years, with half of publications in 2020 at least evaluating one aspect of the authoring tool, arguably, there should be more emphasis on evaluating authoring tools holistically. Therefore, it is important that not only the authoring tool is evaluated, but also the AR constructs that are created with it. E.g., as will be discussed in detail in Chapter 4, it is not self-evident, that scalable concepts for AR trainings, that can be authored through authoring tools, elicit the same benefits as AR trainings specifically and carefully designed for a specific task. It is probable that this is likewise the case for the other deployment purposes. With only 5.8% of all reported authoring tools having evaluated both perspectives, there certainly is a gap in current findings in this regard across deployment purposes and contexts. Furthermore, even if publications reviewed both perspectives, the scope of evaluations was still quite narrow. Not only are sample sizes small, but evaluation efforts are also mostly focused on one specific context. If researchers are interested in learning how their eval-

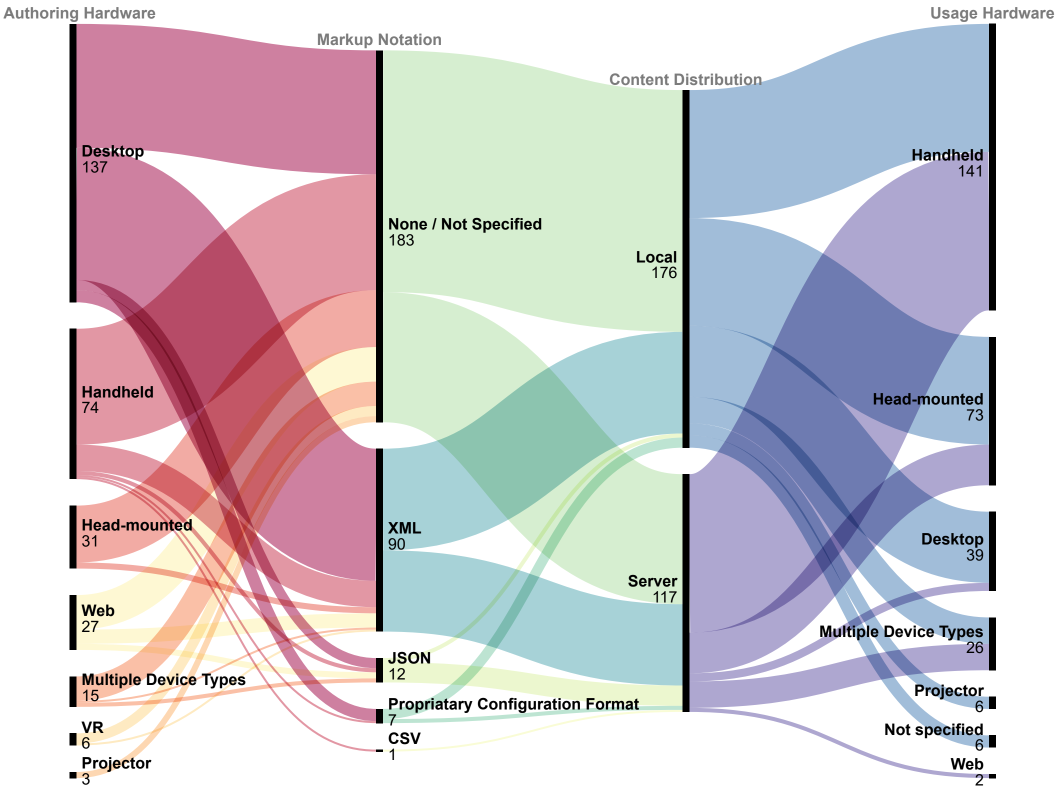


Figure 2.34: An alluvial diagram representing the relationships between four technical dimensions of the map: Authoring Hardware, Markup Notation, Content Distribution, and Usage Hardware.

uation challenges could be addressed, they should consult secondary literature for the systematic analysis of evaluation efforts in the field of AR in general [105, 319]. Ideally, the usability evaluation efforts would not only be conducted for both perspectives and in line with current evaluation guidelines from the secondary literature, but would also be done in a comparable manner. This could be achieved by using standardized usability questionnaires, like the System Usability Scale [65], where convenient analysis and benchmarking toolkits are already available [50].

AR authoring tools create AR constructs that have a predefined structure and are stored and transferred to be used on the same or another AR device. While there might be cases where the content of the constructs should not be openly accessible, for most cases, arguably a common description language should be used. In this, the sustainability of authored AR constructs can be ensured, as the content is not authored one-to-one for a single usage app or device but AR constructs and their usage considerations are separated. Somewhat in line with this gap, it also appears that AR authoring tools proposed as plugin solutions, in previous classification efforts even being their own category [400], are decreasingly addressed by the literature.

Finally, as Richard Stallman famously wrote, "sharing is good, and with digital technology, sharing is easy" [443]. With only 13 of the 293 AR authoring tools actually being available as open source, there clearly is a need for more open tools and concepts in the field of AR authoring. Currently, this might be the field's biggest gap. After all, authoring tools are a technology meant for scaling concepts and technologies and enabling more people to create their own content.

2.8 Limitations & Threats to the Mapping Study

All threats to the validity of the systematic review process, described in Section 2.4.2, are inherently also threats to the mapping study. For example, the potential of human error, or "judgmental errors" as Peterson et al. [374] described it specifically for mapping studies, is also present in the mapping methodology. Especially as 7618 expressions (293 publications that were mapped onto 26 primary dimensions) were manually mapped across the database by a single reviewer. This balance decision towards decreased "review effort" instead of increased review "quality" [327] was made because of monetary and time constraints and the goal to be as comprehensive as possible with the given set of resources. While Budgen et al. [69] discussed the fact that double-reviewer methodologies are often not realistic in the context of theses, Peterson et al. [374] points out that increased comprehensiveness could potentially even elevate the problem of judgmental errors because of increased inclusivity of publications.

Furthermore, the methodology of only textually describing selected publications to concisely depict the expressions of the dimensions, could lead to a distorted presentation based on textual narratives. In line with this, the selection of several visualization approaches with the aim to utilize fitting visualizations for specific dimensions and their expressions, could potentially lead to visual biases in the presentation.

Besides these threats to the validity of the mapping study, there are also limitations, which are specific to the mapping methodology. Firstly, the inductive mapping of dimensions required iterative simplification of expressions (see Section 2.5). While special care was given to not oversimplify dimensions to a point, where expressions are not properly represented, the simplification might have led to classifications of tools, the authors of the publications themselves, having a more nuanced perception of their work, would disagree with. This potential is at least partially amended by sharing the database as a multivariate categorical dataset under the CC-BY license, thus providing the authors and third-parties the possibility to suggest changes or expand upon the work.

Another limitation is the reliance on the honesty and correctness of publication author's reporting of the expressions during the mapping. That this can lead to problems, was already discussed in previous work, where e.g., Petersen et al. [374] discussed an extreme case of this challenge in a mapping study, where "73% of the papers were designated incorrectly, i.e., they for example promised an experiment which was no experiment". While such extreme percentages were neither apparent during the mapping nor are expected, the mapping did rely on the honest reporting of the capabilities of the tools by the authors.

2.9 Current & Future Work

This scoping review and mapping study offers several interesting possibilities for subsequent future work. Most importantly, the review timespan should be expanded upon. Furthermore, some dimensions are mapped abstractly for the purpose of this review but should be investigated in depth in sub-reviews, where only specific expressions of certain dimensions are taken from the database and reviewed for more specific insights, e.g., analyzing the evaluation efforts of AR authoring tools in detail, or building guidelines for Markup Languages in the AR context from reviewing the Markup Languages and Notations used in the context of AR authoring tools in the review. Finally, commercial tools could also be adopted into the mapping study or design space when either reached sufficient foundational certainty.

2.9.1 Expansion of the Review Timeframe

In future work, the timeframe of 2000-2020 of this review should be extended to include more recent publications. This is generally referred to as "second-generation" [505] reviews and can be achieved by adding further iterations of forward snowballing. Forward snowballing is shown to provide high precision in these cases, and is therefore the most efficient approach [135, 505]. In terms of effectiveness, current findings differ, with some indicating it has worse recall compared to additional database searches with restricted timeframe [135], and others indicating it has even better recall [505].

As a Pearson correlation shows that there is a highly statistically significant, very strong, positive, linear correlation between the publication years and number of published papers ($r(19) = 0.91$, p = < 0.001), it can be predicted how many publications would be roughly found in the new search. To predict the number of new publications that will be found, trendlines calculated through least squares regression can be used. When calculating the goodness-of-fit for linear, exponential or polynomial models, they yield R^2 values of 0.82, 0.83, 0.85 respectively. As all R^2 values are similar, the linear model is most appropriate for the short time steps, because of simplicity and yielding the most conservative prediction. Then, the trendline equation of

$$f(x) = \begin{cases} 1.361038961 * x - 2721.735931 & \text{for } x > 2020 \\ \text{undefined} & \text{otherwise} \end{cases}$$

can be used, where x is the year to predict and $f(x)$ is the number of new publications found. This results in a prediction of 28.92 publications for 2021, 30.28 publications for 2022 and 31.65 publications for 2023. In this, it would predict approximately 59 or 91 additional publications to be found, when the review is expanded for the timeframe until 2022 or 2023 (see Figure 2.35).

2.9.2 In-Depth Review of Evaluation Efforts

This mapping study does map papers based on whether they evaluated the usability of the authoring process by differentiating if the AR authoring tool, usage of authored AR constructs, both, or

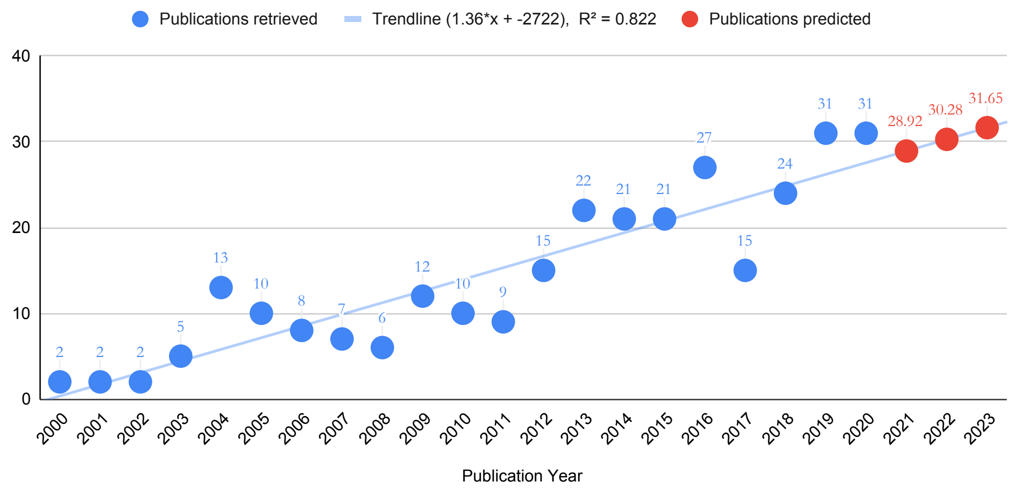


Figure 2.35: The number of publications retrieved from the timeframe of 2000 to 2020 in blue, a linear trendline, and an estimate of publications expected to be found for the upcoming years in red.

neither (see Section 2.6.10) are evaluated for usability. While this creates an overview of what and how much is evaluated and where creators of AR authoring tools could find inspiration on how they could evaluate their tools, this understanding is only a first step. The comparatively abstract approach was chosen deliberately, as the first review and mapping aimed to be rather wide than deep in its scope. Nonetheless, the systematically assembled and published database holds more valuable insights on evaluation efforts to be reviewed.

In future work, analyzing the sub-set of publications that did deploy evaluation efforts, would be of interest. As 59 publications evaluated the authoring aspect, 29 the usage of the authored constructs, and 17 evaluated both, the size of the sub-database would be 105 publications. This already merits its own in-depth review. Here, a more focused review could be performed to differentiate between hedonic or pragmatic usability qualities, to focus on the interpretation of usability principles based on ISO/IEC 9126-4 (Effectiveness, Efficiency, Satisfaction), or to even differentiate specific methodology or questionnaire usage. Some recent and influential reviews in the field address similar challenges in the broader context of AR evaluation efforts[105, 319], e.g., mapping them based on application area, evaluation method, hardware used, participant numbers, dependent measurements, gender balance, and visualization approaches. With AR authoring being a particularly challenging evaluation based on the requirement to ideally evaluate both: the usability of the authoring tool, but also the usability of the authored construct, including inherent interaction effects from the former to the latter, having a systematic “reference work” would likely help researchers developing and evaluating authoring tools.

Ultimately, and likely most importantly, the results and qualitative “lessons learned” from the evaluations in the publications could be synthesized besides the methodological expressions of the evaluation efforts. If schemes emerge, this understanding could lead to the development of guidelines on how AR authoring tools should be designed, what consistent challenges should be addressed, and how previous work was able to successfully or unsuccessfully address them. If differences or contradictory insights emerge, this could lead to explicit roadmaps on what future work should address to advance the field of AR authoring substantially.

2.9.3 Review & Synthesizing of Markup Language Usage

AR authoring tools always create an AR construct, which is then subsequently used by the same or another application to display it to a user. As stated in Section 2.7, it would be appreciated, if this transfer of the AR construct would be performed using a markup language, which only a third of tools in the mapped literature do (see Section 2.6.14). While, in specific cases, this might not be a feasible option, in most cases, not using a human-readable representation of content, is arguably simply a limitation. Moreover, ideally, the transfer should be done through a common markup language across all authoring tools. Why would 293 tools reinvent and use their own language, sometimes even new notations? This limits content to specific platforms, or even implementations, and slows down the development of the field because content cannot be reused or adopted by third-parties if, e.g., new hardware is released. A fact, researchers that proposed AR authoring tools, already discussed over 13 years ago: "In order for our approach to succeed in the future, one of the important aspects to address is the standardization of the content format." [86]

The answer to the question is apparent, when taking a look at the field of AR at large. It is not surprising that the two entangled concepts of AR authoring tools and "backend infrastructure for distribution of AR content and applications" are two of the five big challenges that are holding AR back, which Schmalstieg, Langlotz and Billinghurst discussed in their 2011 paper "Augmented Reality 2.0" [418]. The truth is, that ongoing efforts of developing common languages to describe AR scenes, scenarios or instructions in general, is a research field for more than 20 years. While many researchers have tried and continue to try to address it, the field is still struggling to keep up with ever-changing requirements, developments and technological advancements. In this, the efforts currently never have the chance to sufficiently conclude to provide enough value to be adopted at large, and continuously, the efforts gradually lose momentum and diminish over time.

Previous Work on AR Markup Languages

Besides many applications that simply propose their own AR markup language (e.g., as can be seen in Section 2.6.14) based on JSON or XML-based formats, there were several efforts over the years to develop AR markup languages in different contexts.

Ledermann et al. [263, 264] developed APRIL, the "Augmented Presentation and Interaction Language", an abstract XML-based description language that describes marker-connected scenes hardware-independently. Vitzhum et al. [478, 479] build upon APRIL to develop SSIML/AR, an APRIL compatible visual description language for AR scenes. When developing their ARGON AR web browser, MacIntyre et al. [297] also proposed an AR markup language combining XML and JSON elements to represent AR content based on geolocation and marker usage, which they later named "KHARMA" [195]. Dahne [99] explored how VRML could be extended to be used inside of AR authoring tools. Ruminski et al. [402] developed CARL, the "Contextual Augmented Reality Language", an XML-based markup language to enable the dynamic composition of AR scenes, which they developed towards "CARE", an ontology representation based on semantic web standards [405]. Jung et al. [221] proposed enhancements for the X3D markup language to better represent MR scenes. Müller et al. [334] proposed ARPML, an AR markup language based on the XML notations meeting requirements to be used for "procedural tasks like maintenance or repair". Lee et al. [268] proposed an XML-based AR markup language, named TARML, designed specifically towards the needs of interactive tangible user interfaces in AR. Walczak et al. [483, 484] proposed X-VRML, an XML-based language that incorporates dynamic modeling capabilities to virtual scene description standards such as VRML/X3D. Figueroa et al. [140] proposed a markup language to unify MR 3DUI descriptions. In line with these efforts, Broll et al. [64] proposed MRIML, the "Mixed Reality Interface Markup Language" which is intended to enhance languages like APRIL towards incorporating user interface descriptions. Hill et al. [195] propose KARML, an extension to the Keyhole Markup Language (KML) specifically for geospatial Augmented Reality content in AR browsers. Lechner et al. [260] proposed ARML, the Augmented Reality Markup language, an XML-based markup language for primarily geospatial Augmented Reality content, expanded upon the KML. In a later work, Lechner [259], based on requirement analysis, extended the work towards supporting more visual assets, anchors, and more closely incorporating already established descriptions from KML. Kim et al. [232] proposed an extension of these efforts called the "Mixed Reality Contents Data Markup Representation", that extends the augmentation to also include fixed and movable objects beside loosely geotagged information visualization. Kim et al. [230] also later proposed a specific scheme of descriptions included in metadata of AR content to increase re-usability of geospatially anchored AR scenes. Park et al. [363] proposed a 5W1H metadata scheme for context-specific AR videos. Seo et al. [423] proposed an HTML-based markup language to enable interoperability between MR technologies. Almeida et al. [7] proposed an AR markup language for the specification of assembly steps of "kit format" products, sent to consumers. O'Connor et al. [352] proposed an XML-based language that allows authors to connect interaction metaphors with AR elements. Gonzalez [169] proposed AUIML, the Three-dimensional User Interface Description Language for Augmented Reality in a descriptive, XML-based data format. Mourkoussis et al. [328] proposed ARCO (Augmented Representation of Cultural Objects), a markup language for the representation of Augmented Reality content specifically for museum exhibits. Dominguez et al. [112] proposed the Augmented Reality Service Description language. Ganlin et al. [147] proposed the "Unified modeling language of augmented assembly instructions", an XML-based markup language for assembly procedures. Haller et al. [183] proposed an XML-based AR markup language as part of the AMIRE (Authoring Mixed Reality) framework, that enables "authoring Mixed Reality once, [and] run it anywhere". Looser et al. [289] and Coelho et al. [91] proposed AR-specific extensions to OpenScenceGraph to enable the description of AR-specific information, which were called OSGART and OSGAR.

Furthermore, there were also already efforts to compare existing AR markup languages, like Visser [476], who compared KML, ARML, KARML and Junaio XML based on eight requirements based on potential use cases. In 2014, Ahn et al. [5] compared AREL, ARML 1.0, ARML 2.0, KARML, Layar, Architect, and Webized AR based on their base document structure, registration model, target model, supported content modalities, and rendering compatibility. But these efforts were neither conclusive, nor have additional comparative efforts been made since. Finally, in 2015, ARML 2.0 was even recognized as a standard by the Open Geospatial Consortium [353]. Still, there is virtually no adoption of it or any of the other languages.

AR Authoring as a Realistic Scope for AR Markup Languages

Obviously, it is challenging to develop standards for an emerging research field like AR, where the requirements constantly change and developments are substantially faster than a standardization could realistically take place, simply outpacing the usefulness of a standardized description format.

In the context of AR authoring, a standardized description format is arguably not only especially useful, but there is also chances in the dynamics between AR authoring tools and AR markup languages for the field at large. While the field of AR at large might be an unrealistic scope to cover through a markup language, as applications can differ in many dimensions and sometimes expressions are not clearly formalizable, AR authoring tools inherently have to come with static, thought-out structures, expressions, and capabilities. In easier words: AR authoring tool developers have to already know what the AR construct structure will be like, to develop the tool to create them. With this, there is a fixed set of capabilities and possible expression. Because of this, using the database of AR authoring tools to systematically specify a first set of requirements for AR markup languages could significantly advance the overall field of AR, not only the specific AR authoring subfield. Matching these requirements of AR authoring tools with already existing AR markup languages could furthermore show which features are most important, which markup languages are promising to expand upon, and where more research is required. In this, AR authoring tools could serve as a realistic starting-set of requirements based on a more well-defined subfield with its real requirements, compared to specific cases or developments based on expectations as is currently the case. This set of initial requirements can then be expanded upon to be more inclusive to expressions currently not covered by AR authoring tools, but potential limitations of overfitting a markup language because of biases or context-specific viewpoints from the start, are prevented.

2.9.4 Mapping Commercial AR Authoring Tools

The choice to only scope and subsequently map tools described in scientific literature in the review is in contrast to overlapping reviews, but was fully deliberate. To start, including commercial tools would violate criteria for high-quality systematic reviews, as included tools would not be peer-reviewed (see Section 2.1). Secondly, in line with the first point, the information quality would likely vary significantly. Even in the more heterogeneous context of scientific literature describing authoring tools, some dimensions (e.g., Construct Author, Construct User, Interaction

Techniques utilized) were already challenging to inductively map. Moreover, the available scope of scientific literature was more than sufficient to justify a separate review. Ultimately, it would have been challenging to "scope" non-scientific approaches, as the lack of quality criteria during the inclusion phase of this type of review would likely result in skewed perceptions based on "promises" of publicly available marketing materials for tools, that have no obligation for honest and factual reporting of capabilities compared to the academic context.

Nonetheless, with a sufficiently mature understanding of the expression of the dimensions of the literature map (therefore the levels of the mapped categories), reviewing commercially available AR authoring tools with this already established framework, could add important insights and should be addressed in future work. If comprehensive mapping of commercially available tools is something of value, has to be explored, but there certainly are commercially available AR authoring tools worth mentioning, that are distinct in several dimensions, which coincide with findings from the literature map. While neither comprehensive in scope nor representative of the depth of expression for the map, there are for example:

- **Microsoft Dynamics 365 Guide:** A combination of Desktop and HoloLens-based authoring tool for procedural (step-by-step) instructions on HMDs for assistance purposes, that was released in early 2019.
- **Unity MARS:** Short for mixed and augmented reality studio, an extension for the Unity Engine released in 2020, which allows AR developers to more efficiently and conveniently create static AR scenes, that can be used on Smartphones and HMDs for any purpose.
- **Wikitude Studio:** A desktop-based authoring tool released in 2012, intended to aid designers in the creation of 3D marker-based AR applications for entertainment purposes on Smartphones.
- **Vuforia Studio:** Another desktop-based AR authoring tool, that was released in 2016 that, compared to Wikitude Studio, is more targeted toward industrial use cases and procedural AR action chains for assistance purposes.
- **Apple Reality Composer:** An AR authoring tool released in 2019, available for both desktop PCs but also Smartphones, that enables end users to create static 3D AR scenes for entertainment purposes.
- **Ikea Place:** An AR application for Smartphones, which allows end users to place Ikea furniture into rooms to prototype static room setups in AR, which was released in 2017.
- **Google Maps AR:** A semi-automatic web-based authoring tool, intended to be used by business owners, for location-based AR content for informational purposes, like providing store descriptions to end users. It utilizes already established web interfaces for Google business entries, to automatically display the information in AR.

2.10 Summary

In this chapter, 15692 publications were screened and reviewed through a full-fledged hybrid search strategy to identify 293 scientific publications reporting AR authoring tools. After assessing the review methodology and results, these AR authoring tools were subsequently mapped onto 26 primary dimensions, to create the map of AR authoring tools with its 7618 expression. The map was then visualized to create an understanding and overview of the state of the field. Subsequently, the scope and gaps of the field of AR authoring tools were discussed. Then, limitations of the mapping study were stated. Finally, potentially valuable directions for future work were emphasized: The timeframe should be expanded, evaluation efforts of AR authoring tools should be reviewed in-depth, and commercial tools should be included into the map.

3 Constructing the Design Space for AR Authoring Tools

"The art of progress is to preserve order amid change and to preserve change amid order." — Alfred North Whitehead

In the previous chapter, 293 publications were systematically reviewed, scoped and mapped onto 26 dimensions. While this already provides value, as it structures previous efforts, visualizes them through a high-level mapping study, and reveals the scope, current trends, but also gaps of the field of AR authoring tools, it still does not directly inform researchers and developers interested in creating AR authoring tools beyond suggestions to follow in general: Properly addressing and reporting the "human" factors in Human-Computer Interaction, making tools actually available, evaluating tools, and using common content notations. But, to accomplish this, this entirely theoretical perspective of mapping which design decisions previous researchers have made when developing AR authoring tools, can be utilized and expanded on toward the more practical perspective of creating a framework, or at least a starting point for other researchers to explore their design options for AR authoring tools.

3.1 A Taxonomic, Typologic or Categorized Understanding

The logical subsequent course of action after mapping the literature would be to create a taxonomic, typologic or categorized understanding of the AR authoring tools which were reported. And, as was visualized in the literature map in Section 2.6 and Section 2.7, there are some interconnections between some expressions of the dimensions which were mapped. But these are primarily interconnections between human factors dimensions and hardware/software design decisions. Generally, the 26 dimensions can also be described grouped as 7 bibliographic dimensions (*Authors, Citations, References, Publication Year, Publisher, and Publication Format, Contribution*), 13 hardware/software design decisions made by the creators of the AR authoring tool (*Authoring Hardware, Authoring Interaction Concept, Authoring Contextualization, Authoring Tool Modularity, Internal/External Authoring, Authoring Preview, Content Type, Content Sequentiality, Markup Notation, Distribution, Construct User Hardware, User Interaction Concept, Tracking Type*), and 6 dimensions regarding the human factors (*Availability, Usability Evaluation, Deployment Context, Deployment Purpose, Construct Author, Construct User*).

But for a "taxonomy of AR authoring tools", therefore a taxonomic understanding, the software/hardware design dimensions would have to be classifiable in a hierarchical structure; therefore dimensions would have to be highly associated. For typologic understanding, a hierarchical classification system would not be necessary, but there would still have to be specific traits or associations between multiple mapped dimensions to classify them into a specific type. But is this the case in the literature map?

Through the Python library Dython by Zychlinski [537], heatmaps that display asymmetrical associations between categorical variables can be created to visually explore these potential relationships. Based on calculating the entropy coefficient, Theil's U (denoted in the following as $U(X,Y)$[1]), asymmetrical associations, ranging from 0 to 1, are calculated and visualized for two categorical variables (X and Y) each and then all associations of a given multivariate dataset are plotted. In line with Cramér's V, the results can then be interpreted similarly to correlations, but indicate asymmetrical association ($U(X,Y) \neq U(Y,X)$, while $V(X,Y) = V(Y,X)$) instead of correlation [538]. In simple terms, asymmetrical association means that the entropy coefficient quantifies of how much of knowing the state of Y, reduces the uncertainty about the state of X. In this, 0 indicates no association and 1 indicates perfect association between categorical variables and, similar to a correlation, provides a first indication of "possible relationships" of the variables.

Using this method to create a heatmap for all 13 hardware/software design decisions reported in the literature, the results are visualized in Figure 3.1. Even when utilizing one-hot encoding for the most complex and overlapping dimensions of the AR content type (see Appendix Table 25), there is very little association between the 13 design decisions. While there are some associations, that can be described as "medium" associations, these are decisions like the fact that using the internal or external AR authoring tools are often associated with specific AR hardware ($U(Internal/External\ Authoring, Authoring\ Hardware) = 0.47$), which simply means that internal tools are often associated with mobile hardware and external tools often utilize desktop PCs or web-based approaches. Or the association that the authoring contextualization; if content is authored in-situ or decontextualized, is associated with whether the authoring tool is integrated into the usage application or an external tool ($U(Authoring\ Contextualization, Internal/External\ Authoring) = 0.42$). Without fully exploring all associations beyond these exemplary cases, as can be inquired through Figure 3.1, there is almost no association between the dimensions and the dimensions where, at best, medium associations exist are arguably almost self-evident.

There are three potential explanations for this result. The first explanation is that one or more dimensions are actually distinct categories of tools, and in this, associations would exist, when analyzing them for each category separately. One logical dimensions where this could be the case would be the Deployment Purpose. But arguable, this should have then been apparent in the literature map. Furthermore, exploring this through separate entropy coefficients is challenging, as selective filtration based on specific dimensions reduces the overall dataset size and restricts the available information pool, making it more likely to find associations. The second potential explanation, though unlikely in the HCI context [390], is that the explorations are still preliminary, and perfect associations between design decisions do exist, but were not established yet. The third,

[1] Theil's U was introduced and defined in his 1972 book "Statistical Decomposition Analysis" [459]. For a full but compact definition of Theil's U, see: https://en.wikipedia.org/wiki/Uncertainty_coefficient.

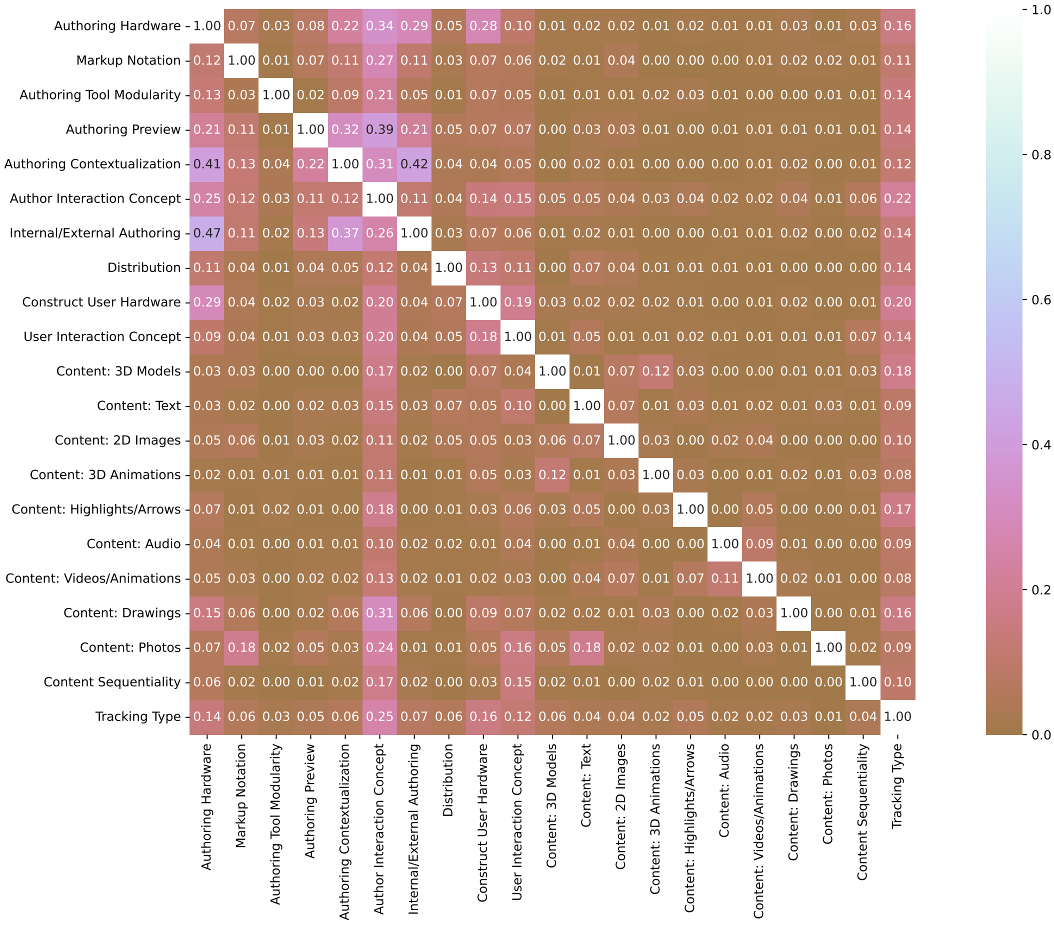


Figure 3.1: Asymmetrical associations based on Theil's U [459] ($U(X, Y) \neq U(Y, X)$) between all 13 software and hardware design decisions mapped from the 293 AR authoring tools reported in the literature between 2000 and 2020.

and arguably the most likely, explanation is that the dimensions are, in fact, not associated and are distinct decisions to be made during the development of an AR authoring tool.

As can be seen, there are no substantial associations in the design decisions, and it is not possible to create a taxonomic or typologic understanding of the developed characteristics. And while previous research has tried to "illustrate aspects of [an AR] authoring taxonomy" [184] this approach only considered three dimensions: The authors' expertise, the constructs complexity and the authoring interface complexity, but this is no longer reasonably applicable when considering more dimensions. In line with this, subsequent classification efforts [400], where tools were classified into distinct categories and then assigned to one of four AR authoring tool types (described as

"Models" in the paper), also further prompt the question: Is this high-level categorization actually sufficiently informing the practical development of new AR authoring tools?

3.2 Constructing a Design Space to Explore

Subsequently, having now established that neither a taxonomic, nor a typologic understanding could be created based on the overall literature map and arguably a classification method would not be appropriate, an alternative approach is necessary. Notably, while not possible for the overall literature map, these structured understandings already exist in the literature for multi of the mapped dimensions in itself, e.g., on AR content [156, 453], or AR interaction metaphors [194, 332]. Therefore, ideally, the approach chosen to systematically elaborate AR authoring tools would be an approach which helps to "bridge from relatively theoretical concerns to the practicalities of design" [299]. One common approach, to accomplish exactly this, is the construction of design spaces in the Design Space Exploration (DSE) methodology. DSE, in early efforts also referred to as Design Space Analysis [299], is an argumentation-based approach to designing constructs and "[producing] an output which can help others understand why the resulting design is the way it is" [299]. Or as Lane [248] described it, "a design space identifies the key functional and structural choices made in creating a system design, and it classifies the alternatives available for each choice. Rules can be formulated to relate choices within a design space".

Therefore, it can be argued that all 293 publications found in the literature did explore the design space of AR authoring tools already, as they made design decisions for all the mapped dimensions. The question is, how informed these design decisions were, how many of the alternative approaches available were systematically considered, and how well others can logically follow the researchers' reasoning. And this is no critique of their work, as this early exploration is necessary, even before knowing all expressions of the dimensions or even which dimensions are of importance. The problem is, that "without fully exploring the design space, the designer cannot be sure whether there exists another approach, which would achieve the goal without any commonality with known approaches. In these situations of sparse requirements, analysts may misrepresent design decisions as requirements, creating an illusion of requirements in software development" [390]. And therefore, as "cumulating design rationale promises to be a powerful way to compile contextualized design knowledge" [299], a very first proposed design space of AR authoring tools is constructed, or technically revealed, based on the design decisions made by previous efforts here.

Importantly, the design space does not inform what "a correct decision" would be for a specific design challenge, but rather serves as a guide for the researchers or developers, trying to develop their AR authoring tool, to follow alongside making their own decisions. However, this also does not imply, that there are no ideal design decisions about specific design aspects. It simply guides which concrete decisions are to be made (dimensions) and reveals which decisions previous efforts made (potential expressions). Comparable efforts of constructing design spaces were already proposed in the AR context, e.g., by Büttner et al. [72] to understand the application of AR for assistive environments in manufacturing, by Wiegand et al. [502] to present a design space for in-car AR applications, or by Ez-Zaouia et al. [521] to propose a design space for AR authoring tools in educational contexts, that comprised of four dimensions.

3.3 The Design Space of AR Authoring Tools

The design space of AR authoring tools [2], informed by the literature map of Chapter 2, therefore has 13 design decisions to be made that can be split into considerations regarding the AR authoring tool itself, the authored AR construct, and the tool that uses the authored AR construct:

- *The AR Authoring Tool*
    1. **Authoring Hardware:** Which hardware (e.g., handheld devices, HMDs or Desktop PCs) is utilized during the authoring process of the AR construct
    2. **Authoring Interaction Concept:** What primary interaction concept is used to author the AR content (e.g., traditional interaction techniques of the hardware choice, gaze interactions, visual programming)
    3. **Authoring Preview:** Whether a 3D preview functionality exists, that shows the author what the final AR construct would look like
    4. **Authoring Contextualization:** Whether the content is authored in-situ, directly in the context or decontextualized (e.g., authored on a desktop PC with or without a 3D preview, but later used as AR content contextualized on a machine)
    5. **Authoring Tool Modularity:** Whether the tool is a standalone application (e.g., it's an executable) or a plug-in used in a host software (e.g., in Blender or Unity)
    6. **Internal/External Authoring:** Whether the authoring tool is its own tool or integrated functionality in the tool that is also utilized to use the AR construct

---

- *The Authored AR Construct*
    7. **Content Type:** Which AR content types are used in the AR construct (e.g., 3D models, Animations, Sprites, Textual content)
    8. **Content Sequentiality:** Whether the AR construct is a constant (possibly interactive) visualization, or a user-controlled sequence of states
    9. **Markup Notation:** Whether, and in which form, content is serialized in a human and computer readable format
    10. **Distribution:** Whether content is distributed locally or through web-based services

[2] This design space is pertaining to the **manual** authoring aspects of **persistent** AR constructs. In this, e.g., live collaboration aspects or automatic content generation is not included in the design space. A full definition is provided through the Inclusion criteria of the literature review in Section 2.3.4

- *The AR Construct Usage Tool*
    11. **Construct User Hardware:** Which hardware (e.g., handheld devices, HMDs or Desktop PCs) is utilized to use the authored AR construct
    12. **User Interaction Concept:** What primary interaction concept is utilized to use the AR content (e.g., viewing, traditional interaction techniques, gesture tracking)
    13. **Tracking Type:** How the AR content is technically tracked (e.g., through AR-Markers, markerless feature point detection, or external tracking sensors)

3.3.1 A Systematic Understanding: Neither Comprehensive Nor Definitive

This design space is neither meant to be comprehensive nor definitive. There are likely more decisions to be made beside the ones which emerged from reviewing the literature of the timespan of 2000 to 2020, and there are definitely more expressions to be explored within each dimension. It is important to note that even if the current expressions within all dimensions were fully mapped, new ones will emerge over time due to the evolving nature of the field of AR. In this, the design space will change with time and dimensions and expressions will have to be revised. Constructing, or rather revealing, the design space here primarily serves as a starting point to create a structured understanding of a comparatively new field, where efforts are disseminated across different publishers and venues and are not as prominently represented in the prime venues as other topics within the AR research efforts [233, 526]. This is to be expected because of the interdisciplinary perspectives involved, but makes finding the efforts and especially identifying trends, learnings, and gaps in the literature challenging for researchers.

In this, the design space is built upon the literature mapping study and only described here on an abstract level. As no recommendations are formulated in the design space itself, part of the exploration of the design space is to actively consult the literature map, e.g., through the multivariate dataset [49], to explore the design space and make informed decision which expression of each dimension in which combination would be appropriate for a specific purpose, context, author, and user. Notably, expressions that were not yet explored together provide no indication that they should not be explored together. One entirely legitimate usage of the design space would be to explicitly search for expressions across dimensions, which were not explored in combination before, to identify research gaps or explore specific combinations empirically.

3.3.2 How to Explore the Design Space: A Guiding Framework

As of now, the design space covers the 13 software/hardware design decisions, but the mapping study also mapped human/context factors, as previously indicated. These dimensions are not part of the design space but are rather the orthogonal perspective, informing the exploration of the design space. Therefore, the 4 human factors (Deployment Purpose, Deployment Context, AR Construct Author, AR Construct User) inform the 13 decisions to be made. Additionally, while not explicitly mapped in the mapping study because of its granularity, the concept of "what" is actually augmented by the AR construct should also be considered in the exploration. In this, a guiding methodology is proposed, which is inspired by the 5W-1H methodology often used in journalism for unbiased and comprehensive reporting, where the five W-questions "who", "what", "when", "where", "why", and "how" are supposed to always be addressed. As creators of AR authoring tools should also objectively reflect on the human factor questions before making their design decisions, while not explored in the literature before, this methodology seems to also be applicable in the DSE methodology. As the design space of AR authoring tools specifically is not concerned with the temporal component of the creation of the AR constructs, the W-question "when" is furthermore replaced with a second "who" question, to explicitly cover the importance of considering the author and the user of the AR constructs separately. Also, the "how" question is the exploration of the design space itself. Therefore, the 5W-1H inspired questions which should be reflected on as a guiding framework to inform the design decisions to be made, are as visualized in Figure 3.2. The order in which the aspects should be reflected on is not rigid and likely dependent on external factors which cannot be predicted from the theoretical perspective.

1. Where: In which context are authoring tool and AR constructs used?
2. What is "augmented" by the AR constructs?
3. Why: For what purpose are the AR constructs authored?
4. Who is the user of the created AR constructs?
5. Who is the author of the AR constructs?

How should the AR authoring tool be designed?

Figure 3.2: The 5W-1H inspired five W-questions, which creators of AR authoring tools should reflect on before exploring the Design Space (H) of AR authoring tools.

3.4 Visual Summary

Visually summarized in Figure 3.3 is the combination of the first proposed Design Space of AR authoring tools (on the right) with the proposed 5W-1H-inspired guiding questions (on the left), with first non-representative reflections on which guiding question likely influences which dimension of the Design Space. Besides serving as a visual representation of the design space, this hopefully helps researchers to not only reflect on the design decisions systematically but also helps to report the design space explorations before contributing AR authoring tools to the scientific literature more clearly, so others can retrace and recreate deliberately made decisions.

Where: In which context are Authoring Tool and AR Constructs used?	What is "Augmented" by the AR Constructs?	Who is the user of the created AR constructs?	Why: For what purpose are the AR constructs authored?	Who is the author of the AR constructs?	How: The Design Space of AR Authoring Tools		
					Dimension	**Expressions**	**Reference Map**
✓	✓		✓	✓	**Authoring Hardware**	Desktop PC	App. Table 14
						Handheld	
						HMD	
						Web-based	
						VR	
						Projector	
✓	✓			✓	**Authoring Interaction Concept**	*See Section 2.6.18 for details*	
✓	✓		✓		**Authoring Contextualization**	In-Situ	App. Table 18
						Decontextualized	
						Partially In-Situ	
✓	✓		✓	✓	**Authoring Tool Modularity**	Plugin	App. Table 16
						Standalone	
						Both	
✓	✓	✓	✓	✓	**Internal/External Authoring**	External	App. Table 20
						Internal	
						Split	
						Both	
✓	✓		✓		**Authoring Preview**	3D Preview	App. Table 17
						No 3D Preview	
✓	✓	✓	✓		**Content Type**	3D Models	App. Table 26
						Text	
						3D Images/Sprites	
						3D Animations	
						Highlights/Arrows	
						Audio	
						Video/Animations	
						Drawings	
						Photos	
✓	✓		✓		**Content Sequentiality**	Sequentiality	App. Table 26
						Constant	
✓			✓		**Markup Notation**	None/Proprietary	App. Table 15
						XML	
						JSON	
						CSV	
✓	✓	✓	✓	✓	**Distribution**	Local	App. Table 21
						Server/Web	
✓	✓	✓	✓		**Construct User Hardware**	Handheld	App. Table 23
						HMD	
						Desktop PC	
						Projector	
						Web-based	
✓	✓	✓	✓		**User Interaction Concept**	*See Section 2.6.23 for details*	
✓	✓		✓		**Tracking Type**	Marker	App. Table 27
						Markerless	
						Object	
						GPS	
						No Tracking	
						QR Code	
						External Sensors	
						RFID	

Figure 3.3: The design space of AR authoring tools with its 5W-1H-inspired guidance to explore it.

4 Heb@AR—Augmented Reality Trainings for Midwifery Education

"In theory, there is no difference between theory and practice. In practice, there is." — Richard P. Feynman

Having now established the Design Space of Augmented Reality Authoring Tools, we are keen to explore this design space and, informed by the expressions of the space's dimensions, develop an AR authoring tool as one contribution of this thesis. While we will do this in the following chapters for procedural task training on handheld devices, we have to establish a solid foundation to prove the usefulness of **scalable** handheld AR trainings for procedural learning first. Additionally, through this, we can facilitate a comprehensive understanding of the conventional AR development process and context-specific challenges and opportunities. Ultimately, this context knowledge will serve as the variables with which we will explore the AR Authoring Design Space afterward. In this, the exploration will be done with real life experiences, and not, expectations. To accomplish this, it is necessary to first diverge from the authoring topic. We will briefly redirect the focus towards AR for procedural task training. In this specific case, the procedural task training on handheld AR devices in the context of academic midwifery education, where we developed the Heb@AR App (see Figure 4.1) during Project Heb@AR. Afterward, we will thoroughly discuss relevant learnings, insights, and implications before returning to the main thread of the thesis, where we will explore the Design Space and develop an AR authoring tool for procedural task training based on the combination of the theoretical Design Space and experiences from practice.

Figure 4.1: A student disposes of a glass ampule after completing the virtual preparation of an emergency tocolysis in the Heb@AR App during curricular implementation at Bielefeld University of Applied Sciences and Arts [381]. Picture by Patrick Pollmeier, licensed under CC-BY 4.0

Structure of this Chapter

Inside this chapter, we will first propose our vision of Augmented Reality-based trainings as one of the logical successors of web-based trainings in Section 4.1. Then we will discuss the academization of midwifery education and Project Heb@AR in Section 4.2. Afterward, the development process of the AR trainings during Project Heb@AR is described in Section 4.3, before the complete Heb@AR App is shown in detail in Section 4.4. Section 4.5 shortly describes the development of usability evaluation utility, before Section 4.6 reports our current evaluation findings of the Heb@AR App in terms of utility and usability. Finally, Section 4.7 discusses the app, its development, and the current findings of its usefulness, before Section 4.8 summarizes and concludes this chapter.

4.1 Toward Augmented Reality-Based Trainings

In their book "The Teaching Gap", Stigler and Hiebert wrote in 1999: "School learning will not improve markedly unless we give teachers the opportunity and support they need to advance their craft by increasing the effectiveness of the methods they use" [444]. Since then, digitization of learning provided new opportunities for teaching, e.g., by introducing asynchronous learning approaches based on e-learning techniques. This not only allows teachers to work more efficiently but also provides benefits to the learner, such as spatially independent communication, self-regulated learning, as well as access to learning anytime and anyplace [74].

4.1.1 Computer- and Web-Based Trainings

Endeavors towards Computer-Based Trainings (CBT) started in the 70s to increase the learners' independence in space and time. In the 90s, the emerging internet yielded advanced Web-Based Trainings (WBT), providing new ways to intensify teacher-learner as well as learner-learner communication and to provide new forms of feedback. WBT approaches in general have been shown to be well accepted by students, reporting self-efficacy as the main factor, independent of the perceived usefulness [369]. Beyond the motivational benefits, though, WBTs struggle to be more effective than traditional approaches, as a meta analysis showed [246]. This might be attributed to the challenge of creating high-quality and sustainable e-learning content [295], which is underlined by the finding that different desired learning outcomes require different kinds of instructions [146].

For procedural training tasks, which primarily consist of a combination of cognitive strategies and motor skills rather than basic declarative knowledge [145], conventional CBTs or WBTs might not be sufficient. The reason becomes apparent when contextualizing the coverage of CBTs and WBTs in Bloom's Taxonomy [242]. It is hard to argue that CBTs can support the learner beyond the levels of *remembering* and *understanding*. While WBTs, with the embracing of social media and communication components, also address the fifth level, *evaluation*, they neither sufficiently support the learners in *applying* (3rd level) or *analyzing* (4th level) procedural task knowledge nor do they provide the freedom of exploration that would be necessary to reach the highest level of Bloom's Taxonomy [242]: *creating*. Though WBTs can arguably be applied to the evaluating level of procedural task learning, it has to be noted that this mostly consists of quite time-consuming, often hand-crafted, methods.

4.1.2 Augmented Reality-Based Trainings

What is needed in terms of technical features to fully address the third and fourth level? One technology potentially able to fill this gap is Augmented Reality. Endeavors towards Augmented Reality-based Training (ARBT) combine the benefits of WBTs with AR's biggest strength of contextualizing information in the physical world (see Figure 4.2). This makes ARBTs interesting for practical training and procedural tasks [54]. Current findings indicate that applying AR as an additional "multimedia source" into existing curricula can already lead to improved retention, attention, and satisfaction [413]. Furthermore, a meta analysis conducted by Ozdemir et al. [359] indicates increased academic achievement compared to traditional learning methods, increased concentration and the enabling of teachers to convey concepts faster and with more clarity through demonstration of connections between concepts and principles. Generally, systematic literature reviews also point towards a consistently positive impact of AR tools used in educational settings [438], especially through interaction, catching the learners' attention and increasing motivation [387]. Many more notable benefits of utilizing AR in educational settings like, e.g., concretization of abstract concepts, flexibility, triggering of creativity, or presenting save learning environments, were synthesized by Yilmaz et al.[516]. Interestingly, while significant differences can be observed for all levels of education, the largest effect size of learning benefits is observed for students of undergraduate level [359].

Figure 4.2: AR as the logical extension of Computer-based Trainings & Web-based Trainings.

4.1.3 Acceptance & Scalability of ARBTs

Despite those apparent didactic benefits, several challenges for a realistic, scalable deployment of ARBTs into training procedures remain. For one, AR-headsets are still expensive and have a half-life period of under 2 years, which renders it almost impossible to deploy larger set-ups at University level. They thus do not scale up to group sizes of today's university-level training of practical skills or vocational training. Moreover, the technology has limitations, such as a narrow field of view, experimental gesture-based interaction methods and unstable tracking under non-optimal conditions. In combination with a lack of media competence in teachers and students with this technology, this can lead to acceptance problems.

As success factors for AR deployment, user experience, stability, adaptability, and independent self learning capabilities have been identified [83, 100]. Technology acceptance models (TAM) applied to potential AR trainings in educational contexts show that students perceive the technology as useful, easy to use [490] and teachers' attitudes imply their intention to utilize AR [208]. Nonetheless, those studies measure perceived use and not actual usage [464]. While behavioral intention can influence morale, disposition and performance, perceived usefulness and perceived ease of use are not reliable indicators for practical acceptance and subsequent usage [464].

Guiding educational theories tailored towards AR training are still ongoing research [83]. However, generalized concepts for AR training that teachers can directly apply into their curriculum are a primary demand from their perspective [466].

Those limitations are, in our view, the reasons AR is still not extensively used in education. While those limitations also apply to handheld AR applications, the users' familiarity with smartphones as well as recent advancements in hardware and tracking solutions (e.g., ARCore [170] for Android & ARKit [14] for iOS smartphones) make them feasible candidates as platforms for AR training applications that can be realistically implemented in educational curricula, e.g., even considering bring-your-own-device (BYOD) approaches. As a consequence, in line with the success factors identified by Dalim et al. [100] and Cheng et al. [83], for us scalability requires:

1. **ubiquitous availability** of devices,
2. **place, and time independence** for self-regulated learning,
3. **high usability and low entry threshold** to compensate low levels of media competency,
4. **clear concepts for interaction and didactic** to maximize the support for teachers in defining new learning materials

Requirement number one is technically met by consumer smartphones of recent years, as has been detailed above. AR-based training is independent of the availability of the teacher, and thus in principle time independent. Whether it is place independent depends on the required context objects: expensive special purpose devices might only be available in laboratories or special training facilities and thus restrict spatial flexibility. Other trainings might not require any additional material, which provides maximum spatial flexibility, as AR applications can then provide alternative virtual proxy environments and context objects in cases the physical ones are inaccessible.

For the last two requirements, comprehensive interaction concepts, including feedback mechanisms and didactic contextualization, are still largely missing. This is especially true for more complex training scenarios, such as procedural training tasks. Systematic literature reviews reveal that most handheld-based AR training scenarios are rather static, only displaying non-procedural information bits with very little to no interaction [57, 362]. They also often only focus on small learning scopes, mostly covering only an isolated topic without long-term focus or feasible scopes of deployment beyond what was necessary for evaluation [387]. While exemplary, scenario-specific AR training applications already elicit the mentioned didactic benefits, there is a need for generalized concepts that work beyond isolated topics for targeted evaluation studies, in particular addressing the challenges of scalability and long-term deployments. Additionally, it is not self-evident that the didactic benefits reported in the literature persist, when moving from prototypical technical implementations with ad-hoc evaluation studies, to the realistically scalable concepts necessary to accomplish this vision of ARBTs today.

4.2 Academization of Midwifery Education & Project Heb@AR

While these challenges and thus our vision of ARBTs are generally true across disciplines, one appropriate starting point to explore the concept is the academic midwifery education. The German midwifery education is currently transitioning towards a full academization. Based on a recent law, since 2022, midwives are exclusively qualified at universities, rather than by vocational training through the German "dual education system", which previously was the standard educational path in Germany [165]. While this is an important step towards increasing the status of midwives in the medical context, it also leads to new challenges. The practical component of the training still has a high priority, with exemplary bachelor degrees consisting of around 4380 hours of theoretical and 2200 hours of practical training. As these practical parts are at least partially covered on-site at the universities, this naturally leads to bottlenecks regarding available practical tutors, training space and scheduling restrictions for trainees in the long run. Furthermore, with the full academization of the midwifery education, the heterogeneity of students magnifies, only increasing the impact of those restrictions due to potential needs for more individual support. Despite those new challenges, the general goals remain: Students have to reach a theoretical and practical expertise and especially the transfer from theory to practice has to succeed. Moreover, students also have to be prepared for important emergency situations that cannot be trained reliably in practice. In this, the academic midwifery education is a discipline that has high amounts of procedural training requirements and, because of the recent academization, is especially receptive to exploring new teaching concepts and models, like the ARBTs. In general, the academic medical education is already moving towards problem-based learning approaches to enable students more self-regulated learning, often in collaborative settings [535]. This is partially done in laboratory training sessions, so called SkillsLabs, and with the utilization of learn-management systems (LMS). These procedural training tasks could benefit from Mixed Reality problem-based learning approaches, but this research area was largely unexplored before our project efforts.

4.2.1 Project Heb@AR

In the research project Heb@AR, funded from November 2019 to December 2022 by the Federal Ministry of Education and Research (BMBF) in Germany, the utilization of handheld AR as a supplementary tool for the practical training components in the academic midwifery education was explored. At the time of starting the project, the utilizing of AR to support the educational goals in midwifery education was not covered by the literature. Specifically of interest during the project was how and where AR can be used effectively in this context, how acceptability and accessibility for tutors and trainees can be ensured and how well emergency situations can be simulated using the technology. Handheld AR was specifically chosen as a technology, in contrast to, for example, Head-Mounted AR devices or Virtual Reality approaches, because of its potential to be scalable immediately because of cost-efficient devices for institutions and BYOD approaches, in line with our vision of ARBTs stated before.

To accomplish this goal, in November 2019, forces were joined with interdisciplinary researchers from Midwifery Science at Hochschule für Gesundheit Bochum [1], Medical Didactics at Ruhr University Bochum [2], and Human-Computer-Interaction (HCI) at University of Applied Sciences Emden/Leer [3]. In short, the project consisted of the development of several exemplary AR training scenarios in the context of midwifery emergency management for handheld AR devices that were iteratively evaluated, improved based on the Design-Based Research methodology [394], and then implemented into the curriculum of the Bachelor of Science midwifery study program at the Hochschule für Gesundheit Bochum. Moreover, the implementation of these AR scenarios into existing structures, e.g., learning-management systems from the technical perspective, but also the onboarding of lectures to use the technology in their teaching, was explored. For this, not only the trainings themselves were developed by the team, but entire supporting structures surrounding the AR trainings were created for both, the lecturers, but also the students to successfully accept and utilize them. Finally, the possibility to create methods to enable lecturers themselves to create new AR trainings was investigated.

4.2.2 The Project's Research Questions

The overarching research question the Heb@AR project team addressed was: *"Is the use of AR to promote the competence to act of prospective midwives meaningful and productive?"*.

Furthermore, within the project, several interesting research questions arose for each of the project partners individually. Some questions, from the perspective of the midwifery researchers and medical didactics researchers, are detailed in our publications on project Heb@AR [54, 481]. The ones with a technical and HCI focus, primarily targeted:

1. **Acceptability:** How can handheld AR be utilized effectively and implemented in a way that it is intuitively usable and perceived as useful by the students?
2. **Scalability:** Which scenarios are suitable to be trained through handheld AR? Do AR scenarios scale better with the increasing number of students compared to on-site trainings?
3. **Viability & Longevity:** How can lecturers be enabled to create their own scenarios using authoring tools, which was identified as a crucial factor for successful longer-term AR im-

[1] Professor Nicola H. Bauer, Professor Annette Bernloehr, Kristina Vogel, Tabea Willmeroth
[2] Dr. Matthias Joswig, Carmen Lewa, Professor Thorsten Schäfer
[3] Jonas Blattgerste, Professor Thies Pfeiffer

plementations in previous work [58]? How can handheld AR be integrated seamlessly into existing technical (e.g. LMS) but also social teaching and training structures and contexts?

4. **Self-regulation & Collaboration:** In which ways can handheld AR support self-regulated and collaborative learning at home and in training facilities?

While we will not directly address the project's research questions inside this thesis, they are stated here as the developed Heb@AR App is built with careful considerations for its scalability, collaboration factors, and acceptability. It is also evaluated in Section 4.6.3 regarding its usability and perceived usefulness (acceptability) but also objective usefulness. Furthermore, informed by these research questions, factors of longevity and viability will be addressed in later chapters, where we propose the authoring tool TrainAR. Therefore, the questions are implicitly addressed through the reported design decisions.

4.3 Development of the AR Trainings

Overall, we developed five AR trainings during Project Heb@AR, which followed a Design-Based Research methodology in accordance with [394]. Therefore, we tried to work with prototypes which we iteratively evaluated with experts, lecturers, and students, to gather impressions, feedback, insights, and especially their diverse perspective, to subsequently improve the prototype towards the final versions used in the Heb@AR App, that is desribed in the following Section 4.4.

Because of the nature of the medical context, with its complexity and nuances to take into consideration during development, the methodology that was used for bootstrapping the initial prototypes differed slightly and was more linear than we initially anticipated. As one learning during the early stages of development, we had to develop our own methodology to ensure that the action sequences were correct from the midwifery perspective (e.g., in line with treatment/standard-of-care guidelines), addressed all didactic considerations, were still technical possible to develop, and retained good overall usability of the AR training itself. Therefore, a more linear approach for this initial bootstrapping was created. As we believe that this methodology, which proved to work for us during Project Heb@AR, might be valuable for other researchers or developers trying to transfer procedures into practical Mixed Reality trainings, it is shortly described here. Furthermore, learnings from developing this methodology are later used to explore the design space of AR authoring tools and for the technical and didactic development of the TrainAR authoring tool.

Describing this methodology abstractly, the training procedure that was supposed to be implemented was first observed in detail and recorded on video. Based on the observed and recorded action sequence, a task process analysis [216] was created, formalizing the action sequence. Based on the task process analysis, the learning goals and didactic considerations were discussed and formalized. Utilizing this, the process analysis was then translated into a stateflow diagram, specifying the technical carrying out more closely and taking into account the technical feasibility and potential limitations. Then, the training was technically implemented as a first prototype before they were iteratively improved. This transfer methodology is visualized in Figure 4.3, described in more detail in [54], and described in the following with a focus on the HCI perspective.

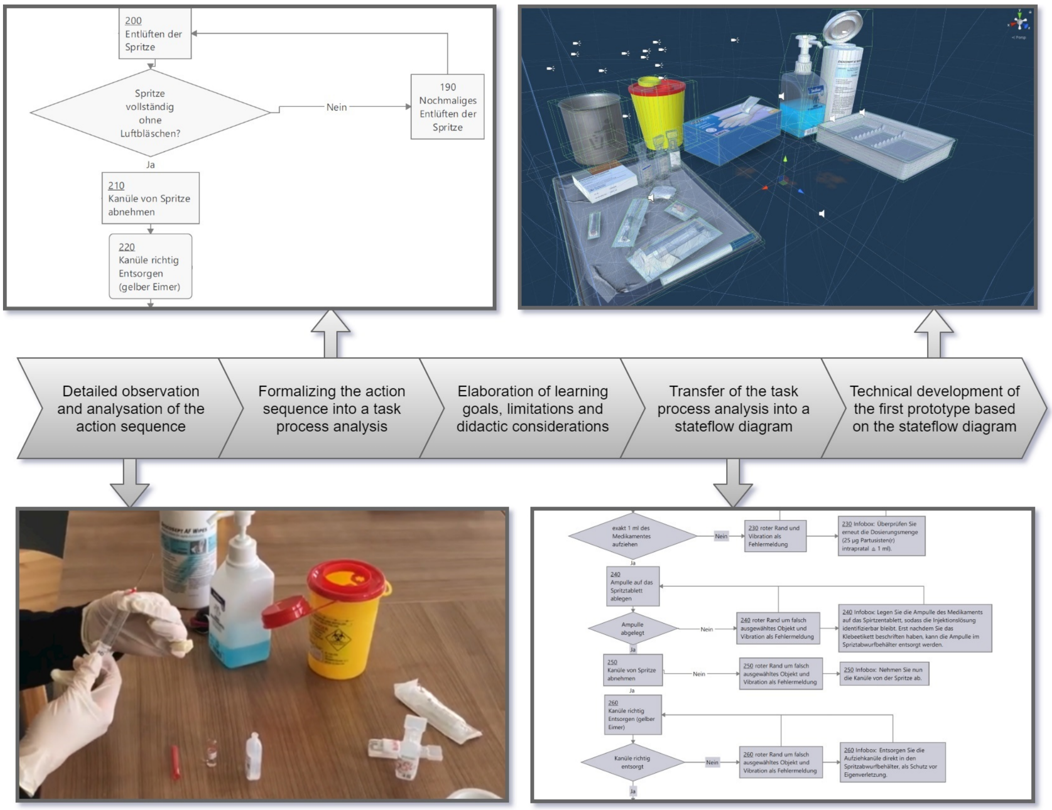


Figure 4.3: The initial "bootstrapping" development methodology established during Project Heb@AR for the structured development of medical AR trainings in interdisciplinary teams.

4.3.1 Observation & Documentation of the Procedure

After analyzing and elaborating a training task to be implemented as an AR training based on guidelines and findings of evidence-based medicine (EBM), the complete procedure of the task was first demonstrated to the team by one of the midwifery researchers. After questions and uncertainties were resolved, the procedure was then documented on video. Depending on the training, we either used smartphone cameras or, when the perspective of the trainee was important, head-mounted cameras to capture the procedural training. During the execution, the midwifery researcher verbalized individual action steps and important aspects. This recording was then shared with the medical didactics and HCI researchers.

4.3.2 Formalization of the Task Process Analysis

The medical didactics researchers then formalized the task process analysis [216] based on the video recordings in consultation with the midwifery researchers, using a software called MindManager. This task process analysis is a description of the sequence of actions and does not take into account learning goals or considerations on how the procedure could be represented in an AR training.

4.3.3 Establishment of Learning Goals

Subsequently, the work process analysis was used to derive, among other things, cognitive and psychomotor learning objectives based on taxonomy levels and clinical competence levels by the medical didactic researchers. They carried out the assignment of the learning and competence levels, that are suitable for teaching practice, to AR learning activities, with reference to the MARE framework [529]. Based on this, the general requirements for AR trainings were formulated and scenario- and location-specific AR implementation recommendations were elaborated.

After finalizing the didactic design of the AR training, based on the work process analysis, the competency-based learning objectives were coordinated within the team based on what would be didactically valuable, technically possible to implement, and midwifery-professionally accurate.

4.3.4 Procedure Transfer into Stateflow Diagrams

After the learning goals of the AR training were established, they were combined with the task process analysis and a stateflow diagram was created by the team, using the charting software Miro. In our case, the midwifery researchers developed the initial versions of the stateflow diagram from their perspective, which were then iterated on by the team. In contrast to the task process analysis, which describes the actual procedure in reality, this stateflow diagram describes the procedure from the perspective of how it would be implemented using one of two proposed interaction concepts (called TrainAR and Decide-Freeze-Imitate, explained in the following subsection) in the AR training itself. In this, it is considerably closer to the final training and could therefore be used as a reference not only to initially bootstrap the AR training but, even more importantly, during the iterative formative stages as a reference to where changes are necessary. From the HCI perspective, this was helpful because of the complexity of the procedures in general, but also more specifically helped to create an early shared perspective of how the final AR training could be realized. While the stateflow diagrams developed during Project Heb@AR are not openly available, the Appendix of this thesis includes exemplary stateflow diagrams (Figure 1, 2, and 3) that were used for the evaluation of the TrainAR authoring tool.

4.3.5 Technical Development of the Initial Prototypes

Based on these provisional stateflow diagrams in the Miro software, the initial prototypes of the AR trainings were then technically implemented by the HCI team. The AR trainings were hereby developed mainly using the Unity game engine [4] and version-controlled through the Gitflow branching model with feature-branches [115] on GitHub. All trainings were technically implemented based on the ARFoundation 4.2.0 [5], ARCore 4.2.0 [6], ARKit 4.2.0 [7] libraries, as this allows them to be deployed for Android and iOS for both mobile devices, but also tablets. Furthermore, Vuforia 9.8.8 [8] was used for trainings that required AR-marker tracking, as early technical explorations revealed that the AR-marker tracking of Vuforia was considerably more robust than the AR-marker tracking functionality provided through the already used ARFoundation library. Moreover, the Unity localization package 1.2.1 [9] was used for all textual content of the AR trainings, to provide them in multiple languages. For the multi-user trainings, Proton BOLT [10] was used for the implementation of the smartphone connectivity, and ZXing.Net [11] was used for the creation and interpretation of QR codes that are used by the multi-user trainings to connect multiple smartphones.

Outlining the components of technical development efforts, they were split into three categories: The creation of the AR training's assets (like 3D models, sprites, images, animations, videos, audio files, and physical AR-markers), the development of the interaction concepts for the training, and the connection of the flow of states, using the assets and interaction concepts, inside statemachines.

The assets were created through the usage of a multitude of tools like Blender, Adobe Photoshop, and Adobe Premier with a deliberate focus on ensuring their performant usage on handheld AR devices. For hard-surface objects, the assets were generally first conceptually drawn, then implemented as low-poly versions for the simultaneous development of the interaction concepts, and then based on the real physical objects measurement, realized as high-poly versions with complex interactions, physics, shadows, and transparency effects (see Figure 4.4). The organic assets' modelling and the modelling of more complex hard-surface assets was done based on the videos of the documentation procedure and additional resources, like medical procedure reference books. If measurements of objects were important, e.g., to contextualize size and orientation of the assets on the training dummies, 3D scanning was utilized.

The development of two entirely novel interaction concepts was necessary, due to the combination of specific learning goals and the given limitations of smartphone-based AR, e.g., having a smartphone in one of the hands but having to perform two-handed medical procedures. Additionally, existing interaction concepts were rather static in scope and, even for purely virtual AR trainings, offered little support to realistically implement procedural chains of actions for task

[4] https://unity.com/
[5] https://docs.unity3d.com/Packages/com.unity.xr.arfoundation@4.2/manual/
[6] https://docs.unity3d.com/Packages/com.unity.xr.arcore@4.2/manual/
[7] https://docs.unity3d.com/Packages/com.unity.xr.arkit@4.2/manual/
[8] https://library-archive.vuforia.com/articles/Solution/vuforia-engine-package-hosting-for-unity.html
[9] https://docs.unity3d.com/Packages/com.unity.localization@1.2/manual/index.html
[10] https://www.photonengine.com/bolt
[11] https://github.com/micjahn/ZXing.Net

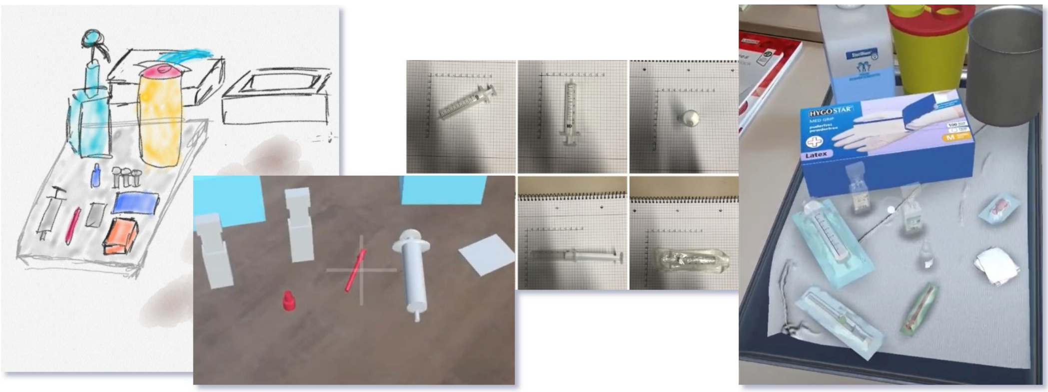


Figure 4.4: Examples of the prototyping process of models and their interactions from paper prototyping, through low-poly models with simple interactions, to modeling the final versions.

training. The combination of the AR trainings' assets and the interaction concepts was then done through, depending on the scenario, visual-scripting based or manually coded statemachines.

Development of the TrainAR Interaction Concept

The first interaction concept we developed as a means of interacting with the AR training's content was the TrainAR interaction concept. This purely virtual AR implementation lets the trainee place a virtual training assembly with virtual medical material onto a table, without the need for any physical medical consumables or AR markers. Afterward, as described in Figure 4.5, the trainee can grab, move, rotate, interact with, and combine the virtual AR objects through the usage of the smartphones position in physical space and the usage of onscreen buttons to complete a procedural chain of actions.

The idea behind TrainAR, in contrast to visualization-only approaches utilizing AR, is to use the purely virtual form of AR for contextualized visualization purposes and its interactivity to enable embodied interactions with the learning content by requiring the user to actively and deliberately utilize arm- and hand-movements in physical space to accomplish the tasks during the training. In contrast to conventional (non-AR) gamified procedural learning tasks, in this, psychomotor learning is included into the learning task to provide a more holistic learning experience. Or, as Lindgren et al. [286] described it, "the design rationale is that having learners act out and physicalize the systems, processes, relationships, etc., that they are trying to understand [...] will create conceptual anchors from which new knowledge can be built." Therefore, the AR training incorporates all three domains of learning: cognitive, affective, and psychomotor. This should not only gamify the training, but also increase intrinsic motivation for students to engage with it. Additionally, there is a good body of work indicating that increased interactivity of immersive technologies leads to improved learning outcomes [6, 9, 153].

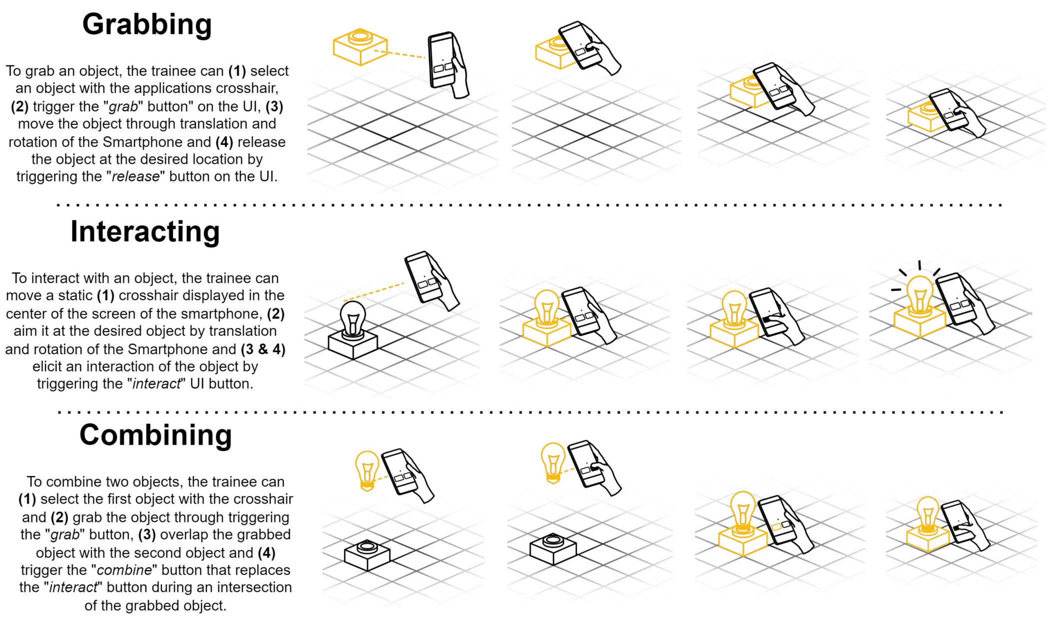


Figure 4.5: The TrainAR interaction concept: Grabbing & moving, interacting with and combining virtual AR objects to complete a procedural chain of actions.

TrainAR, with its purely virtual AR design, furthermore draws upon insights from Knierim et al. [236], who found that tangibility in AR trainings had no significant impact on learning outcomes and knowledge transfer, while significantly increasing setup-times. Therefore, while we still define TrainAR trainings as AR, we utilize this “VR-like” approach of purely virtual learning content in AR and still utilize Smartphones as the hardware choice without requiring additional physical material. Ultimately, this is what should increase the scalability of the AR intervention, enable BYOD approaches, enable self-determined trainings at home, and increase usability because of familiarity and the introduction of deliberately few obstacles for first-time users.

> TrainAR inside this Heb@AR chapter refers to the TrainAR interaction metaphors. In later chapters TrainAR is described in more detail and the term refers to the TrainAR Framework, which is the interaction concept paired with a didactic framework, and an AR authoring tool as a holistic solution for procedural AR trainings on handheld AR devices. (see Chapter 6 & Chapter 7)

Development of the Decide-Freeze-Imitate Interaction Concept

The second interaction concept we developed is called the “Decide-Freeze-Imitate” concept, which is described in Figure 4.6. It is an interaction concept designed to implement trainings that supplement the practical trainings in SkillsLab environments. Here, the trainees have to make **decisions** in a non-linear procedural flow on which actions they would want to perform next. They are then provided feedback based on that decision and, if the decision was correct, can observe the action visualized in AR, contextualized on a physical training dummy. If the decision was incorrect, they are provided feedback why this decision would not be appropriate and asked to reconsider. After the correct decision, they can **“freeze”** the AR view from any angle so that the action is still clearly visible. Trainees are then asked to put the device aside and **“imitate”**, therefore actually perform, the observed action on a physical training dummy using the physical consumables provided during the SkillsLab session. Furthermore, trainees can be instructed to actively collect parameters, such as respiratory or heart rate, or to train communication aspects. In these cases, during the training, the app takes over the role of the visual/auditory simulation.

The idea behind the Decide-Freeze-Imitate interaction concept is to be able to train both cognitive strategies and the corresponding motor components of a procedural chain of actions (e.g., not only knowing what to do next but also knowing the specific grips used during a reanimation procedure), while still retaining some of the BYOD methodology benefits and the scalability of cost-effective AR device usage. Additionally, the contextualized AR visualizations allow the demonstration of steps and principles, which would not otherwise be available to the trainee and the AR training takes the role of a simulation but also examination of the students’ decisions. On the other hand, these benefits come at the cost of having to displace the smartphone during the training, e.g., compared to using HMD-based AR approaches, being more location-dependent, dependent on expensive training dummies, and having to consume medical training material.

4.3.6 Iterative Improvements & Amalgamation of AR Trainings

After the initial bootstrapping of the prototypes, they were iteratively evaluated and improved upon, following the Desing-Based Research methodology [394], until they were finalized. The stakeholders during this methodology were the students, the lecturers, external medical domain-experts (e.g., a pediatrician) and other researchers. Here a focus was on ensuring the interaction concept’s usability, the content’s correctness, and ensuring a sufficiently understandable onboarding utility. One exemplary formative evaluation is described in the TrainAR chapter in Section 6.3.

When an AR training reached sufficient maturity, it was finally merged into an overarching cross-platform compatible project in Unity, from which the Heb@AR App was ultimately built. In the Gitflow branching model, this would be considered a release branch [115]. To provide a rough estimate of the size of the final amalgamation of all AR trainings and their supporting app structure: At the time of writing this thesis, this includes custom C# functionality written by the team with over 26067 lines of code (excluding statemachine files, localization scripts, 3824 blank lines and 3193 lines of comments) in 292 C# files, 14018 words of fully localized written textual content (counted is the German localization only), about 80 audio files with over 20 minutes of

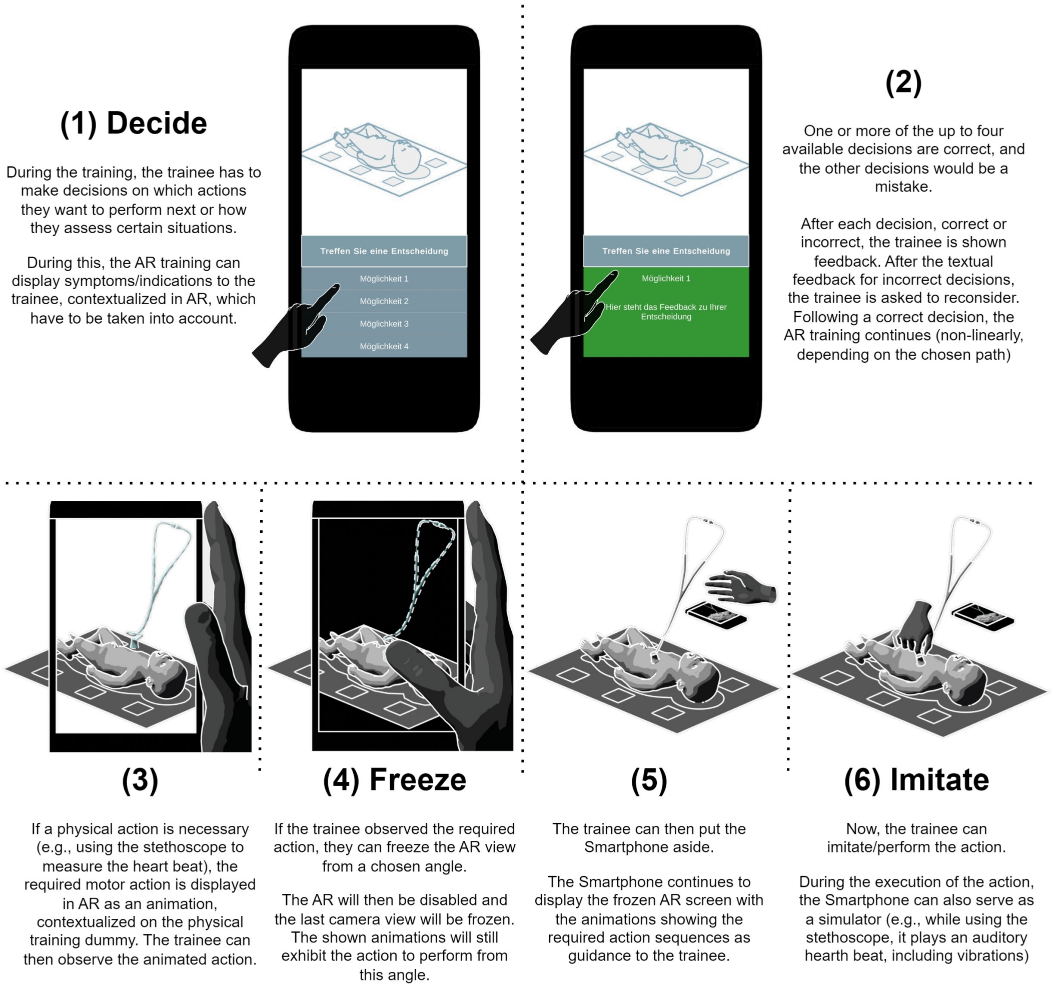


Figure 4.6: The Decide-Freeze-Imitate interaction concept: Deciding which action to perform next, getting detailed feedback on the decision, viewing the required motor component, freezing the AR view, putting the Smartphone aside, and performing the observed motor action.

total playtime, over 100 custom-made figures/images/sprites, 21 2D animations/videos, and over 60 custom-made 3D models with 74 3D animation clips. (Calculated on the 25.05.2023, based on the Heb@AR App GitHub repository, using CLOC and the GitHub repository search)

4.3.7 Outline of the Development Timeline

A rough outline of the development timeline is visualized in Figure 4.7. While the GANTT bars do not necessarily represent full-time efforts spent towards this specific task and the iterative improvement efforts are combined into the "Development" and "Formative Evaluation" bars, it provides a rough impression of how long the development from initial conceptualization to curricular implementation of each training took from the perspective of the previously described methodology. Additionally, it shows, when they were evaluated by which group. We will reference back to this information from the evaluation section, discussion section, and the following chapter where we explore the design space of AR authoring tools.

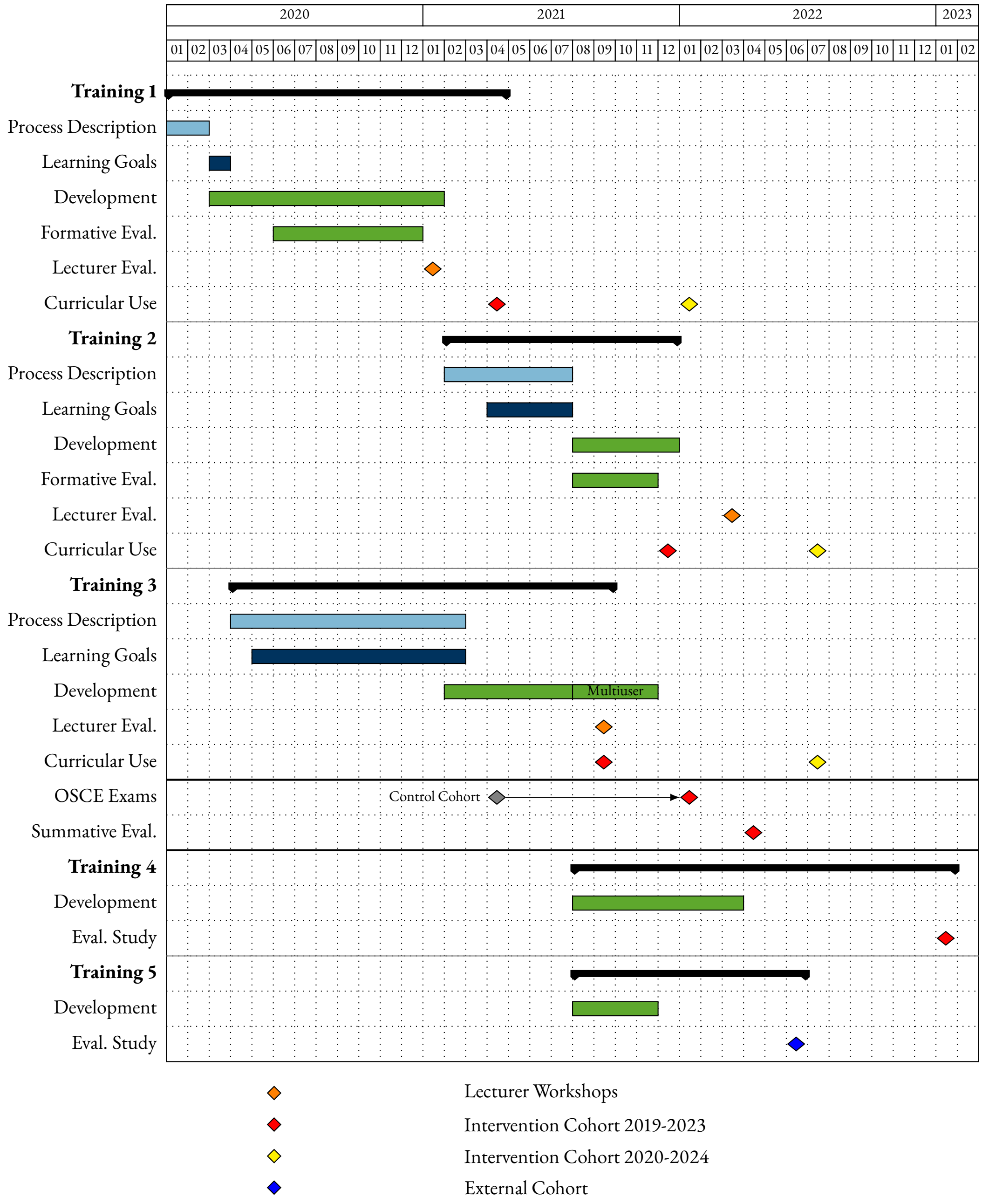


Figure 4.7: An outline of the development and evaluation timeline of all AR trainigns developed during Project Heb@AR. Visualized from process description, over learning goal formulation, to technical development and formative evaluation as GANTT bars. Milestones visualize the evaluations. Black bars show the time between starting development to it's first curricular use.

4.4 The Heb@AR App

One of the final results of Project Heb@AR, the developed Heb@AR App, is more than just a prototype. It is already freely available as an Open Educational Resources (OER) in the Android and iOS App stores, and supplementary material, like AR markers, medical material lists, reference videos, and feedback forms, are available as OERs on GitHub under the CC-BY 4.0 license. The App contains the five iteratively improved midwifery-specific AR trainings with consistent interaction, multi-layered feedback, and self-assessment concepts, trainings with adjustable difficulty levels, multi-user trainings, and supplementary materials such as AR markers, material lists, and instructions on how to use the app for lecturers who want to deploy it. Each individual AR training is always contextualized in a realistic midwifery training situation, offers training-specific onboarding for the AR interactions, and allows trainees to choose whether they would like to receive additional optional expert insights and tips from practice in audio and textual form.

Besides these deliberately consistent concepts across all included trainings, the five AR trainings in the app also differ in other dimensions (see Table 4.1). For example, three of the apps trainings are virtual "Training@Home" trainings, based on the newly created "TrainAR" interaction concept (as desribed in Section 4.3.5). These trainings can be used by students as location-independent, self-controlled SkillsLab preparation opportunities or as retention trainings afterward, to independently rehearse the knowledge learned in practical sessions or lectures. They train the declarative and procedural knowledge of action sequences but are purely virtual in form, therefore do not need any physical material. With this, on the other hand, they also do not train the motor components needed to perform the trained sequence actions [145]. These three trainings are the *"Preparation of an emergency tocolysis"*, the *"Resuscitation of a newborn using a Resuscitation Unit"* and finally the non-procedural training *"Anatomy of the female pelvis"*.

The more comprehensive trainings included in the Heb@AR App are the AR-SkillsLab exercises *"Resuscitation of a Newborn"* and *"Preparing a Pregnant Woman for a Cesarean Section"*. These two trainings are intended to be directly integrated into the curricula of universities through SkillsLab session in terms of time and space. They serve as simulations during the training sessions, where students receive visual and auditory information from the App directly contextualized on training dummies, have to make decisions on what actions to perform next and put the AR device aside to actually physically perform the action afterward. In this, the AR-SkillsLab exercises train both, the cognitive skills and strategies of a course of action, but also the associated motor components to perform them [145]. To accomplish this, training dummies and physical material are necessary during the training, and they are not location-independent. These trainings are based on the newly developed Decide-Freeze-Imitate concept (as described in Section 4.3.5).

4.4.1 Central Design Concepts

In line with the strategies for the design of effective user interfaces proposed by Schneiderman et al. [435], the central design concepts of the Heb@AR App strive for consistency in colors, button and information placement, naming of actions and form of features. It furthermore tries to be

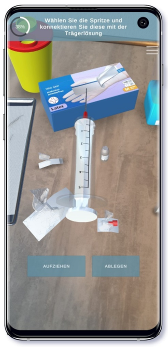
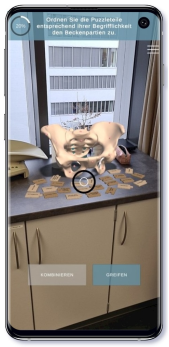
Preview					
Description	(1) Preparation of an emergency tocolysis	(2) Preparing a Pregnant Woman for a Cesarean Section	(3) Resuscitation of a Newborn	(4) Resuscitation of a newborn (Resuscitation Unit)	(5) Anatomy of the female pelvis
Type	Training@Home	SkillsLab Exercise	SkillsLab Exercise	Training@Home	Training@Home
Learning Goal	Procedure	Procedure, Motor component	Procedure, Motor component	Procedure	Concept
Needs Material		✓	✓		
Duration	15 minutes	45 - 60 minutes	30 - 45 minutes	20 minutes	15 minutes
Other Features	Includes 3 difficulty settings		Includes an individual and multi-user version		

Table 4.1: The 5 AR trainings included in the Heb@AR app, with their training name, training type, abstract learning goal, whether they need physical material during the training, a rough estimate of the training duration, and individual other features of the trainings.

universally usable, while always providing feedback to the user of the app to keep them informed at all times. It strives to keep the user as much in control as possible, though, because of the nature of medical procedural trainings, inside the AR trainings itself, actions that do elicit consequences are often not reversible, inherently breaking one of the principles. Our concept is furthermore in line with recent findings, that for learning tools, the most important UX quality aspects according to their users are "quality of content", "usefulness", "efficiency" and "clarity" [420]. Moreover, it is informed by the usability principles for AR applications on smartphones by Ko et al. [238], e.g., by creating defaults, familiarity, and visibility, using direct manipulation for interaction with virtual objects (ideally with as little physical effort as possible), while providing technical error handling and usage help/documentation at all times.

To accomplish this, five central design decisions were made, as visualized in Figure 4.8. Firstly, a medical inspired color palette was chosen for the app (see Figure 4.9). The primary color, pastel blue (#6895AC), is intended to convey calmness, tranquility, and trust. The other primary color, white to light gray (#EEEEEE) is intended to convey simplicity, and cleanliness. The other colors of the palette (#776DC3, #6DC299, #64AB35, #5B1261, #1B4165) are contrast colors, to distinguish different content and contexts. Especially, the contrast color orange (#ECB736) is used

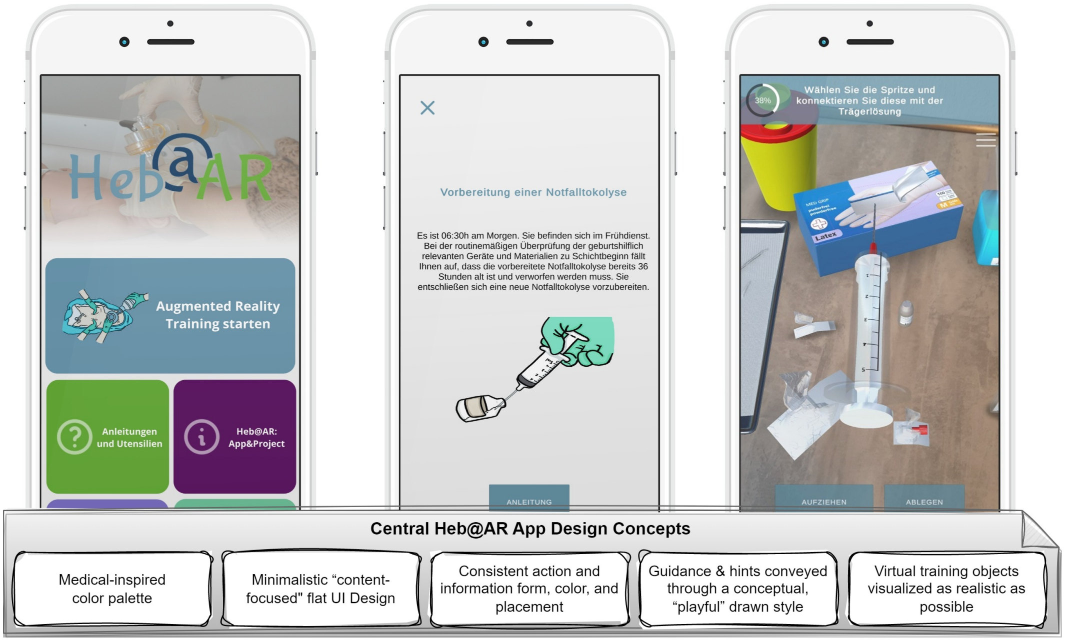


Figure 4.8: Central design aspects of the app: A medical-inspired color pallete, flat UI design, consistent button and information placement, playful guidance and context, and realistic AR objects.

Figure 4.9: The medical inspired color palette used for the Heb@AR App: Two main colors, pastel blue (#6895AC) and white/light gray (#EEEEEE), combined with six supplementary colors (#776DC3, #6DC299, #64AB35, #5B1261, #1B4165, #ECB736).

throughout the app consistently to guide the trainees' attention toward changes or important visualizations. Secondly, a minimalistic flat UI design was chosen in this color palette, to keep the UI content-focused, as the AR components of the app are likely already sufficiently stimulating to the user. Thirdly, buttons, information, and helpers are always consistent in size, placement, form, color, and function throughout the app. As Nielson described it, "consistency is one of the most powerful usability principles" [347], and as the AR components are likely already challenging for the users, we tried to maximize the UI consistency. Besides the modern flat UI design, the technical and contextual onboarding, expert hints, and training summaries are always conveyed in a stylized comic/drawn style to communicate the conceptual nature of the information and add at least somewhat playful aspects to the otherwise quite serious context and contrasting design aspects. Finally, the virtual Objects inside the AR training themselves are always modelled as realistic and close to the real object as possible with the limited hardware. This is done to ensure that objects found in the training are recognizable as the actual physical equivalent in reality.

As the main font, “Liberation Sans”, an open-source font out of the Liberation font family, which is intended as substitute fonts for proprietary fonts like Ariel or Times New Roman, was used throughout the app. In some action buttons, a capitalized version is used. Finally, the UI is fully responsive and automatically adjusts to different screen sizes, resolutions, and aspect ratios. With this, while designed for smartphones, the app is also fully functional on tablets.

4.4.2 Menus & Peripheral Features of the App

The Heb@AR App is built around the included AR trainings and only serves as a vehicle to bundle the trainings. In this, the app is structured fairly simple. When the app is started, a home screen shows the Heb@AR logo with a preview image showing a reanimation of a newborn. Right below this logo, the most prominent button is “Start Augmented Reality Training” (see Figure 4.10a), which, when clicked, leads the trainee to a menu where they can select which AR training they want to train and in which version (e.g., which difficulty for Training 1 or if they want to use the multi-user version of Training 3). To support the trainee in this decision, descriptions, duration estimates, and icons classifying the training are provided to them (see Figure 4.10b). When one of the trainings is started, the user is provided context onboarding and technical onboarding before the AR training starts. After an AR training is completed, the trainee returns to the main menu of the Heb@AR App. Besides starting the AR trainings, the trainee can also access “Tutorials and other utils” (see Figure 4.10c) from the main menu, e.g., to download the AR markers or lists of medical material necessary to train the AR SkillsLab exercises. Moreover, trainees can read more about Project Heb@AR (see Figure 4.10d), learn about how to create their own AR trainings similar to the Heb@AR app and open the settings menu (see Figure 4.10e).

Figure 4.10: The main menu (a) and the training selection menu (b) of the Heb@AR App. Moreover, menus for the supplementary materials for the trainings (c), information about project Heb@AR (d) and the settings menu (e).

4.4.3 Training 1: Emergency Tocolysis

The first training included in the Heb@AR App, that is started when the trainee selects it from the training selection menu (see Figure 4.10b), is the "Preparation of an emergency tocolysis". This training is intended as an introductory training session in which a labor inhibiting injection has to be prepared by the trainee. To accomplish this task, the trainee has to complete a sequence of actions in the correct order, by picking up, combining, and interaction with virtual material inside the app. This training's duration is roughly 15 minutes, and it is based on the TrainAR interaction concept. Therefore, it can be trained location-independent and trains the procedural component of the action sequence, but not the motor actions associated with it.

Learning Goals

Summarizing the more comprehensive and detailed elaborations of the learning goals that the midwifery [35] and medical didactic [217] project partners developed during Project Heb@AR, the learning goal of this training is for the trainee to know the correct preparation, sequence of actions, dosage of medications, and follow-up procedures and its documentation for the preparation of an emergency tocolysis. Additionally, the secondary goal is to furthermore know and correctly apply the hygiene and occupational safety measures within the action sequence.

Flow of the Training

When the trainee starts the training, they are first shown a case-based training description, which contextualizes the AR training into a hypothetical scenario to enable self-directed learning, critical thinking, and application of the trainee's knowledge. Afterward, the trainee can decide if they want to receive additional expert hints, described as "knowledge from practice" to the trainee, by a virtual expert midwife during the training. If the trainee decides to receive these expert hints, they will be displayed textually under the top UI element during the AR training and played as audio clips (see Figure 4.11). Moreover, the trainee is then shown technical onboarding on how the TrainAR interaction concept is utilized to interact with objects during the training based on textual explanations and animations in line with the concept figure 4.5. Then, the AR component is started, and the trainee is guided by animations to place the virtual training assembly onto a table in front of them. Afterward, the training starts and the trainee has to prepare the tocolytic injection, where they have to disinfect their working area and hands and put on gloves. Now trainees have to select the material out of a set of appropriate objects and distractors. The distractors include, for example, the wrong needle, the wrong carrier solution, and a carrier solution which expiration date is exceeded. Afterward, the student has to pull up the syringe with the carrier solution and medication in the correct order, label the syringe, and perform all follow-up procedures. Instructions on what action, or group of actions, the trainee has to perform next and indications of the trainee progress are provided on the top UI element (see Figure 4.11). At the end of the training, the trainee is shown a training summary, which shows the required time, numbers of errors made, and specific textual feedback for severe errors (e.g., trying to pull up the medication before the carrier solution would be a severe mistake). A full reference video of the training is available on YouTube [12], licensed under CC-BY 4.0.

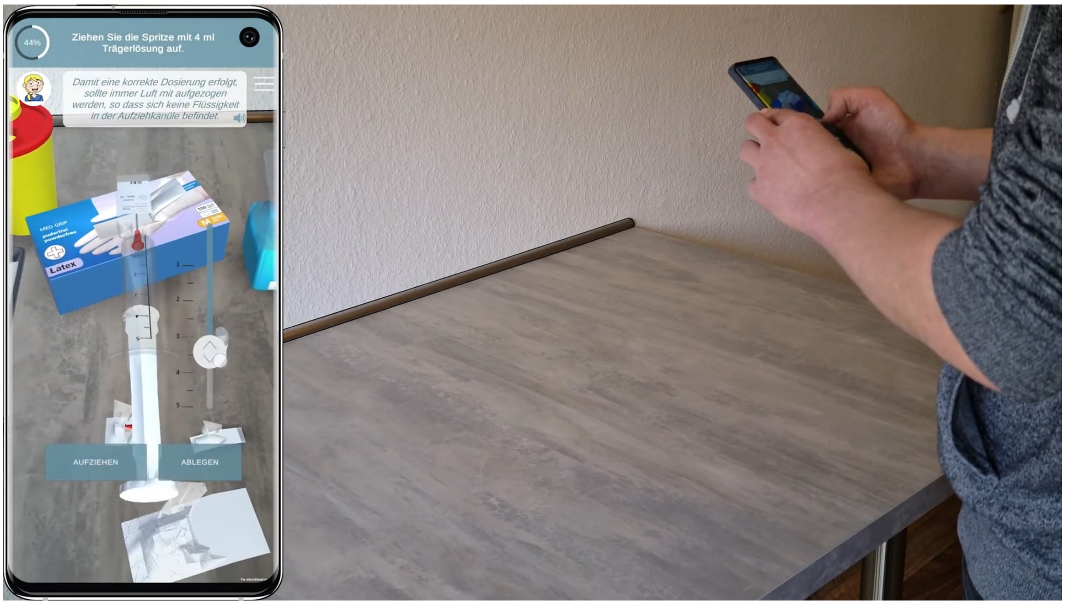


Figure 4.11: Training 1, "the preparation of an emergency tocolysis" in the "guidance mode". Shown in the figure is the step of drawing up the carrier solution into the syringe. The top UI element guides the user on what to do next, while the expert midwife provides tips on what is important to take into consideration when pulling up the carrier solution in reality.

Additional Feature: Difficultly Settings

The "preparation of an emergency tocolysis" AR training was the training we selected during project Heb@AR to explore the implementation of decreasing guidance during repeated uses of the AR trainings. The training therefore includes 3 difficulty modes that decrease in provided instructions and feedback. In the first mode, the "guidance mode", the user receives instructions on each action to perform next. If they make repeated mistakes during the training, they are shown additional textual guidance on which objects specifically would be correct to use for the step. In the "training mode", the trainee only receives instructions on which groups of actions they have to perform, but is no longer instructed specifically which step of the action sequence they would have to perform next. This information is not given until the trainee makes repeated mistakes, where the more granular instructions are provided as feedback. In the "exam mode", the trainee receives no instructions, they are simply instructed to prepare the emergency tocolysis. Furthermore, only guidance about potential technical challenges is provided to the trainee as feedback, but it would

[12] Reference video of Training 1: https://www.youtube.com/watch?v=CUyuzIkvvuk

also not provide hints about which specific action to perform next. The "exam mode" is not intended to be used as an actual examination of students' knowledge, but rather as a self-check retention test opportunity for the students.

4.4.4 Training 2: Preparation for a Cesarean Section

The second training of the Heb@AR App is the "preparation of a birthing person for a cesarean section" (see Figure 4.12), the first emergency training intended to be used in the SkillsLab as an on-site exercise. This training is only partially covered by the AR component of the app and is reliant on the supporting teaching concept using WBTs. The AR training part alone is still the largest AR training of the app with a duration of roughly 45–60 minutes, is focused on the preparation of placing a permanent bladder catheter in preparation for the cesarean section, and is intended to be used during a practical training session in the SkillsLab of a university. The Decide-Freeze-Imitate interaction concept is utilized for this training, which enables the incorporation of motor aspect of the procedural action sequence into the AR training, but also makes the training location-dependent and requires the usage of specific training dummies and medical disposables.

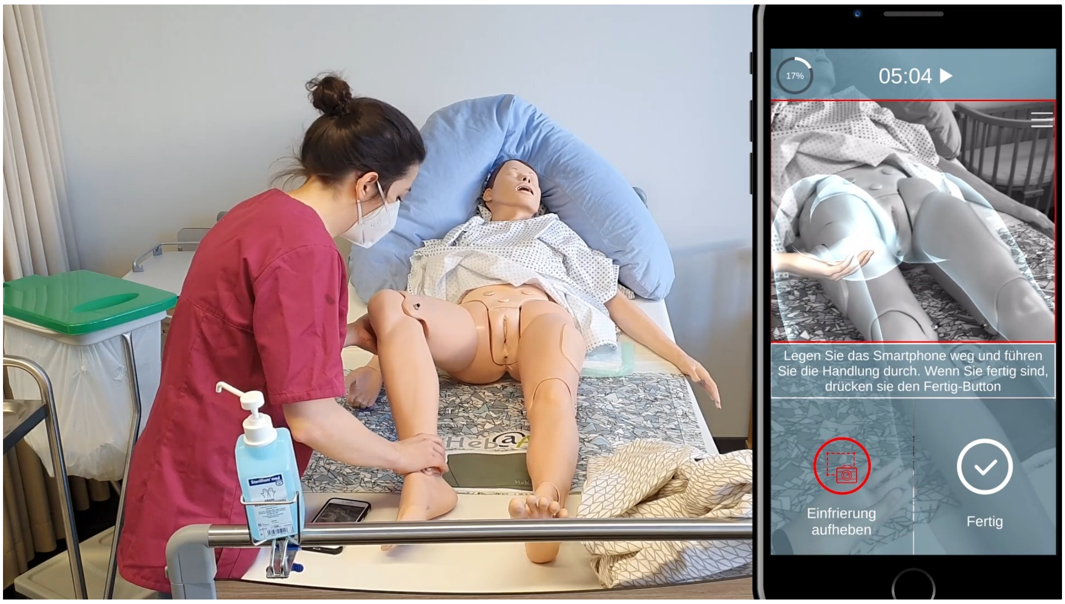


Figure 4.12: Training 2, "the preparation of a birthing person for a cesarean section". Shown in the figure is the step of positioning the legs of the birthing person in preparation for the placement of the permanent catheter.

Learning Goals

Summarizing the overarching learning goal, developed by the project partners [35, 217], the main learning goal of the training is for students to be able to identify the needs for action in preparation of a birthing person for an urgent cesarean section and to be able to initiate and perform appropriate, adequate emergency management.

The specific learning goals for the AR training component in this are to know and apply the procedures (preparation, execution, post-processing) for placing a transurethral permanent bladder catheter within a meaningful sequence of actions. For this, trainees have to correctly identify and physically use the required materials for placing the catheter within the action sequence. Trainees furthermore have to know and apply hygiene measures within the action sequence.

Flow of the Training

After starting the training, the trainee is first shown all the physical material necessary to complete the training (e.g., which training dummy, catheters, and types of gloves should be used). Afterward, the trainee is shown the case-based training description and asked if they want to receive expert input during the training. The user is then given an explanation on how the AR component of the training works and how to use the app. When the training itself starts, the trainee has to make decisions on which action they want to perform next. After making a decision, the trainee is always provided feedback. Afterward, the trainee is shown how to perform the action, either as an onscreen video (e.g., for bringing the bed in the correct position or disposing of material) or as an AR animation contextualized on the training dummy (e.g., for properly disinfecting the area and inserting the catheter). They can then place the smartphone aside and physically perform the action. Afterward, they can pick the smartphone back up and continue with the training. After the training concludes, the trainee is given a detailed training summary, showing their needed time, incorrect decisions, and detailed feedback when specific incorrect decisions were made during the procedure. A full reference video of the training is available on YouTube [13], licensed under CC-BY 4.0.

4.4.5 Training 3: Reanimation of a Newborn

The "reanimation of a newborn" is the third training of the app and is available in two versions: An individual AR training and a multi-user training. During the AR training, the trainees have to reanimate a newborn with transitional difficulties after birth, based on elicited symptoms, in accordance to the ERC guidelines on the reanimation of a newborn. This AR training is a SkillsLab exercise, based on the Decide-Freeze-Imitate interaction concept, which takes roughly 30–45 minutes to complete and requires the use of physical training dummies and medical disposables.

[13] Reference video of Training 2: `https://www.youtube.com/watch?v=dw4dJnryNMs`

Learning Goals

According to the specifications developed by the project partners [35, 217], the learning goal of the training is to be able to assess and evaluate the newborn immediately after birth. Therefore, students have to be able to identify the need for action in case of transitional difficulties of the newborn and initiate and execute appropriate, adequate emergency management. Importantly, beside being able to perform the correct "algorithm" according to ERC 2021 guidelines [300], this means, they also have to know how to execute the necessary motor movements, e.g., the c-grip to properly place the mouthpiece of the respiration bag onto the newborn's mouth and nose.

Figure 4.13: The individual training version Training 3, the "reanimation of a newborn". Shown in the figure is the task of using the stethoscope to determine the newborn's heart rate. The Smartphone's AR view is frozen, while the action is performed, and the smartphone simulates the heart beat through audio and vibration.

Flow of the Training

In line with the flow of Training 2, the trainee is again shown all the necessary physical material, the case-based training is contextualized, and the usage of the AR app is explained. Afterward, the training is started, and the user is instructed to overlay an outline of a newborn with the newborn dummy on the AR-marker pad. After the AR-marker is recognized by the smartphone, the training begins. During the training, the trainee has to make decisions on which action they want to perform next according to the ERC guidelines to reanimate a newborn, based on simulated symptoms. They are then always provided feedback to their decision. If they have to perform physical actions themselves, the trainee can watch the necessary action (e.g., using a stethoscope, as shown in Figure 4.13), freeze the AR view, put the smartphone aside, and then perform that action themselves, while the Heb@AR App acts as a simulator (e.g., in this case using audio and vibrations to simulate the heartbeat). Symptoms are generated during the training; therefore the decisions have to be made based on the elicited symptoms and the correct procedural chain of actions can change between training runs. After the newborn is reanimated, the trainee is instructed to inform the parents and is shown a training summary with their time, errors, and textual hints pointing toward severe errors. A full reference video of the training is available on YouTube [14], licensed under CC-BY 4.0.

Additional Feature: Multi-User Training

The "reanimation of a newborn" is the AR training we selected during Project Heb@AR to explore the potential of multi-user AR trainings. In the implementation we chose, the training's interaction concept, content, and sequence are equivalent to the individual training version. The difference is that multiple smartphones can connect to each other using QR-codes (see Figure 4.14 (left)) and then individually make decisions on which action the trainees think would be the correct action to perform next. After all participants of the training made their decision, all decisions are displayed to the participants (see Figure 4.14 (right)). Users of the training, which made the wrong decision, are shown feedback on why the decision would be incorrect and are then automatically guided to the correct decision. If multiple decisions were correct, the decision which was selected the most is used.

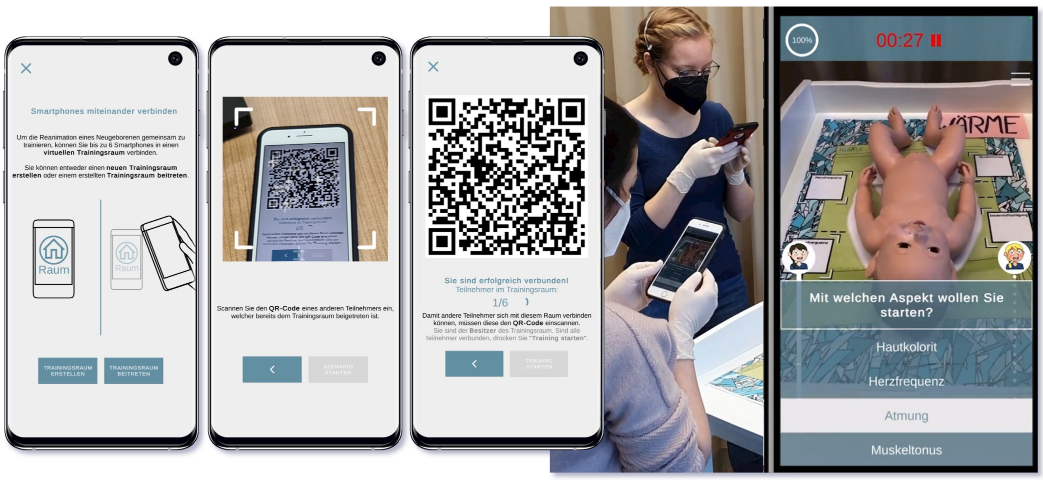


Figure 4.14: To use the multi-user version, up to 6 smartphones can connect to each other using QR-Codes. (left) During the training, all participants make decisions on which action they would perform next. After everyone submitted their decision, they are shown on the UI. (right)

[14]Reference video of Training 3: `https://www.youtube.com/watch?v=KoGDslW4abM`

While this procedure of the multi-user version of the training is explained during onboarding of this version, before starting the training, we deliberately do not state whether it is intended to be used competitively or cooperatively. As both approaches of using this training are supported by the implementation, we leave this decision up to the trainees or educators.

In line with the individual training version of Training 3, during the multi-user versions, the trainees have to actually perform the motor actions, which would be required during the algorithm of reanimating a newborn. Here, one of the trainees is selected to freeze their AR view and actually perform the action on the training dummies, while the other trainees are instructed to compare the instructions they see in the AR view with the actions performed by that trainee. Technically, a token system is used, where a trainee earns one token after performing the motor action and the next trainee to perform an action is always selected among the trainees with the lowest number of tokens, to ensure that the motor components are evenly distributed among the trainees. As visualized in Figure 4.15, for some motor components of the AR training, like the combination of the chest compression and bag-valve-mask ventilation, multiple trainees are selected to perform the action cooperatively. There is also a full reference video of the multi-user version available on YouTube [15], licensed under CC-BY 4.0.

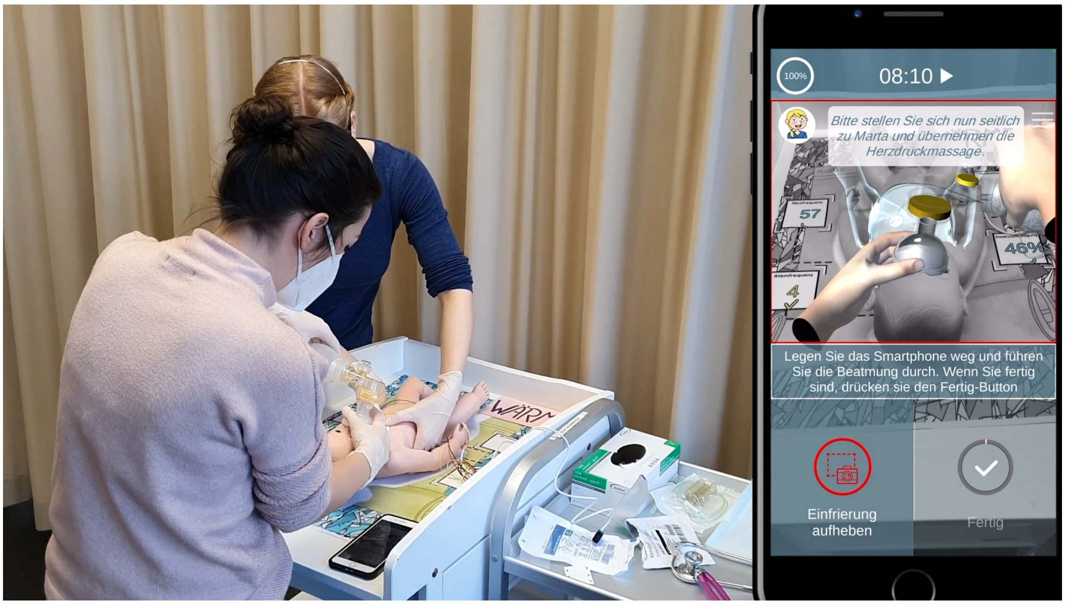


Figure 4.15: The multi-user version Training 3, the "reanimation of a newborn". Shown in the figure is the cooperative task, where one trainee is instructed to perform chest compression, while the other is instructed to perform bag-valve-mask ventilation.

[15] Reference video of multi-user version of Training 3: https://www.youtube.com/watch?v=mEp0kCC-Sug

4.4.6 Training 4: Reanimation of a Newborn (Virtual)

Training 4, the virtual version of the "reanimation of a newborn" is a Training@Home variant of the SkillsLab emergency situation exercise of reanimating a newborn. While the main procedure of the reanimation itself and the way trainees have to make decisions during the training through the Decide-Freeze-Imitate interaction concept is equivalent, the training is purely virtual, and therefore the motor actions do not have to be performed on a physical training dummy. When motor actions are necessary for the current step of the action sequence, they are simply visualized and can be observed by the trainee, before proceeding, as visualized in Figure 4.16. In this, the motor components are not actively trained, but no disposable material is used, and the training is location-independent through the combination of both interaction concepts. Furthermore, the virtual version of the training includes a reanimation unit. Trainees can optionally use the reanimation unit during the training, e.g., to display vitals on the patient monitor or use the suction of the unit to suction mucus that clogs the respiratory tract. Notably, they can still use the manual methods (e.g., using a stethoscope) as learned in the SkillsLab exercise, increasing the trainees degree of freedom in terms of possible correct choices.

Figure 4.16: Training 4, the virtual version of the "reanimation of a newborn using a reanimation unit". Shown in the figure is the step of using a stethoscope to determine the breathing rate of the newborn. While the trainee is not required to actually use the stethoscope, they have to listen to and count the breathing rate, which will be queried in the next step by the app.

Beside the actual motor movements not being necessary, the learning goals are largely in line with Training 3, as it is intended as a Training@Home retention opportunity of the AR SkillsLab exercise. Some symptoms during the training deliberately deviate from the SkillsLab version, e.g., while the respiratory tract is free in the SkillsLab version, in the virtual version it is not, and therefore it is necessary to suction the mucus before ventilating the newborn. Moreover, the virtual version has an optional pre-training, which allows the trainees to familiarize with the preparation of the reanimation unit, before its usage in the emergency situation. A full reference video of the training is available on YouTube [16], licensed under CC-BY 4.0.

4.4.7 Training 5: Female Pelvis

The fifth training, that we developed for the Heb@AR App, is the "denomination of the female pelvis". It is a serious game that allows trainees to interactively learn the regions and bone structures of the female pelvis (see Figure 4.17). To ensure the game's scalability and self-determined usage at home, it is based on the TrainAR interaction concept (see Section 4.3.5). During the training, which takes roughly 15 minutes to complete, students have to contextualize German-Latin word pairs in the form of puzzle pieces to the correct region or bone structure of a virtual 3D model of the female pelvis. To accomplish this task, they can either resolve the German-Latin pairing first and then combine the correct pair with the pelvis model, or combine the individual German and Latin pieces individually with the 3D pelvis model. The AR component of the game is used herein to enable embodied interactions, by requiring users to use deliberate hand and arm movements in physical space to pick up puzzle pieces and connect them with the correct area or part of the pelvis. The idea is that this added interactivity is not only more fun for the user, but also increases retention, as described in Section 4.5.

Learning Goal

While the game is primarily targeted at midwifery bachelor students as a self-directed retention opportunity for the historically unpopular and "dry" subject of learning the bones and regions of the female pelvis, it can also be used by nursing or medical students. The learning goal of this training is for students to be able to correctly identify and name all regions and bones of the female pelvis in German and Latin. No previous knowledge is required to complete the training.

Flow of the training

After reading a case-based training description, the trainee completes the technical onboarding that explains the mechanics of AR and chooses if they want to receive hints by a virtual training partner agent during the game. Then, they start the training by placing the virtual assembly with the pelvis and all puzzle pieces on a desk using markerless AR, in line with Training 1. The game itself is split into 2 levels. In the first level, the learning goal is to familiarize with the three bone regions and their German and Latin names. This is implemented as a labeling puzzle, where the two corresponding names (German and Latin) have to be identified within six pieces that are scattered around the pelvis. The trainees have to literally pick up the pieces by approaching them

[16] Reference video of Training 4: https://www.youtube.com/watch?v=FB7izeXjDWo

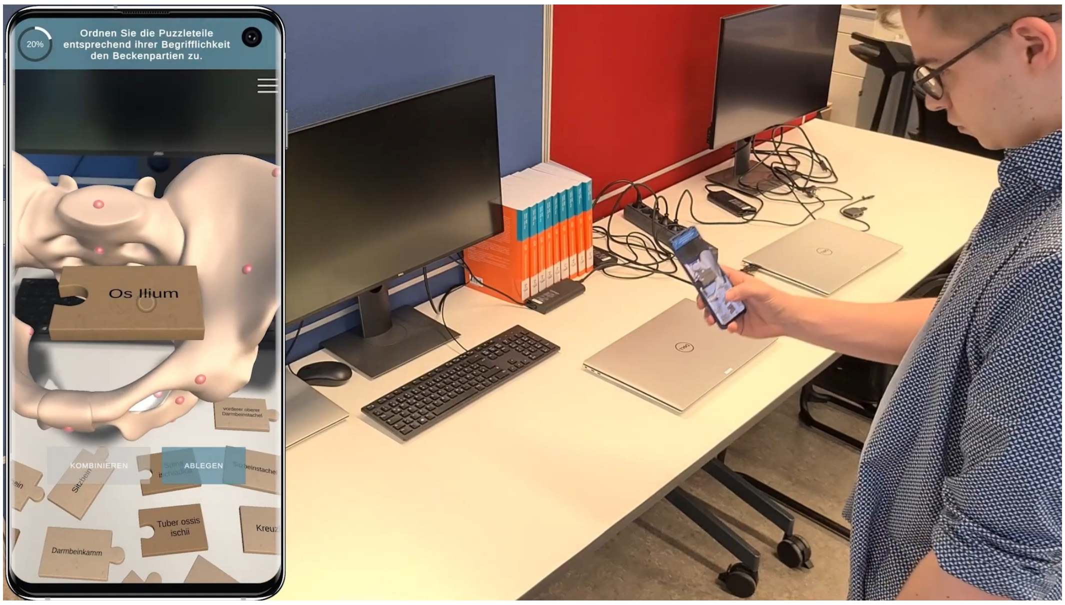


Figure 4.17: Training 5, the "denomination of the female pelvis". Shown in the figure is a grabbed Latin puzzle piece to be connected with the correct bone structure of the virtual pelvis model.

with the smartphone and grabbing them by the press of a button when they are in proximity to the center of the screen, which is signalled by a reticle. The grabbed piece then has to be overlapped with the matching piece or the matching region, and can be combined with the press of a button.

Afterward, the more comprehensive task of contextualizing the 12 German and 12 Latin puzzle pieces for the bone-structures has to be completed using the same process. During the game, the optional agent, which serves as the virtual training partner, can provide textual and auditory tips to the user. If the app detects multiple or repeated incorrect interactions, an error overlay can provide feedback for actions that are not possible. After the game is completed, the app displays a summary and performance assessment to the user for self-reflection. A full reference video of the training is available on YouTube [17] licensed under CC-BY 4.0.

4.4.8 Supplementary Heb@AR Material

As some of the trainings require additional physical medical material and are contextualized onto the training dummies using AR markers, a GitHub repository [18] contains the AR markers for Training 2 and 3 as .pdf and .png files and two material lists, detailing which medical utensils are necessary to use each training. This supplementary material is also accessible through the

[17] Reference video of Training 5: https://www.youtube.com/watch?v=arTJ3lrHRkw
[18] Heb@AR GitHub Repository: https://github.com/Mixality/HebAR

Heb@AR App itself, directly linking to the GitHub repository. Through this, it is ensured that the supplementary material used is always up-to-date.

Furthermore, this repository also includes the privacy policy of the Heb@AR App, which is necessary to publish it to the app stores, forms to request changes to specific AR trainings, a review procedure on how these requests would be evaluated, and a changelog for all released versions of the app into the app stores. Additionally, the Android manual installation files (.apk) are provided in the repository, for example, to manually install them on institutional phones, which do not necessarily have access to the commercial stores.

4.4.9 Long-term Vision for the App & Analytics Functionality

One of the core ideas of the Heb@AR App is bundling all the AR trainings developed during Project Heb@AR into a single application, to enable BYOD approaches. By using students' smartphones, the app can be used independently by other institutions in curricular teaching as well as by midwifery students independently, making the Heb@AR App scalable today. We recommend the Training@Home trainings for self-determined preparation or consolidation of, in the SkillsLab taught learning content, and the AR-SkillsLab exercises as a multimedia supplements to the SkillsLab exercises themselves, accompanied by qualified tutors. The Heb@AR app is available for free in the Android & iOS App Store and uses the common AR interfaces ARCore and ARKit. Thus, the Heb@AR app can be used on 48 [19] iOS and 707 [20] Android models (as of 06.2023). On the one hand, this makes the app accessible as one of the results of Project Heb@AR, which can be used sustainably as an OER, expanded upon, and implemented into curricular of universities beyond the scope of the evaluation. As visualized in Figure 4.18, the app was, for example, already implemented in curricular teaching of the midwifery bachelors program at Bielefeld University of Applied Sciences and Arts [381].

On the other hand, this approach paves the way for long-term usage analytics and evaluations of AR training application usage, which in terms of length, depth and overall scope, is missing in the AR training literature today. Especially, the evaluation of actual usage of provided AR trainings for self-determined learning at home could contribute significantly to our understanding of the benefits of AR learning applications of that kind. While it is not currently active in the app store version of the app because of privacy considerations and potential legal liabilities, the app does include features, which allow tracking several aspects, e.g., AR training times, training durations, how often specific trainings were used, which decisions trainees made during the training, where inside the training trainees usually made errors or quit the training, whether there were specific technical problems during the training, and many more technical aspects to understand the landscape of devices using the app. With the implemented functionality, it is even possible to perform process mining to understand how trainees acted during the Heb@AR trainings, e.g., which decisions were most common in the non-linear action chains, on a large-scale basis.

[19] Manually counted based on: https://en.wikipedia.org/wiki/Timeline_of_Apple_Inc._products
[20] https://github.com/androidtrackers/arcore-devices

Ultimately, the results we will gather from this long-term analytics evaluation approach will likely provide unique contributions to the field of AR learning, but as of now, will not be part of the scope of this thesis. Nonetheless, from evaluations of the individual trainings during Project Heb@AR, we can gather preliminary insight whether the Heb@AR App is a useful learning application, which we will approach in the following sections.

Figure 4.18: Students train the preparation of a tocolytic injection, using the Heb@AR App, during the curricular implementation of the Heb@AR App at Bielefeld University of Applied Sciences and Arts [381]. Picture by Patrick Pollmeier, licensed under CC-BY 4.0

4.5 Development of Efficient Usability Benchmarking Utility

Already, during the formative stages of the evaluation of the early training prototypes, we quickly encountered usability challenges. Through the user-centered design process, e.g., creating personas, and the initial usability empirical evaluations of prototypes, it became clear that the specific target group in the context of midwifery education will require careful considerations for the usability of this novel technology because of comparatively low levels of media competency. Additionally, mainly because of the pandemic situation of the time, we had to rapidly iterate over prototypes and implementations, sometimes being able to actually measure the usability in curricular use for the first time. Here, the evaluation efforts from a didactic and midwifery perspective would not allow for lengthy questionnaire usage or measurement of objective data. In the end, these evaluations were the project goal, but this was standing in stark contrast to the early realization that the shared perspective of how Handheld AR could be used in the context of procedural midwifery trainings from the research proposal of the Heb@AR project, was not so shared after all, as we needed to develop entirely novel interaction concepts to properly address the learning goals. When we did start to create these missing interaction concepts, though, opportunities emerged as well. Other researchers, having similar struggles with handheld AR, showed interest in deploying our AR interaction concepts to their contexts.

Combining challenges and opportunities, we needed a tool to efficiently gather and interpret perceived usability data, ideally with a concise and focused design, little required effort, and implementation and context independence. Ideally, we needed something that allows us to quickly, but with validity, "benchmark" the usability, when comparative evaluations were not feasible.

A famous usability questionnaire, which we though might be fitting, was the System Usability Scale (SUS) by Brooke [65]. Originally envisioned and self-described as a one-dimensional "quick and dirty" approach, SUS questionnaires accounted for about 43% of post-study usability questionnaires used in the experiments identified in a meta analysis conducted by Lewis et al. [279] in 2009. Throughout the last 25 years, the initial validation of the questionnaire with $n = 20$ participants increased to $n = >10{,}000$, making the SUS a "fairly quick, but apparently not that dirty" approach as Lewis described it [276]. With recent developments towards the application of the SUS in new contexts such as elderly people or people with cognitive impairments [197] and the validation efforts of the SUS for various languages [149], this trend shows no sign of slowing down. Ultimately, the SUS has good reliability with a coefficient alpha usually around 0.92, high correlations with *likelihood to recommend* (0.75) and high correlations of *overall experience* (0.80) [28].

Besides its use in studies and specific validation endeavors of the SUS questionnaire itself, multiple researchers proposed approaches to contextualize SUS scores, trying to answer the question of what a specific SUS score actually means. As SUS scores, spanning between 0 and 100, follow neither a normal nor a uniform distribution, they cannot be interpreted linearly and especially not as a percentage value. Consequently, researchers calculated percentile curves of SUS scores from SUS study datasets, tried to contextualize SUS scores on adjectives, grading, net promoter score, quartile and acceptability scales, calculated at which point SUS scores become conclusive, and investigated the dimensionality of the SUS questionnaire by deriving learnability as a secondary dimension besides the usability of a system. All these contextualization and interpretation insights potentially add value over reporting pure SUS scores and allow for quick benchmarking.

But at the time, only a handful of mostly commercial tools existed that helped to calculate SUS scores. While these would have helped, the calculation of SUS scores itself is fairly simple and most tools do not provide further support, such as allowing researchers to compare different conditions, to plot graphs, or to contextualize the calculated results regarding the aforementioned interpretation scales. Notably, comparatively sophisticated, free toolkits do exist for competing questionnaires, like the User Experience Questionnaire (UEQ) by Laugwitz et al. [257]. But the UEQ is focused on hedonic qualities, while our needs were purely pragmatic. As we believed a tool combining such features would not only help us to efficiently ensure pragmatic usability during Project Heb@AR, but also beyond our evaluations, and usability researchers and practitioners using the SUS outside the AR context, we developed the open source web-based analysis toolkit for the SUS. This tool, visualized in Figure 4.19, can calculate SUS scores, create different SUS plots and contextualize results on the interpretation scales developed in previous works.

The tool provides an ad-hoc way to calculate all relevant SUS measurements and create clearly visualized, interactable, and customizable graphs for entire provided SUS datasets. This makes perceived usability results comprehensible across disciplines. The SUS Analysis Toolkit in its current state is fully functional, hosted on a server of the Mixed Reality Research Group at University of Applied Sciences Emden/Leer [21], and fully open-sourced on GitHub [22]. As described in detail and evaluated in a corresponding publication [50], it can be used to calculate, analyze, interpret, contextualize, and plot SUS scores from either singular, comparative, or iterative usability studies. Throughout the rest of this thesis, it is used to report all analysis, results, and figures regarding the perceived usability.

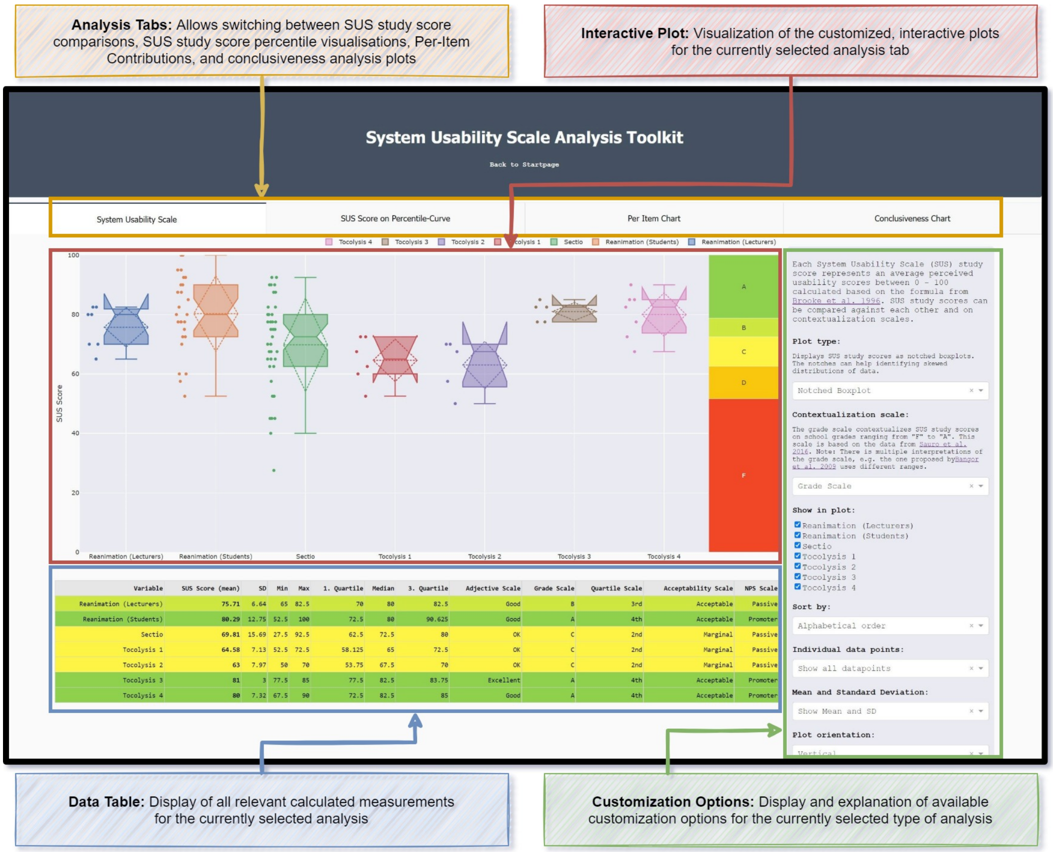


Figure 4.19: The four basic components of the System Usability Scale Analysis Toolkit User Interface: The Analysis Tabs, Interactive Plots, Data Tables, and Customization Options. In this case, the SUS Score study comparison analysis tab is selected and shows the interactive notched box plot and corresponding data table according to the chosen customization options.

[21] https://sus.mixality.de

[22] https://github.com/jblattgerste/sus-analysis-toolkit

4.6 Evaluation of the Heb@AR App

The evaluation of the Heb@AR App was conducted using an exploratory sequential mixed-methods quasi-experimental design with integrated curricular within-subject evaluations and summative cohort comparisons. Grounded in self-efficacy theory, the focus of the evaluations was on student's perceived competency and academic performance. Additionally, within-subject studies, grounded in self-determination theory, were deployed for some evaluations of the Training@Home trainings, with a focus on students' autonomy and intrinsic motivation to engage with subjects. For the purpose of this thesis, a subset of the overarching evaluations conducted by the entire Heb@AR team, is used to explore questions from the perspective of the accepted user experience principle that a useful learning construct has to be usable by the target group and provide utility to them. As conceptually visualized in Figure 4.20, the utility of one specific training can hereby be a combination of several learning benefits from either theory.

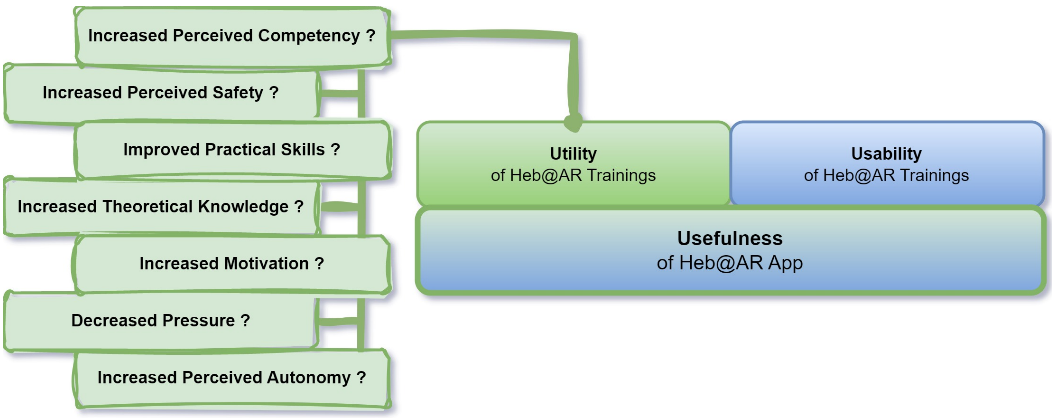


Figure 4.20: Conceptual visualization of the accepted user experience principle, that the usefulness of the Heb@AR App is dependent on its usability, but also its utility, which can be a combination of several learning benefits.

Therefore, while more granular evaluations of specific learning aspects, e.g., how well the AR trainings addressed specific learning goals from the midwifery and medical didactics perspective, which were explicitly operationalized with hypotheses, are forthcoming, the abstract questions we will exploratorily address in this section are as follows:

1. Are there indications, that the five trainings of the Heb@AR elicited learning benefits?
2. Were the Heb@AR Trainings usable by the target group of midwifery students?

To accomplish this, several questions and scales from multiple evaluations were selected post-hoc to tailor the analysis towards those specific questions. As visualized in Figure 4.21, this chapter reports evaluation results of Training 1 (see Section 4.4.3), Training 2 (see Section 4.4.4), and Training 3 (see Section 4.4.5) during curricular implementation of the App and within-subject evaluations for Training 4 (see Section 4.4.6) and 5 (see Section 4.4.7), which are independent of the curricular evaluations. Furthermore, it reports the lecturers' perspective on the curricular trainings of the Heb@AR App, which were gathered during lecturer workshops. Finally, it reports parts of the summative evaluation efforts of the curricular Heb@AR trainings in the form of objective exam comparisons to a control cohort, but also students' retrospective self-assessment of the AR intervention's benefits. As the AR intervention was implemented into two midwifery cohorts (2019-2023 and 2020-2024) at Hochschule für Gesundheit Bochum during the project's lifetime, the results are combined from evaluations in both cohorts.

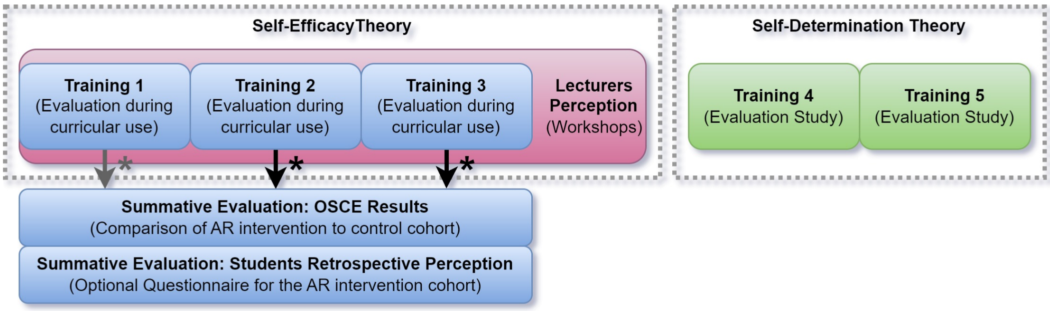


Figure 4.21: An overview of the evaluation efforts of the Heb@AR App reported in this section. self-efficacy + usability evaluations during curricular implementations for Training 1, 2, and 3 are combined with lecturer feedback, and summative evaluations. For Training 4 and 5, evaluations grounded in self-determination-theory + usability were supplemented. A timeline of the evaluations is included in the GANTT chart in Figure 4.7.
**As multiple cohorts received the AR intervention, the results of the evaluation of Training 2 and 3 are combined results for both cohorts. The results of the evaluation of Training 1 are not the cohort included in the summative evaluation.*

4.6.1 Disclaimer

The following Subsections 4.6.2 to 4.6.4 and 4.6.7 to 4.6.9 describe results from evaluations conducted during Project Heb@AR and therefore also analyze results operationalized, instrumentalized and recorded by the project partners form the Hochschule für Gesundheit Bochum and the Ruhr University Bochum. To distinguish their, in some cases unpublished and forthcoming, contributions, they will be explicitly marked as:

- *The medical didactics researchers [217]* for results which were operationalized and recorded by Matthias Joswig, Carmen Lewa, and Thorsten Schäfer
- *The midwifery researchers [35]* for results operationalized and recorded by Nicola H. Bauer, Annette Bernloehr, Kristina Vogel, and Tabea Willmeroth

The raw data from these evaluations was provided by the project partners for the purpose of this thesis. **All translation, statistical analysis, plotting and interpretations were performed by me** and do not necessarily represent interpretations or opinions of the project partners. Furthermore, scales, items, and questions were selectively chosen post-hoc for this thesis, before conducting the analysis. The project partners often asked substantially more questions or operationalized scales differently. Because of this selective variable analysis, the research questions were formulated in an exploratory manner, and hypotheses are not generated post-hoc. Non-significant or retrospectively less important insights were furthermore not removed to ensure an unbiased approach in the analysis. The evaluations for the Training 4 and Training 5 of the Heb@AR App were operationalized and evaluated independently. These two evaluations reported in Subsections 4.6.5 and 4.6.6 are therefore reported in full, including hypotheses, and represent independent contributions of this thesis.

4.6.2 Evaluation 1: Preparation of an Emergency Tocolysis

Training 1, 2, and 3 were all implemented into the midwifery curriculum of two cohorts (2019-2023 and 2020-2024). Due to the pandemic situation during the implementation of the first training, Evaluation 1, corresponding to Training 1 (The "preparation of an Emergency Tocolysis"), only reports the results from the evaluation during the curricular implementation into the second AR intervention cohort (2020-2024) in January 2022. Here, the experiment was designed as a non-controlled cohort within-subject before-and-after learning intervention study. Students first filled out a pre-study questionnaire, then completed the guidance mode of the training using the Heb@AR App, and afterward filled out a post-study questionnaire. During this curricular implementation, 33 students took part in the experiment.

Results: Self-Reported Perceived Competency

To measure, if the AR training influenced the perceived competency of students, *the medical didactics researchers [217]* asked students, "[Before/After] the AR training, how would you rate your competence in correctly performing the preparation of an emergency tocolysis on a scale of 0 to 100?" in the pre-study questionnaire before they started the AR training, and in the post-study questionnaire after completing the training.

Students reported their perceived competency as on average 44.03 (Mdn = 50, SD = 25.01, Skewness = 0.13) before the AR training and an average of 81.24 (Mdn = 85, SD = 20.57, Skewness = -2.25) after completing the AR training for the preparing the tocolytic injection (see Figure 4.22).

A Shapiro-Wilk test indicated that the residuals are normally distributed (W = 0.96, p = 0.34). Tukey Fence (k = 1.5) indicates that there are no outliers in the data. Therefore, a parametric test was used. A paired t-test indicated that there is a statistically significant large difference between their self-reported perceived competency before and after completing the AR training ($t(28) = 7$, $p < 0.001$, Cohen's d = 1.3).

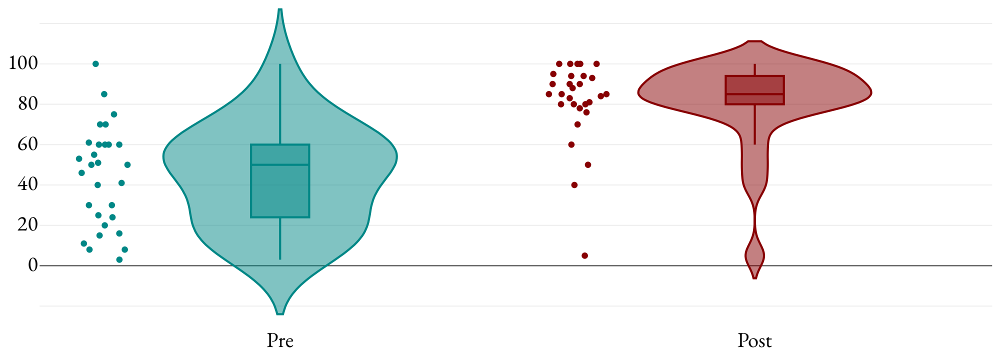


Figure 4.22: The reported perceived competency to prepare a tocolytic injection on a scale from 0 to 100, before (pre) and after (post) completing the Training 1 of the Heb@AR app during curricular implementation.

Results: Perceived Usability

To record the students' perceived usability of the training from a pragmatic perspective, we furthermore included the SUS in the post-study questionnaire. The SUS version used during the study consisted of 10 questions: Question 1: "*I think that I would like to use this product frequently.*", Question 2: "*I found the product unnecessarily complex.*", Question 3: "*I thought this product was easy to use.*", Question 4: "*I think that I would need the support of a technical person to be able to use this product.*", Question 5: "*I found the various functions in this product were well integrated.*", Question 6: "*I thought there was too much inconsistency in this product.*", Question 7: "*I would imagine that most people would learn to use this product very quickly.*", Question 8: "*I found this product very awkward to use.*", Question 9: "*I felt very confident using this product*", Question 10: "*I needed to learn a lot of things before I could get going with this product.*". A validated German translation of this revised version [149] of the SUS was used for the experiment.

Overall, a SUS study score of 83.11 (SD = 12.9) was reported. This SUS study score would correlate with the adjective usability description of "Excellent" according to Bangor et al. [27], graded an "A" [416] and would be classified as an acceptable usability [28]. Furthermore, this SUS study score surpasses the non-empirical but commonly used industry benchmark, to surpass SUS study scores of 80 [278]. With a sample size of $n = 33$, this result should be conclusive according to Tullis et al. [463]. The results are visualized in Figure 4.23. The Appendix Table 29 includes the complete analysis from the SUS Analysis Toolkit [50].

4.6.3 Evaluation 2: Permanent Catheter Placement

The evaluation of Training 2, in line with the evaluation of Training 1, was designed as a non-controlled cohort within-subject before-and-after learning intervention study. The Competency results hereby combine results from evaluations in both cohorts (2019-2023 and 2020-2024) combined, which were conducted in December 2021 and July 2022 respectively, and 39 and 25

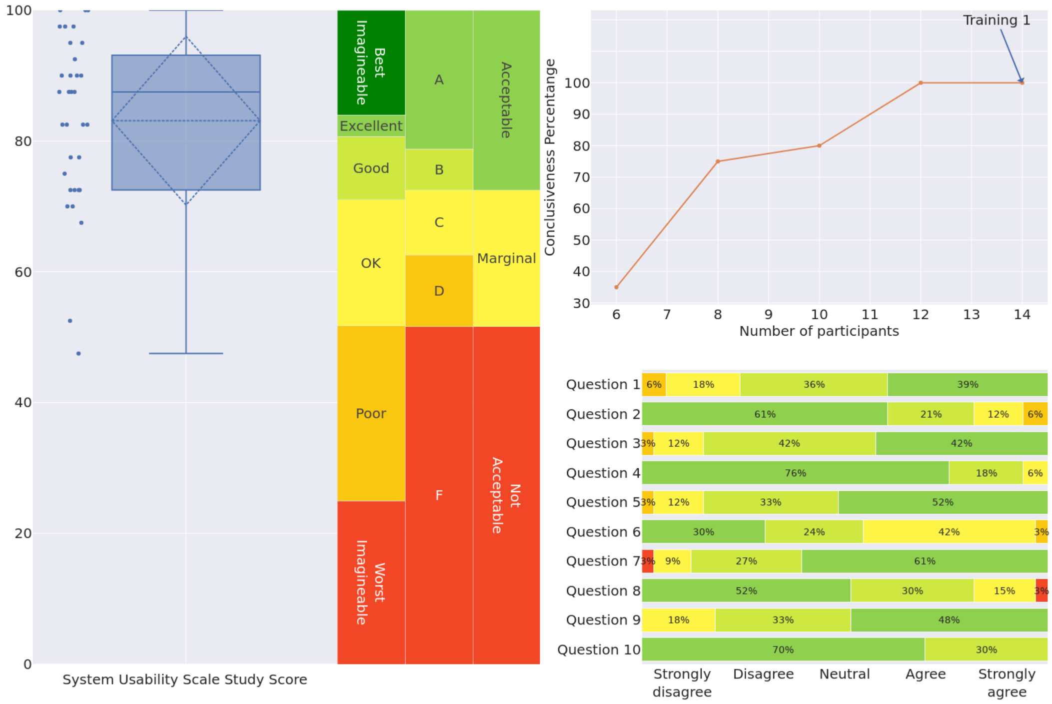


Figure 4.23: The individual SUS scores of Training 1 as data points, its SUS study score as a box plot and the SUS contextualization scales: “Adjective Scale” [28], “Grade Scale” [416] and “Acceptability Scale” [27] for their interpretation (left). The conclusiveness percentage of the SUS study score based on the number of participants [463] (upper right) and the response distribution on the Likert scales of the ten questions of the SUS as stacked bar chart [50] (lower right).

students took part in the experiment during the curricular implementations of the same training. The usability results were only recorded for the first cohort.

Results: Self-Reported Perceived Competency

To measure, if the training influenced the self-reported perceived competency of students, *the medical didactics researchers [217]* asked students, “[Before/After] the AR training, how would you rate your competence in correctly performing the placement of a permanent transurethral bladder catheter on a scale of 0 to 100?” before starting the AR training and after completing it.

As shown in Figure 4.24, students reported an average competency score of 53.27 (SD = 28.11, Skewness = -0.46) before starting the training and an average competency score of 71.5 (SD = 20.03, Skewness = -1.24) after completing the AR training.

A significant Shapiro-Wilk test indicated that the residuals were not normally distributed (W = 0.92, $p < 0.001$). As the assumption of normality was violated, a Wilcoxon-Test was used. Tukey

Fence (k = 1.5), indicated that the data contains 3 outliers, accounting for 4.69% of the observation. The Wilcoxon-Test indicated that the perceived competency before the training (Mdn = 58) was significantly lower than the perceived competency after the training (Mdn = 75). Here, the standardized effect size was large (0.73), indicating a substantial difference between before and after the training and the common language effect size was 0.081. This suggests that there is an 8.1% probability that a random value from before the training is greater than its corresponding value from after completing it.

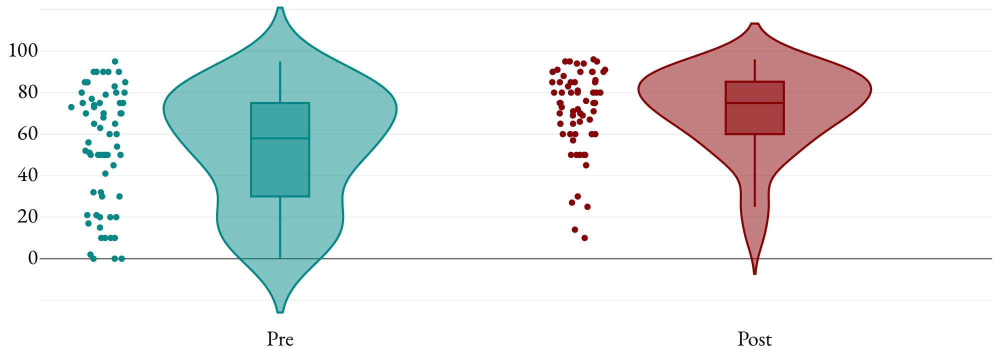


Figure 4.24: The reported perceived competency to place a permanent catheter in preparation for a c-section, on a scale from 0 to 100, before (pre) and after (post) completing the Training 2 of the Heb@AR app during curricular implementation.

Notably, when exploring cohort differences using Mann-Whitney U-Tests, the residuals between groups show a small statistically significant difference (U = 254, p = 0.02, r = 0.31) but the differences in the post measures are not statistically significant (U = 287, p = 0.071, r = 0.24).

Results: Perceived Usability

In terms of perceived usability, the reported SUS study score of the first cohort was 69.81 (SD = 15.69). This would be classified as "OK" usability according to Bangor et al. [27], and while it is still an above-average SUS study score, it would not surpass the commonly used non-empirical benchmark ok 80 [278]. The SUS study score of 69.81 would only indicate "marginally acceptable" [28] usability. With a sample size of n = 39 participants, this result should be 100% conclusive, according to Tullis et al. [463]. The SUS results are visualized in Figure 4.25. The Appendix Table 29 includes the complete analysis from the SUS Analysis Toolkit [50].

4.6.4 Evaluation 3: Reanimation of a Newbown

The curricular evaluations of Training 3, in line with the other evaluations, were also designed as non-controlled cohort within-subject before-and-after learning intervention studies, which were conducted in September 2021 and July 2022. For the first cohort, the single-user version of Training 3 was evaluated, and 26 participants took part in the experiment. For the second cohort, the

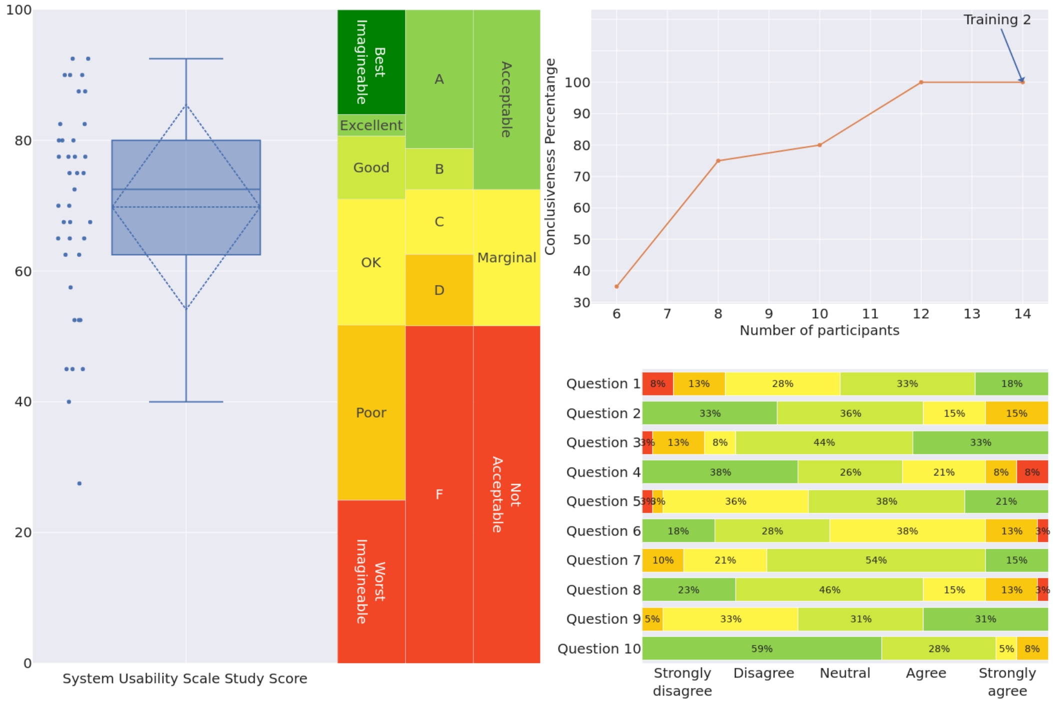


Figure 4.25: The individual SUS scores of Training 2 as data points, its SUS study score as a box plot and the SUS contextualization scales: "Adjective Scale" [28], "Grade Scale" [416] and "Acceptability Scale" [27] for their interpretation (left). The conclusiveness percentage of the SUS study score based on the number of participants [463] (upper right) and the response distribution on the Likert scales of the ten questions of the SUS as stacked bar chart [50] (lower right).

multi-user version was used, and 33 participants took part in the experiment. Results regarding potential influences on the competency were recorded from the first cohort, usability results were recorded for both cohorts to have data on both versions of Training 3 of the Heb@AR App.

Results: Self-Assessed Influence on Competency

While *the medical didactics researchers [217]* did not ask the pre- / post-questions about the students' perceived competency in line with the evaluations for the two previous trainings, *the midwifery researchers [35]* did operationalize learning-goal specific competency pre- / post-scales, which we will publish in a forthcoming publication. For the purpose of this thesis, to at least report tendencies of the AR training's influence on the students' perceived competencies, we can combine self-assessments of the students, recorded after completing Training 3 during its curricular implementation, with qualitative feedback reported by that cohort in their "logbooks", similar to learning diaries, after completing their practical phase.

As reported in our publication on the reanimation of a newborn [293], after completing the AR training during curricular implementation, the students strongly agreed with both statements, *"The AR training reanimation of a newborn contributes to the advancement of my professional competencies"* (m = 4.46, SD = 0.58, n = 26) and *"I was able to further my knowledge [about the reanimation of a newborn]"* (m = 4.46, SD = 0.71, n = 26) in their post-study questionnaires.

In line with these results, when *the midwifery researchers [35]* asked the students the qualitative question *"If [you participated in the AR training for the reanimation of a newborn], did the AR training make you feel well-prepared for the situation [in your practical phase]?"* in their learning diary, that they had to fill out after completing their practical study phases, the n = 31 students who did complete this AR training at some point (not necessarily during curricular implementation, as there were additional optional AR training opportunities), stated feedback indicating the following: Most of the students provided feedback indicating that they felt well-prepared by the AR training (19 students). For their reasons, they stated that it helped to clarify, structure, sort, and deepen the necessary action steps of the reanimation (8 students) and that the AR training supplemented their previous knowledge well (3 students). 4 Students provided feedback indicating that they felt somewhat, or "in principle", well-prepared by the AR training. As reasons, they state that they would have liked to repeat the reanimation procedure more often with support from a lecturer in the SkillsLab sessions (1 student) or at home in their self-study time (1 student). Of the remaining students that somewhat agreed, 1 student stated that they simply think, the training dummies are not comparable to a real reanimation. Finally, 7 students indicated that they did not feel well-prepared by the AR training or cannot answer this question because they did not encounter the reanimation of a newborn in their practical study phase.

Results: Perceived Usability

The first cohort, where the single-user version of Training 3 was implemented, reported an SUS study score of 80.29 (SD = 12.75). This would correlate with the adjective description of "good" usability according to Bangor et al. [27], graded as an "A" [416] and indicates acceptable usability [28]. Being above 80, this SUS study score is furthermore surpassing the non-empirical but commonly used industry benchmark for the SUS [278]. With a sample size of n = 26, this result should be conclusive according to Tullis et al. [463].

On the other hand, the cohort, where the multi-user version was implemented, only reported an SUS study score of 60.08 (SD = 14.33). This is a below-average SUS study score, which is around 68 [278] to 70 [27], depending on the source, and would indicate only marginally acceptable usability according to Bangor et al. [28], while being graded a "D" [416]. Described using adjectives, this would only indicate "OK" usability [27]. With a sample size of n = 33 participants, this result should be conclusive based on the sample size as well [463].

As visualized in Figure 4.27, the results of the descriptive statistics indicate that the users of Training 3 in the single-user variant reported higher SUS scores (M = 80.29, SD = 13.01) compared to the users of the multi-user version of Training 3 (M = 60.08, SD = 14.56). A Shapiro-Wilk test indicates that the residuals are normally distributed (p = 0.31), and Levenes test indicates equality

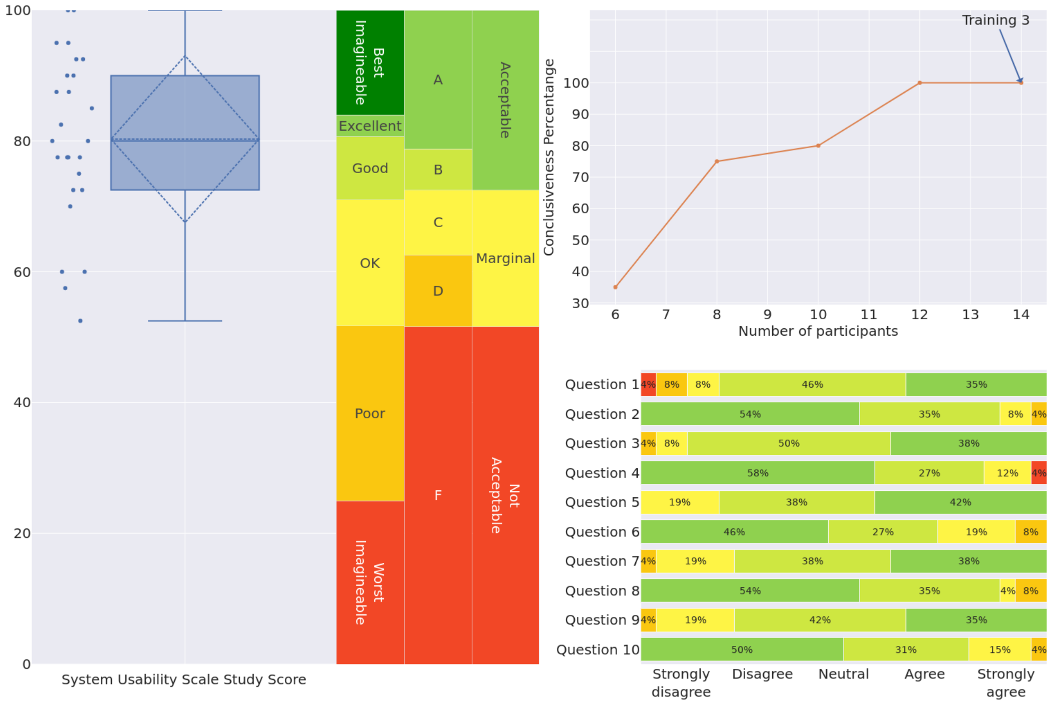


Figure 4.26: The individual SUS scores of Training 3 as data points, its SUS study score as a box plot and the SUS contextualization scales: "Adjective Scale" [28], "Grade Scale" [416] and "Acceptability Scale" [27] for their interpretation (left). The conclusiveness percentage of the SUS study score based on the number of participants [463] (upper right) and the response distribution on the Likert scales of the ten questions of the SUS as stacked bar chart [50] (lower right).

of variance. A two-tailed t-test for independent samples shows that the difference between the single-user and multi-user Training variants with respect to the perceived usability was statistically significant, $t(57) = 5.55$, $p = < 0.001$. The Appendix 29 includes the complete results for both versions from the SUS Analysis Toolkit [50].

To investigate where this significant difference in usability between the two versions originates from, we decided to review a subset of the qualitative feedback provided by the participants, regarding technical problems. As *the midwifery researchers [35]* asked the participants qualitative questions like *"The following difficulties were encountered in using the AR exercise:"* and *"Do you have a suggestion to improve the AR application, "resuscitation of a newborn"?"* after completing the AR training during the curricular implementation, we can combine the results to explore the potential reasons. Roughly combining the impressions of the qualitative feedback with non-representative observations by the experimenters, most students from the cohort that used the single-user version focused their feedback on specific aspects of the training, e.g., suggesting textual or color changes or which midwifery-specific aspects they would have liked to be explained

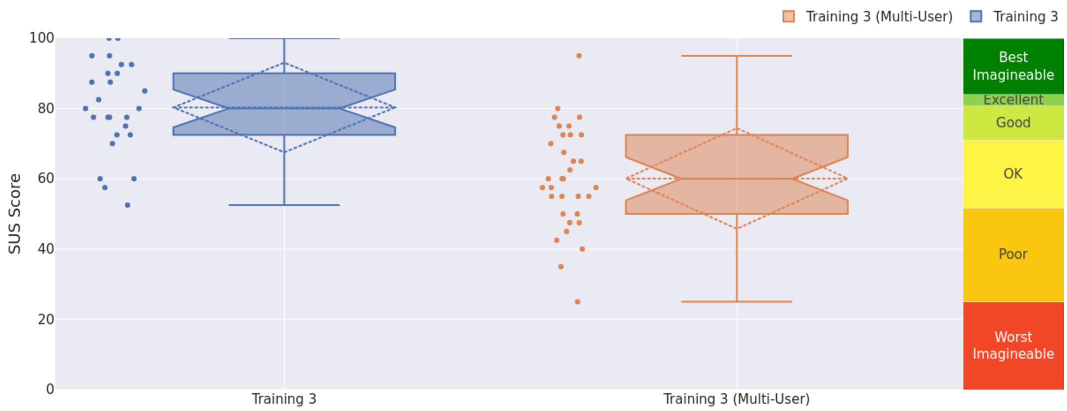


Figure 4.27: The individual SUS scores and SOS study scores as box plots for Training 3 in the single- and multi-user variants. Contextualized on the Adjective Scale by Bangor et al. [27].

in more detail. Nonetheless, some did report crashes of the app. Students from the cohort that used the multi-user application, on the other hand, reported a substantial number of problems. Out of the 35 participants who provided feedback to this question, 24 reported having technical problems and most reported that these problems were in relation to the “connectivity” of the smartphones during the multi-user training, which, according to them, sometimes even resulted in the rest of the group not being able to complete the training. Smartphones that were disconnected, were unable to rejoin the training, according to many accounts, likely also influencing the perceived usability of participants that were not directly effected. Likely because of this, even when excluding the SUS results from students that reported technical difficulties, the SUS study score only slightly improves (m = 65.45, Mdn = 65, SD = 13.64), which would still only indicate a marginally acceptable [28], “OK” [27] perceived usability.

4.6.5 Evaluation 5: The Female Pelvis

The evaluation of Training 5, learning the bones of the female pelvis, was designed as an optional, non-controlled cohort within-subject before-and-after learning intervention with the perspective of “Usability + Utility = Usefulness”, where the utility aspect was grounded in Self-Determination Theory. It was conducted during a practical “SkillsLab” tutorial session by Prof. Annette Bernloehr in the academic midwifery bachelor study program of the Bielefeld University of Applied Sciences and Arts in June 2022. The evaluation of Training 5 is reported before the evaluation of Training 4, as it was performed beforehand, has the larger sample size, and is used to report the internal consistency of the operationalization. It is still described as “Evaluation 5” for clarity, to be consistent with the numbering of the training it evaluates.

While the previous trainings were evaluated in the context of the self-efficacy theory, focused on students’ perceived competency, Training 4 and 5 were evaluated contextualized in self-determination theory. Here, while perceived competency and perceived pressure are influencing factors, the focus is on the students' intrinsic motivation to engage with a specific subject. Therefore, a non-validated German translation of the validated Intrinsic Motivation Inventory (IMI) [406], a validated German translation of the System Usability Scale (SUS) by Gao et al. [149] and qualitative feedback questionnaires were used in the pre- and post-study questionnaires to answer the following research questions:

1. Does the Pelvis Termini training increase the intrinsic motivation of students to engage in the anatomical terminology topic, which is historically perceived as boring, compared to their previous experiences with a conventional learning method?
2. Is the Pelvis Termini Training usable by midwifery students?
3. Do students perceive the training as a valuable addition to existing learning methods? Could it potentially even replace the conventional methods?

Hypothesis

Our hypothesis was that the interactive properties of the Pelvis AR training increase students' intrinsic motivation to engage with the topic of anatomical terminology compared to conventional methods significantly. To be more specific, this hypothesis can be split into the following three hypotheses: H1.1: the interactive Pelvis AR training increases perceived competence among students compared to previous experiences in a traditional memorizing exercise. H1.2: the interactive Pelvis AR training reduces the perceived pressure in studying content compared to previous experiences of the students. H1.3: the Pelvis AR training increases the intrinsic motivation of students compared to their previous experiences. Furthermore, as similar usability evaluations for other applications using the TrainAR framework had promising usability evaluations [18, 55, 114], we expected the usability of the interactive Pelvis AR training to be excellent for the target group and to not influence the motivational effects negatively by complicating the interaction unnecessarily (H2). When it comes to the students' acceptance, we expected them to perceive interactive handheld AR trainings as a useful optional addition to existing learning methods, but indicate skepticism regarding it potentially replacing conventional methods outright (H3).

Procedure

In a second semester practical midwifery lecture, in which students engaged with a handheld Augmented Reality procedural training for the first time, they were offered to participate in this study after completing the lecture's obligatory learning content. After scanning a QR code, participants first completed a pre-study survey on their smartphone. Here, they were asked for their consent, a demographic questionnaire, their experience with Latin and AR technology and the IMI [406] in relation to their conventional learning approach for anatomical terminology learning. Subsequently, students completed the AR Pelvis training, either through their own smartphones by downloading and then completing the training, or by using institutional smartphones that were made available to them. Finally, participants were asked to complete the post-study survey, where they answered the System Usability Scale, the IMI on the Pelvis AR training, and qualitative questions on the perceived usefulness of the training. Participants completed the training independently and were not helped during the training. An experimenter was available in case of technical difficulties, and a midwifery professor was available in case of subject-related questions. Due to time and space constraints during the lecture, some participants had to complete the Pelvis AR training before filling out the pre-study survey.

Participants

The experiment was carried out with 36 primary qualifying midwifery bachelor students aged between 18 and 40, with an average age of 21.81 (SD = 4.52). All participants were female. Out of the 36 participants, 8 had the advanced Latin certificate (German: "Großes Latinum"), 4 had the intermediate Latin certificate (German: "Kleines Latinum") and the 24 remaining participants had no formal Latin certification. When asked how much experience they had with Augmented Reality, 33 participants reported having no experience with the technology, and 3 participants reported having very little experience.

While participation in the study was optional, it was attached to a practical "SkillsLab" training session of their curriculum, and students were not compensated for their participation. All participants in the study successfully completed the Pelvis AR training. 33 participants completed it on Apple iPhones, ranging from the iPhone 8, over the iPhone SE to the iPhone 12 Pro. Of the remaining participants, 2 used Android smartphones and 1 participant used an Android tablet to participate in the study.

Results: Intrinsic Motivation

To measure the motivation of students, the three subscales interest/enjoyment (the self-reported measure of intrinsic motivation, and therefore referred to as "intrinsic motivation" from here on), perceived competence, and pressure/tension of the IMI [406] were administered before and after the intervention. Hereby, 7-Point Likert Scales ranging from "strongly disagree" to "strongly agree" were utilized. For all three subscales, the Cronbach's alpha indicated sufficient internal consistency, ranging from $\alpha = 0.77$ to $\alpha = 0.92$. Additionally, for checking the differences between the pre- and post-measure, the assumption for normality was satisfied and therefore three paired t-tests were conducted. To control for Type I error based on multiple comparisons, a p-value based on the Bonferroni correction ($p = 0.017$) was applied.

The results of the subscales are visualized as Box plots in Figure 4.28 and the results of the t-tests are displayed in Table 4.2. They show significant changes for all three subscales between the pre- and post-measures. Perceived competence increased significantly after completing the Pelvis AR training compared to the students' previous experiences with traditional memorizing approaches $t(31) = -4.56$, $p = <0.001$. Moreover, also the perceived pressure/tension increased significantly after completing the Pelvis AR training $t(31) = -2.64$, $p = 0.01$. Finally, the result of the remaining paired t-test indicates that students' intrinsic motivation increased after the intervention (M =

5.50, SD = 0.63) compared to before the intervention (M = 4.60, SD = 1.22). This difference was highly significant t(31) = -3.94, p = <0.001. Both perceived competence (d = 0.81) and intrinsic motivation (d = 0.70) represented strong effect sizes, whereas the differences in pressure/tension represented a medium effect size of d = 0.47.

Bonferroni corrected (p = 0.017) Pearson Correlations showed a significant high positive correlation between pre- and post-measures for the perceived competence (r(34) = 0.58, p = <.001). No significant correlations between pre- and post-measures were found for pressure/tension (r(34) = 0.3, p = .072) or intrinsic motivation (r(34) = 0.19, p = 0.275).

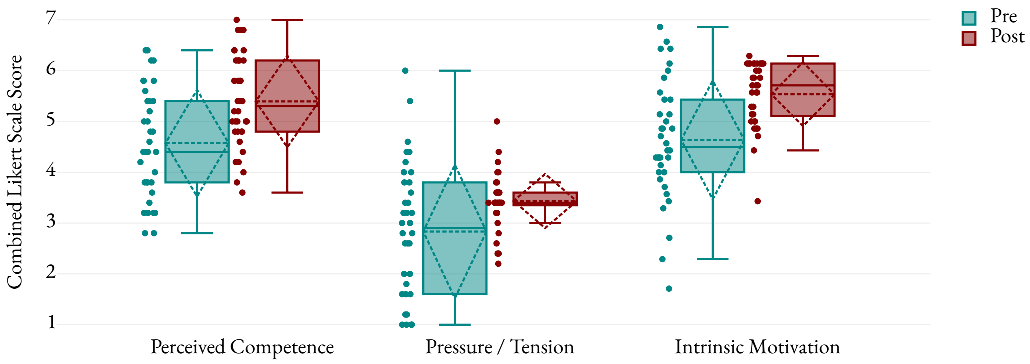


Figure 4.28: Measures of the motivational variables of the perceived competence, pressure/tension and intrinsic motivation of the IMI [406] before and after completing Training 5 (Female Pelvis).

Parameter	**Pre** M	SD	**Post** M	SD	**t(31)**	**p**	**d**
Perceived Competence	4.61	1.09	5.34	0.89	-4.56	<0.001	0.81
Pressure / Tension	2.83	1.35	3.43	0.57	-2.64	0.01	0.47
Intrinsic Motivation	4.60	1.22	5.50	0.63	-3.94	<0.001	0.67

Table 4.2: Results of the three paired t-test examining the change of the motivational variables of the perceived competence, pressure/tension and intrinsic motivation of the IMI [406] before and after completing the Pelvis AR training.

Results: Perceived Usability

Calculating the results of the SUS [28] using the System Usability Scale Analysis Toolkit [50], revealed a SUS study score of 84.79 (SD = 13.51) with a minimum score of 55, a maximum score of 100 and a median score of 90. This SUS study score is considered an "Acceptable" usability according to Bangor et al. (2008) [28], would be graded an "A" on the empirical grading scale by Sauro et al. [416] and categorized as "Best Imaginable" usability when described using adjectives, according to Bangor et al. (2009) [27] (see Fig. 4.29, left). With a sample size of 36 participants, this result is 100% conclusive according to Tullis et al. [463] (see Fig. 4.29, upper right). Furthermore, when normalizing the 10 individual question scores to their average contribution towards the SUS study score according to Blattgerste et al. [50], there were no distinctive insights or deviations stemming from individual questions (see Fig. 4.29, lower right). The complete SUS analysis is included in the Appendix Table 30.

A one-way ANOVA revealed no statistical differences in perceived usability in form of SUS scores as a result of pre-existing Latin certification, F (2, 33) = 1.915, p = 0.163.

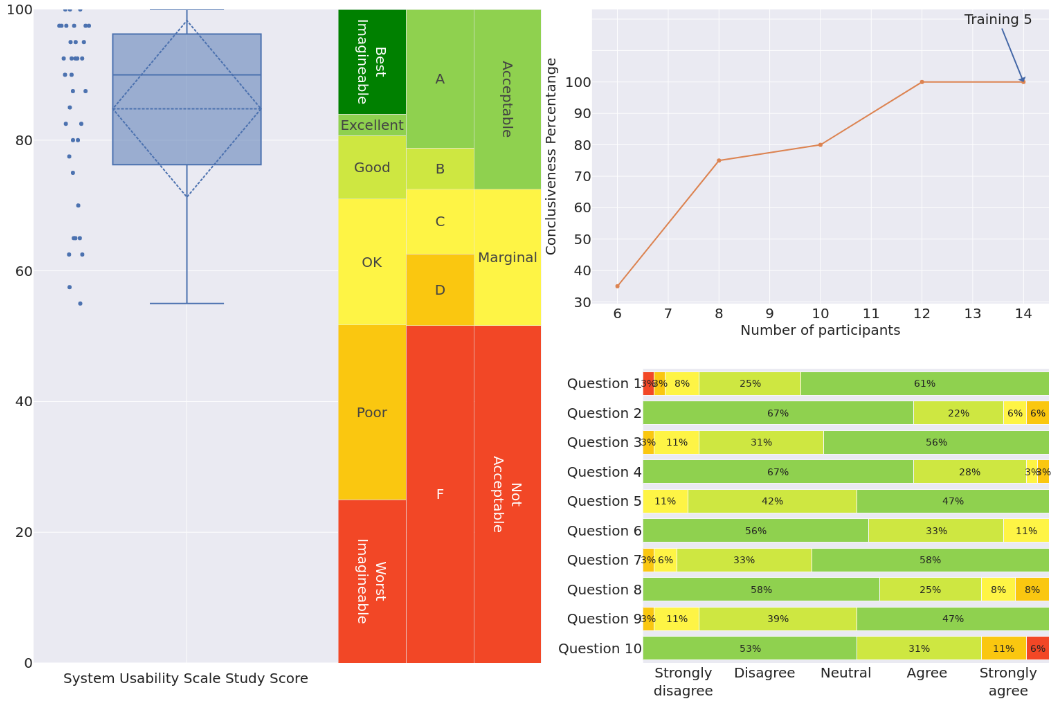


Figure 4.29: The individual SUS scores of Training 5 as data points, its SUS study score as a box plot and the SUS contextualization scales: "Adjective Scale" [28], "Grade Scale" [416] and "Acceptability Scale" [27] for their interpretation (left). The conclusiveness percentage of the SUS study score based on the number of participants [463] (upper right) and the response distribution on the Likert scales of the ten questions of the SUS as stacked bar chart [50] (lower right).

Results: Perceived Usefulness

Finally, to gather additional feedback on the perceived usefulness of the Pelvis AR training, we asked participants how much they agree with two statements on 7-Point Likert Scales and provide the reasoning for their answer.

When asked how much the participants agreed with the statement *"I think the Pelvis AR training would help me learn the pelvic bones and pelvic spaces."*, an average Likert scale value of 6.389 (SD = 1.231) was reported. This would indicate that the participants strongly agree with this statement. When asked for their reasoning, 14 participants provided no answer. Out of the remaining 22 participants, 8 indicated that they especially liked the 3D visualization as it would be particularly helpful for them to "envision the pelvis model spatially, as opposed to on paper". Additionally, 8 participants expressed their agreement with the concept of "active" learning in terms of naming and memorizing the parts and repeated correction in case of errors. One participant stated that the training would "make learning easier" and "have a long-term effect". Furthermore, three participants explained the perceived benefits mainly in connection with the fun and "playful" aspects of the training, which they expect to lead to "higher motivation" and the possibility of "using the training at home". Two participants stated that they had already acquired the knowledge through the conventional learning methods "books and index cards" or by using a "pelvis model" and the training would therefore not be helpful. Another participant stated that "the app is only useful with previous knowledge".

Subsequently, we asked how much participants agreed with the statement *"I would rather use the Pelvis AR training instead of learning the pelvic bones and pelvic spaces the conventional way."* and reported a Likert scale value of 5.222 (SD = 1.669), which would indicate that they agree with the statement. Sixteen participants provided no reasoning for their answer. Of the remaining participants, six noted that the advantage of training stems from the combination of realistic visualizations, which make it easier to "mentally associate and remember terms" compared to conventional learning methods. In this context, one of them pointed out that it is "less theoretical" (likely meaning less "dry" as a subject), which would "increase motivation" for her. Another participant mentioned that the advantage of this learning method is that it is "more convenient" when learning on the go. Two participants explained their reasoning for preferring the Pelvis AR training, with "it is more fun" and "exciting learning". The remaining participants were more critical. Two of them stated that they believe traditional learning methods to be "just as effective". In line with this perspective, five participants see "mixing" their traditional learning with the Pelvis AR training as promising so that the scenario rather serves "as a supplement" or complementary to "consolidate knowledge". One of the participants stated that although this learning method would not help her "learn the Latin terms", it would still help her in learning to contextualize the terminology onto the correct bones-structures and areas. Finally, two participants stated that they would prefer traditional learning methods because they did not appreciate "technical learning methods" and there is no possibility to "physically interact" with the Pelvis model.

Finally, when asked if participants had further feedback, notes or suggestions, three participants emphasized the meaningfulness of the "training" and thanked the developers. One participant noted that she had a lot of fun during the training.

4.6.6 Evaluation 4: Virtual Reanimation of a Newborn

The evaluation of Training 4, the virtual version of the reanimation of a newborn, was designed as an optional, non-controlled cohort within-subject before-and-after learning intervention, in line with the evaluation reported in Section 4.6.5. In this, it also uses the same instruments and asked the same research questions as the evaluation of Training 5. It was conducted before a practical "SkillsLab" session for exam preparation in the academic midwifery bachelor study program of the Hochschule für Gesundheit Bochum in January 2023.

Hypothesis

Our hypothesis regarding the intrinsic motivation, usability and perceived usefulness, are largely in line with the hypothesis stated in Section 4.6.5 (H1.1, H1.3, H2, H3 are equal). Only our hypothesis for the sub-scale pressure/tension of the IMI [406] (H1.2) is adjusted to the observations from the evaluation of Training 5. Therefore, we expect the perceived pressure to increase, compared to previous experiences of the students.

Procedure

After scanning a QR code, participants first completed the pre-study questionnaire on their smartphone. Here, they were asked for their consent, a demographic questionnaire, how they would normally learn the reanimation procedure, and the IMI [406] in relation to their conventional learning approach. Subsequently, students completed the AR training, either through their own smartphones by downloading and then completing the training, or by using institutional smartphones that were made available to them. After the training was concluded, participants were asked to complete a post-study questionnaire, where they answered the System Usability Scale, the IMI, and qualitative questions on the perceived usefulness of the AR training. Participants completed the training independently and were not helped during the experiment. Two experimenters were available in case of technical difficulties.

Participants

The experiment was carried out by 10 participants with an average age of 21.11 (SD = 0.78), 8 were female, 1 identified as diverse, and 1 did not provide an answer to the question. While 9 participants filled out all questionnaires, 1 participant was excluded for the IMI [406] as data points were missing for the pre-study questionnaire, but is included in the SUS results [65]. The participation in the study was optional, students had previous contact with the AR through the trainings 1, 2 and 3, and students were not compensated for their participation. When asked what method, they would normally use to consolidate the knowledge of the reanimation procedure of a newborn, four participants stated they normally use "Learner's note", four stated that they would learn based on the lecturers notes, and 2 stated that they would use the SkillsLab.

Results: Intrinsic Motivation

The results of the subscales are visualized as Box plots in Figure 4.30 and the results of the t-tests are displayed in Table 4.3. According to Shapiro-Wilk tests, the residuals of the subscales perceived competency (p = 0.809), pressure/tension (p = 0.08), and intrinsic motivation (p = 0.509) all follow a normal distribution. Therefore, for comparability to the results from Evaluation 5 in Section 4.6.5, Bonferroni corrected (p = 0.017) parametric tests are used, despite the sample size.

According to the data, the perceived competence of students decreased (M = 4.13, SD = 0.87) after completing the Virtual Reanimation AR training compared to the students' previous experiences (M = 4.60, SD = 1.06) with their traditional learning method, but this difference was not statistically significant (t(8) = 0.8 p = 0.447). Furthermore, the perceived pressure/tension decreased (M = 2.84, SD = 0.85) after completing the AR training (M = 3.18, SD = 1.12). This difference was also not statistically significant (t(8) = 1.08 p = 0.312). Finally, the result of the paired t-test for students' intrinsic motivation indicates that it increased after the intervention (M = 5.38, SD = 0.78) compared to before the intervention (M = 4.05, SD = 0.66). This difference was statistically significant, t(8) = -3.45 p = 0.009.

The non-significant decrease in perceived competence (d = 0.15) represents a small effect size, the non-significant decrease in pressure/tension represented a medium effect size of d = 0.4. The statistically significant increase in intrinsic motivation (d = 0.72) represented a strong effect size.

Bonferroni-corrected (p = 0.017) Pearson Correlations showed a non-significant high negative correlation between pre- and post-measures for the perceived competence subscale (r(7) = -0.65, p = 0.058). Furthermore, non-significant positive high correlations were found for pressure/tension (r(7) = 0.59, p = 0.094.). Finally, the pre- and post-measures for the intrinsic motivation subscale showed a non-significant small negative correlation (r(7) = -0.29, p = 0.447).

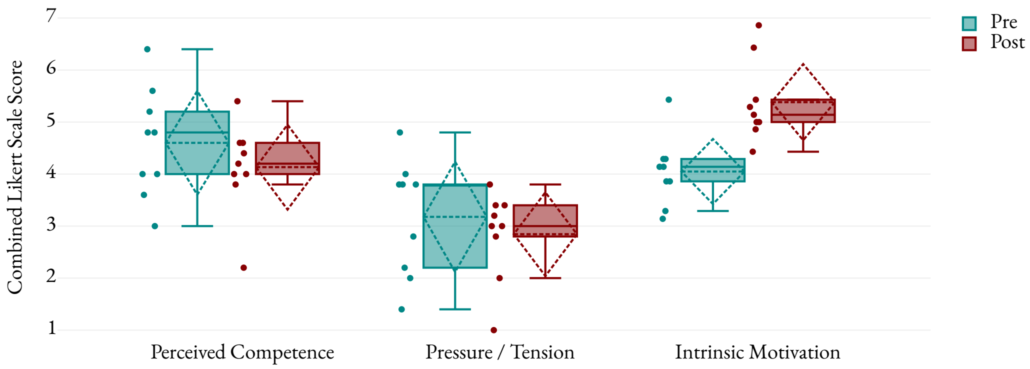


Figure 4.30: Measures of the motivational variables of the perceived competence, pressure/tension and intrinsic motivation of the IMI [406] before and after completing Training 4, the Virtual Reanimation, using a reanimation unit.

Parameter	Pre		Post		t(8)	p	d
	M	SD	M	SD			
Perceived Competence	4.60	1.06	4.13	0.87	0.8	0.447	0.15
Pressure / Tension	3.18	1.12	2.84	0.85	1.08	0.312	0.40
Intrinsic Motivation	4.05	0.66	5.38	0.78	-3.54	0.009	0.72

Table 4.3: Results of the paired t-tests examining the change of the motivational variables (Competence, pressure, and intrinsic motivation) of the IMI [406] before and after completing Training 4.

Results: Perceived Usability

Calculating the results of the SUS [28] using the System Usability Scale Analysis Toolkit [50], revealed a SUS study score of 73 (SD = 15.6) with a minimum score of 40, a maximum score of 92.5 and a median score of 71.25. This SUS study score is considered an "Acceptable" usability according to Bangor et al. (2008) [28], would be graded an "B" on the empirical grading scale by Sauro et al. [416] and categorized as "Good" usability when described using adjectives, according to Bangor et al. (2009) [27] (see Fig. 4.31, left). With a sample size of 10 participants, this result should be roughly 80% conclusive according to Tullis et al. [463] (see Fig. 4.31, upper right). The Appendix 30 includes the complete SUS analysis from the SUS Analysis Toolkit in the form of tables [50].

While it is generally not advised to interpret individual questions of the SUS independently as they are not diagnostic [66], it should be noted that manual inspection of the Likert scale values of Question 9: *"I felt very confident using this product"* (see Figure 4.31), suggests inconsistencies in comparison to the other items. Analyzing the average Likert scale value of 3.1 (SD = 0.99) for this item, it also doesn't surpass the benchmark calculated through the linear regression function $Item\_9\_Benchmark = 0.6992487 + 0.04435754 * SUS\_Study\_Score$ [R^2=0.85] suggested by Lewis et al. [278], which for an SUS study score 73 would suggest a benchmark of 3.93. This might have been caused by students rating their confidence with the contents of the AR training, rather than the confidence regarding the applications' usability, especially when viewed with the surprising decrease in perceived competency measured by the IMI [406]. Ultimately, this was not observed in any other of our evaluations and is therefore only noted here; the item is not excluded from the calculation, as suggested by Lewis et al. [277].

Results: Perceived Usefulness

To gather feedback on the perceived usefulness of AR Training 4 "Resuscitation of a newborn: Training case resuscitation unit", we asked participants how much they agree with the statements *"I think this AR training is an effective follow-up for the AR training (Resuscitation of a Newborn), which I previously trained in the SkillsLab."* and *"I think this purely virtual AR training (Resuscitation of a newborn: Training case resuscitation unit) could completely replace the SkillsLab AR training resuscitation of a newborn, which has to be trained on site in the SkillsLab."* on 7-point Likert scales. A Likert scale value of 5.77 (SD = 1.47), indicates that they agree with the first state-

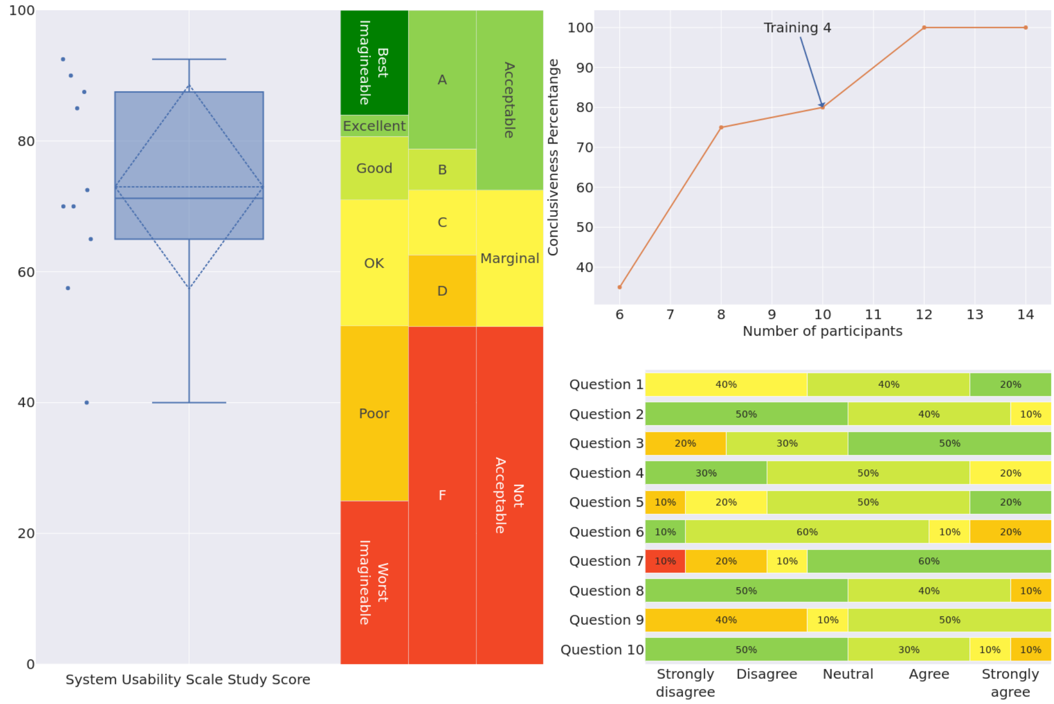


Figure 4.31: The individual SUS scores of Training 4 as data points, its SUS study score as a box plot and the SUS contextualization scales: "Adjective Scale" [28], "Grade Scale" [416] and "Acceptability Scale" [27] for their interpretation (left). The conclusiveness percentage of the SUS study score based on the number of participants [463] (upper right) and the response distribution on the Likert scales of the ten questions of the SUS as stacked bar chart [50] (lower right).

ment, but a Likert scale value of 1.33 (SD = 0.71) indicates they strongly disagree with the second statement. When asked for the reasoning behind their decision, six participants stated that the "handgrips" (Note: as in the motor components) are missing, and that the "handling of the newborn" is important. Two participants stated that they felt like it was a good opportunity to "repeat the whole procedure at home, without pressure" but that the "atmosphere"/"realism of the situation" was missing.

When asked how much they agreed with the statement, *"I would rather use this AR training than rehearsing the same content with my conventional method."* on a 7-point Likert scale, an average Likert scale value of 4.45 (SD = 2.13) was reported, indicating that they somewhat agreed with that statement. When asked for the reason, four participants stated that they would rather "combine approaches" and use the AR training as a "complement to normal learning". The two participants that strongly agreed, stated that they think it's "more fun" and good to "consolidate knowledge". Two participants that disagreed, stated as their reasons that the time aspect was unrealistic and that they "appreciate the exchange with other people" and "do not like virtual learning".

Finally, when asked if participants had further feedback, notes or suggestions, they stated that they liked the virtual presentation, that "you have to pay attention to everything", that the auditive breathing and heart assessments train your "feel for it", and that they felt like they are "actually an actor in the scenario". While one participant noted that they liked that there was "no pressure", another participant stated that the training is "missing a time limit" to assess how well they did during specific steps of the reanimation procedure.

4.6.7 Evaluation of the Lecturers Perception

Throughout the timespan of project Heb@AR, the team conducted six, three to four hours long, workshops for lecturers, in which 7 to 11 female lecturers took part each, as described in Vogel et al. [481]. The workshops were designed as a hybrid format, depending on the current pandemic situation and if practical AR training parts were required. The workshops themes ranged from "Introduction to Augmented Reality", over "Integrating AR into teaching environments", to "creating your own AR trainings" in the concluding workshop. *The medical didactics researchers [217]* and *the midwifery researchers [35]* instrumentalized and recorded a wide range of more specific feedback regarding lecturers perspectives through observations, collaborative visual thinking techniques, and questionnaires. Beside the fact that less than half of the lecturers had any contact with AR before the workshops at all and that the workshops statistically significantly improved their attitude towards implementing AR into their teaching in general [481], these findings will be reported in forthcoming publications. Beyond this, the workshops were used to introduce the three trainings Tocolysis (Training 1), Sectio (Training 2), and Reanimation (Training 3) to the lecturers in the practical workshop parts, where they would complete the AR trainings. They would then provide expert feedback, to ensure technical accuracy, didactic value and realization of learning goals for all trainings within the Design-Based Research methodology and report the usability. Reported in the following are a subset of items regarding the Heb@AR App, specifically: the usability after introducing the trainings, their expert assessment of the didactic utility of the trainings, and their assessment of how and when it should be used in the curriculum.

Lecturers Usability of the AR Trainings

We recorded the perceived usability of lectures after they completed Training 2 and 3 (Sectio and Reanimation), using the SUS. Perceived usability data was not recorded from the perspective of the lecturers for Training 1, as the training introduction was done in a remote format, which would have influenced the results. As the recorded SUS instrumentalization is in line with the instrumentalization used to record the perceived usability for the students, we can directly compare the perceived usability of the trainings with the student groups (see Figure 4.32).

While the students reported a SUS study score of 69.81 (SD = 15.69) for Training 2, the preparation of a pregnant person for a c-section, which would indicate only "Ok" [27] and "marginally acceptable" usability [28], lecturers reported a SUS study score of 75.42 (SD = 10.25), which indicates "good" Usability [27], and would be considered "acceptable" usability [28]. With 6 participants, the lecturer SUS study score would only be considered around 35% conclusive, according

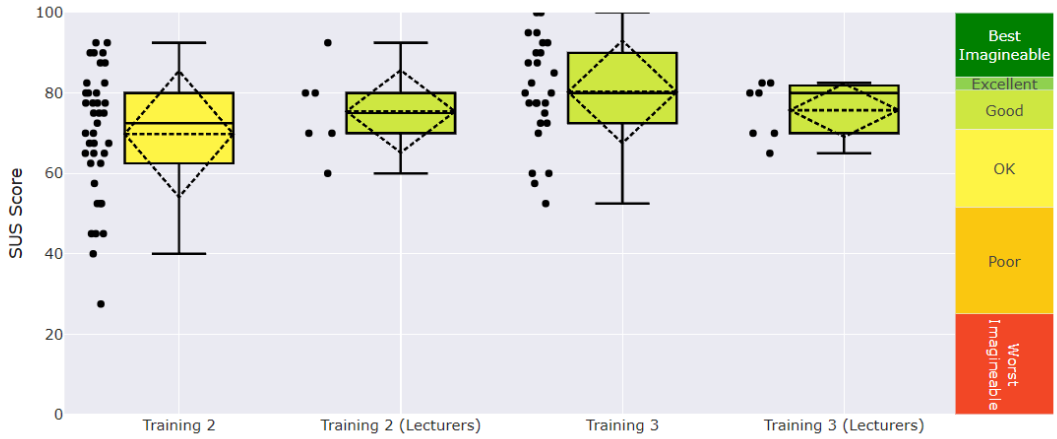


Figure 4.32: SUS Results from the Lectures for Training 2 and 3, recorded after the practical usage during the workshops, compared to the SUS results from the students recorded after the curricular implementation of the scenario.

to Tullis et al. [463]. When inspecting the averages of the SUS on item-level, differences between the groups were primarily visible in Question 1 (*"I think that I would like to use this system frequently."*), 6 (*"I thought there was too much inconsistency in this system."*) and 7 ("I would imagine that most people would learn to use this system very quickly.").

For AR Training 3, the reanimation of a newborn, the lecturers again reported higher SUS study scores with 80.29 (SD = 12.75), compared to the SUS study score of students with 75.71 (SD = 6.64), but both scores are contextualized equally, as "good" usability [27], which is considered "acceptable" according to Bangor et al. [28]. According to Tullis et al. [463], these results from the lecturers are 55% conclusive with a sample size of 7. All results calculated by the SUS Analysis Toolkit regarding the lecturers SUS scores are included in Appendix Table 31.

Moreover, we explored if the perceived usability in form of SUS scores significantly differed between the lecturers and student groups. To account for the two pairwise comparisons, a Bonferroni correction (p = 0.025) was applied. While a Shapiro-Wilk tests considers all four distributions to be normally distributed, non-parametric tests were used because of the small sample size for the lecturers, which neither follow a normal distribution when manually inspected, nor indicate homogeneity of variance. A Mann-Whitney U-Test indicated that there is no statistically significant difference between the SUS study scores of the lecturers (Mdn = 75) and students (Mdn = 72.5) for the Training 2 (U =94, n_1 = 6, n_2 = 39, p = 0.454). Another Mann-Whitney U test was conducted to compare the differences of SUS study scores between the groups for Training 3. There was no significant difference in SUS study scores (U =68, n_1 = 7, n_2 = 26, p = 0.325) between lecturers (Mdn = 80) and students (Mdn = 80).

Expert Assessment of the Didactic Utility

In the questionnaires during the workshops, *the medical didactics researchers [217]* asked very detailed questions about the lecturers' perception of the realization of individual, midwifery-specific learning goals and implementation aspects for each of the AR trainings. As these results are forthcoming and for the readability of this section, only the concluding question will be reported. In this question, *the medical didactics researchers [217]* asked how much lecturers would agree with the statement *"Overall, I find the didactic approach of the [...] in the AR app well done."* on a 5 point likert scale for all three training scenarios right after they completed the trainings with the AR app during the lecturer workshops. For all three trainings, lecturers strongly agreed with the statement: training 1 (m = 4.25, SD = 0.71, n = 8), training 2 (m = 4.29, SD = 0.49, n = 7), and training 3 (m = 4.89, SD = 0.38, n = 7). As can be seen in Figure 4.33, there is a slight trend that lecturers were especially fond of the didactic approach of Training 3. As not all lecturers attended every workshop over the months, not every lecturer answered each question for all trainings, but some answered for several trainings. Because of this, combined with the small sample sizes, potential dependencies cannot be accounted for in a statistical analysis. Therefore, the results of this question are only provided descriptively.

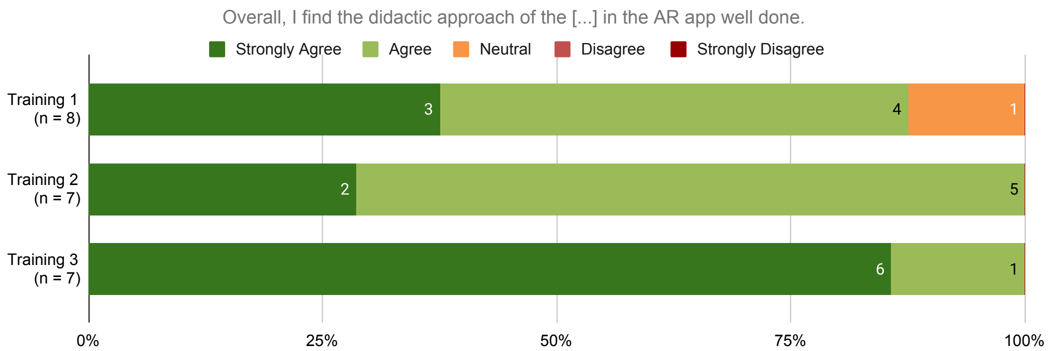


Figure 4.33: Lecturs agreement with the statement *"Overall, I find the didactic approach of the [...] in the AR app well done."* on a 5 point likert scale, directly after completing each of the 3 trainings.

Assessment of Possible Usage of the AR Trainings In Their Teaching

The medical didactics researchers [217] and *the midwifery researchers [35]* subsequently also asked them, how much they would agree with the statement *"The AR [...] opens up new perspectives for my teaching methods."* on a 5 point likert scale. Here, they agreed with the statement for training scenario 1 (m = 4, SD = 1.12, n = 9), strongly agreed for training scenario 2 (m = 4.57, SD = 0.53, n = 7), and again agreed for training 3 (m = 3.82, SD = 1.08, n = 11). The likert scale values are visualized in Figure 4.34. As not all lecturers attended every workshop over the months, not every lecturer answered each question for all trainings, but some answered for several trainings.

Because of this, combined with the small and uneven sample sizes, potential dependencies cannot be accounted for in a statistical analysis. Therefore, the results are only provided descriptively.

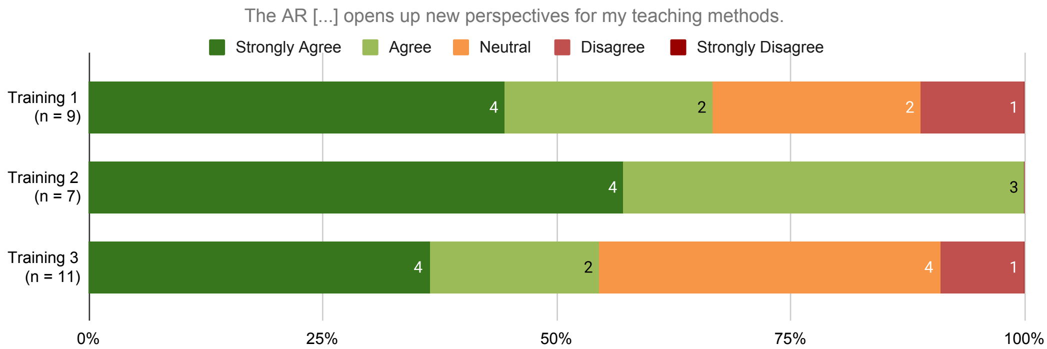


Figure 4.34: Lecturs agreement with the statement *"The AR [...] opens up new perspectives for my teaching methods."* on a 5 point likert scale, directly after completing each of the 3 trainings.

The medical didactics researchers [217] then asked the lecturers, *"With which objective would you want to carry out the AR training?"* (As a preparation for the practical training in the SkillsLab, or as a Consolidation after the practical training in the SkillsLab), for all three trainings. As visualized in Figure 4.35, they would primarily want to use the trainings as consolidation opportunities, rather than in preparation for practical SkillsLab trainings, across all three trainings.

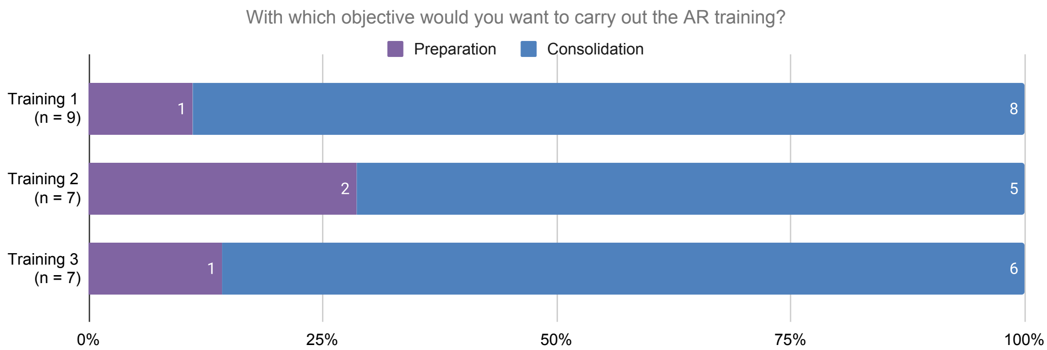


Figure 4.35: Lectures choices, with which objective they would want to carry out the AR trainings, either in preparation for, or as consolidation of, practical SkillsLab trainings.

Finally, across all AR trainings, *the medical didactics researchers [217]* furthermore asked the lecturers *"What support do you need to use AR in your teaching?"* as a qualitative question. The lecturers stated that they need "technical support" or "technical instructions" (3 lecturers), as they would not know "what to do, if they cannot progress during an AR training". They also stated that they would need close contact and exchange with developers of the Heb@AR App for changes to existing or new AR trainings (3 lecturers) and likely more financial support (2 lecturers). Ideally, they would like "somebody who is employed in the department for further development of trainings" as support. Furthermore, they emphasized that they would need continuous knowledge exchange with other lecturers and time and space to independently train with the AR trainings themselves, "to help students with technical problems". Combining this feedback with results from the collaborative visual thinking and non-representative observations during the workshops, we clustered this in Vogel et al. [481] as the four main support needs: *"practical introductions"*, *"time"*, *"technical support"*, and *"knowledge exchange"*.

4.6.8 Summative Evaluation: Objective Final Exam Comparisons

In addition to within-subject evaluations during the implementations of Training 1, 2, and 3 into the curriculum, *the midwifery researchers [35]* also evaluated the effect of the AR intervention on the OSCE results of one entire cohort, which used the app during their study (the "AR intervention cohort", n = 44), compared to a control cohort (n = 27) with the same course structure.

The OSCE, short for O(bjective) S(tructured) C(linical) E(xamination), is a structured examination to assess a student's competency reliably and objectively [185]. To achieve this, the OSCE consists of a series of timed stations with specific tasks or scenarios that the students must perform. These stations are then graded on a scale by an examiner. To explore potential effects of the Heb@AR App on these objective final exam results, they recorded the results for 4 stations during the OSCE of both cohorts:

- *"The preparation of an emergency tocolysis"*, subsequently referred to as **Station 1**, as it examens procedures trained using Training 1 of the Heb@AR App
- *"The preparation of a birthing person for cesarean section"*, referred to as **Station 2**, as it examens procedures trained using Training 2
- *"The reanimation of a newborn"*, subsequently referred to as **Station 3**, as it examens procedures trained using Training 3 of the app
- *"Positioning of a birthing person"*, subsequently referred to as the **control station**, as it serves as the baseline to compare the cohorts' overall performance.

While this methodology enables objective comparisons of the AR intervention, compared to the conventional training of procedures, it is limited in several ways and should be considered exploratory. Firstly, the sample size is small, and only two cohorts are compared. Secondly, the sample sizes are uneven. Thirdly, the AR intervention cohort was likely influenced by the pandemic situation during 2020 and 2021, which necessitated changes in teaching models in general and lead to fewer chances for actual practical application of the procedures in the students' practical phases. Most importantly, due to privacy considerations and ethical clearance, pseudonymization, which could be used to link final exam results to previous results, was not recorded. Moreover, no demographic data was recorded for the same reason. In this, the cohort performances are compared, but it is not known which students of the AR intervention cohort actually utilized which AR trainings, either during curricular usage or as a self-determined retention training. This likely weakens the statistical power to find actual differences substantially and increases the risk for Type II errors. Finally, Station 1 and 3 were examined by different examiners for each cohort, which could lead to examiner-related effects, e.g., because of different scoring criteria. This effect, called the "rater variability", is repeatedly shown to "explain more of the variability seen in trainees' scores than the trainees' own performances" [158]. Station 2 was examined by the same examiner and should therefore not be influenced by this effect, but was examined by somebody involved in the Heb@AR Project, which potentially could have led to unconscious biases.

To facilitate comparative statistical analysis across stations, which utilized different scoring scales, raw scores were transformed using a normalization process. This process rescaled the scores to a common range from 0 to 100 to make them directly comparable.

Baseline Comparison of the Cohorts Exam Performances

Analyzing the descriptive statistics for the OSCE results of the station "Positioning of a birthing person", used as a baseline for the exploration of the performance differences between both cohorts (see Figure 4.36), the control group had lower scores at the baseline task (m = 42.41, SD = 18.68, Mdn = 35) compared to the AR Intervention cohort (m = 48.47, SD = 17.94, Mdn = 50).

While Levenes test assumes homogeneity of variance in the sample (p = 0.589), a Shapiro-Wilk test indicated that the control cohorts OSCE results are not following a normal distribution (p = 0.043). Therefore, a non-parametric test was used to compare the differences between the independent samples. Tukey Fence indicated that results from the control cohort contain 1 and the AR intervention cohort 2 outliers. A Mann-Whitney U-Test, including the outliers, was deployed, which indicated that the difference between the control and AR intervention cohort with respect to the OSCE results at the station "Positioning of a birthing person" was not statistically significant (U = 451, p = 0.092, r = 0.2). The standardized effect size of r = 0.2 indicates that the magnitude of the non-significant difference between the values is small.

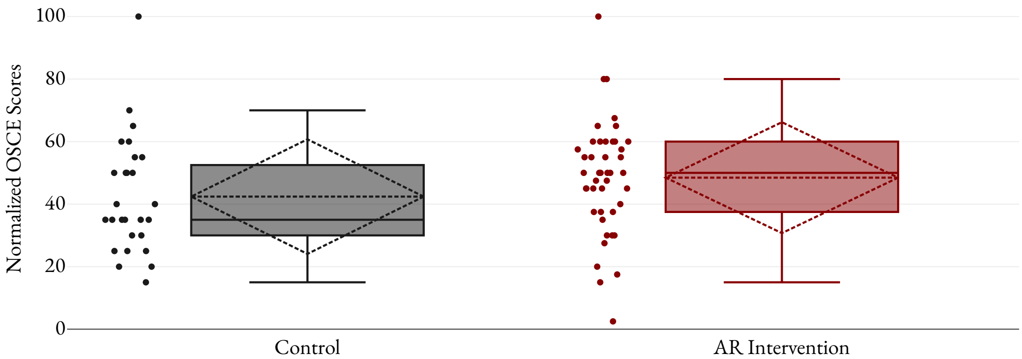


Figure 4.36: The normalized OSCE scores of the control station for control and AR intervention cohort.

Effects of the AR Intervention on the Cohorts OSCE Results

Analyzing the descriptive statistics of the exam results for Station 1 ("Preparation of an emergency tocolysis"), the control cohort achieved an average score of 77.37 (SD = 11.5, Mdn = 78), while the AR intervention cohort achieved an average score of 82.67 (SD = 13.01, Mdn = 85). For the OSCE Station 2 ("Preparation of a pregnant woman for a c-section"), the control cohort achieved an average score of 56.35 (SD = 15.73, Mdn = 58), and the AR intervention cohort an average score of 67.63 (SD = 17.59, 70.25). Finally, on the "Reanimation of a newborn" station, OSCE Station 3, the control cohort achieved an average score of 82 (SD = 9.08, Mdn = 84), while the AR intervention cohort achieved an average score of 59.55 (SD = 21.38, Mdn = 64).

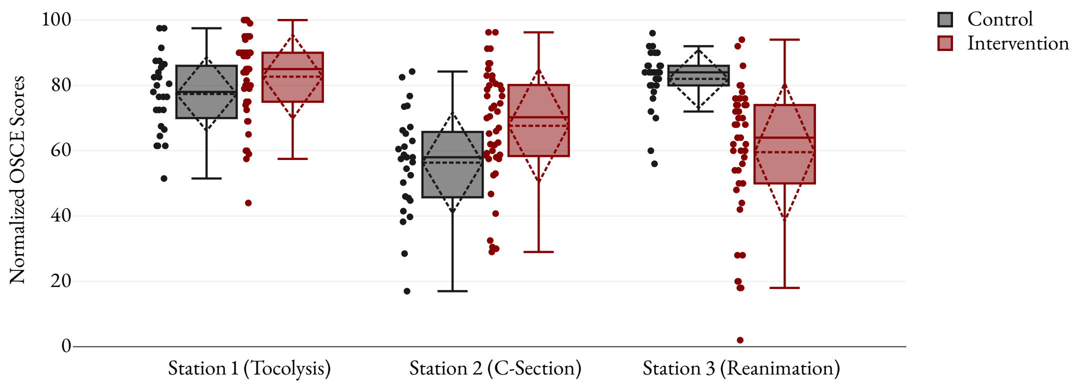


Figure 4.37: The normalized OSCE scores of the Station 1, 2, and 3, grouped by the cohort.

While manual inspection of all distributions indicated reasonable homogeneity of variance for most examined stations across the cohorts, the distribution of the results from Station 3 of the control cohort exhibits characteristics that deviate from the expected pattern (see Figure 4.37 for comparison and Figure 4.38 for a detailed visualization), raising concerns about the underlying assumptions or a potential anomaly. To further investigate, Levene's test for equality of variances was conducted to assess the homogeneity of variances assumption across all six groups. The results of the test indicated that there were significant differences in variances between the groups, F(5, 207) = 4.1183, p = 0.0014. Given the significant result, the assumption of homogeneity of variances is violated. Post-hoc pairwise comparisons were performed using Tukey's HSD Tests. Out of the 15 comparisons, two pairs demonstrated significant differences of variance. The first pair consisted of the results from station 3 of the control cohort (M = 82, SD = 9.08) and station 2 of the intervention cohort (M = 67.63, SD = 17), with a mean difference of 8.03, 95% CI [0.675, 15.393], p = 0.0234. The second pair included the results of station 3 of the control cohort and station 3 of the intervention cohort (M = 59.55, SD = 21.38), with a mean difference of 10, 95% CI [2.641, 17.359], p = 0.0017. No other pairs demonstrated significant differences of variance.

After careful consideration and examination of these results, it was decided to exclude the comparison between the control and intervention cohort for the results on station 3 from the subse-

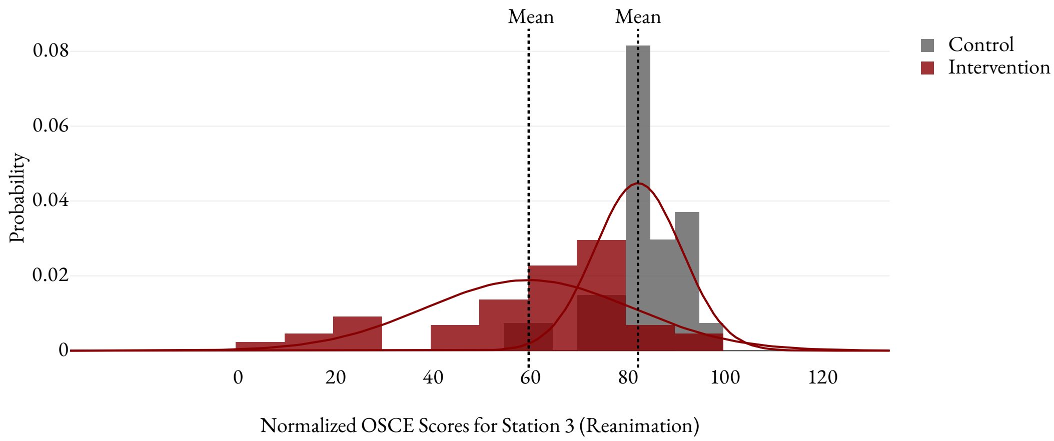


Figure 4.38: The probabilities of normalized OSCE scores for Station 3, the distribution's means, and closest normal distribution, visualized for both cohorts (control cohort, AR intervention cohort).

quent analysis. Not only is the observed variance of the results on station 3 for the control cohort not in line with variances observed for the other stations, but also the trend for this specific level was contrary to both the expectations and the trends observed in the remaining stations between cohorts. Additionally, the standardized effect size of the difference between control and AR interaction cohort is substantially higher than in any other comparison. In this, we believe the differences to be primarily caused by external (e.g., examiner-related) factors, rather than an actual difference in the cohort's performance of the task. Nonetheless, it should be noted for transparency, that when performing a pair-wise comparison of the results, the control cohort performed significantly better than the AR intervention cohort ($U = 156.5$, $p = < 0.001$, $r = 0.62$) at station 3, with $r = 0.62$ indicating there to be a large difference. With the exclusion of one of the three repeated levels because of concerns about underlying assumptions and external factors influencing the results, the remaining analysis should be considered entirely exploratory. For this reason, no corrections are applied, and results should be interpreted with caution.

A Shapiro-Wilk test indicated the remaining residuals to not follow a normal distribution ($p = 0.002$) but manual inspection using Q-Q plots suggested that the data is reasonably symmetric around the average and ANOVAs are considered robust for moderate violations of the normality assumption. Therefore, a two-way mixed ANOVA with one between-subjects factor (the cohort) and one within-subjects factor (the examination task) was performed. The results showed a significant difference between cohorts, $F(1, 52) = 9.22$, $p = 0.0037$. They also showed a significant difference between examination tasks, $F(1, 86) = 105.59$, $p < 0.001$. However, the interaction between both factors was not significant, $F(1, 86) = 2.96$, $p = 0.089$. The non-significant interaction suggests that the significant differences between cohorts are consistent and do not depend on the examination task, and vice versa.

As the parametric pair-wise comparisons are not robust for even moderate violations of the normality assumption and to be consistent with the previous pairwise comparisons for the other stations (Control and Station 3), non-parametric tests were used to check for pair-wise differences. For the OSCE Station 1, a non-corrected Mann-Whitney U-Test showed that the AR intervention cohort performed significantly better than the control cohort (U = 419.5, p = 0.04, r = 0.25). A second Mann-Whitney U-Test showed that the intervention cohort also performed significantly better than the control cohort on OSCE Station 2 (U = 357, p = 0.005, r = 0.33).

4.6.9 Summative Evaluation: Students' Retrospective Feedback

Beside the objective summative results, which were gathered in the OSCE exams, *the medical didactics researchers [217]* also gathered feedback from the students in a concluding summative evaluation questionnaire after the OSCE evaluations. When we analyzed these results (see Lewa et al. [275]), among other insights, they showed that 77—94% of students were satisfied with the work processes within the individual trainings and 80—96% found the Heb@AR App helpful as a learning medium (n = 26-40, across 3 trainings). Additionally, over half of the students indicated that they found the motor exercises in the AR app particularly helpful for theory-practice transfer. When asking them how well specific learning goals were supported by the app, feedback indicated that they perceived it as helpful for procedural knowledge, identification of pathological progressions, and medical hygiene concepts, but less helpful for communication and documentation aspects, as described in more detail in Lewa et al. [275].

In this questionnaire, *the medical didactics researchers [217]* also asked students to retrospectively assess their competency improvements across the implemented trainings and to assess how much the Heb@AR App, didactic choices, and specific implementation choices in AR, supported them in attaining their learning goals. Furthermore, *the midwifery researchers [35]* asked how well this competency transferred into practice. Out of the full cohort, 28 female participants completed the optional summative evaluation questionnaire.

Retrospective Competency Self-Assessment

As part of this questionnaire, *the medical didactics researchers [217]* asked, *"How do you rate your competencies for independent correct performance of this task on a scale from 0 to 100? (0 = no practical skills and 100 = complete practical skills acquired)"* for all three trainings in a retrospective pre-post design [46]. Out of 28 participants, 22 completed all the trainings, and therefore provided complete retrospective assessments for all three trainings. For simplicity reasons, the other 6 participants were excluded for the analysis of this question. Those 22 participants assessed their competency to prepare a tocolytic injection (Training 1) as 57.73 (SD = 27.56) before and 87.23 (SD = 11.38) after the training. For the preparation of a pregnant woman for a c-section (Training 2), their self-assessment before the training was 68.5 (SD = 24.27) and after the training 81.36 (SD = 16.94). Finally, they assessed their competency regarding the reanimation of a newborn (Training 3) as 35.91 (SD = 27.47) before and 72.05 (SD = 22.76) after the training (see Figure 4.39).

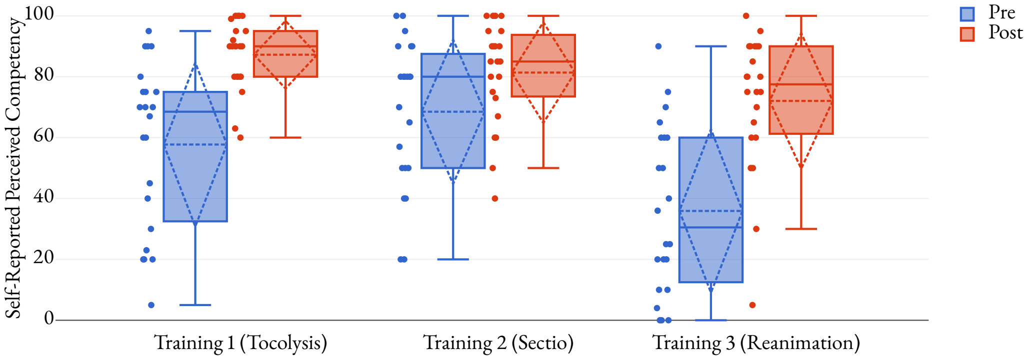


Figure 4.39: The retrospectively self-reported competency assessments of students which completed all three AR trainings before and after the training.

A two-factor ANOVA with repeated measures was utilized to detect differences between the pre-post assessments across all three scenarios, with interaction effects. A Shapiro-Wilk test indicated the residuals to not follow a normal distribution (p = 0.009). While they significantly deviate from normality, manual inspection of Q-Q plots, indicated the severity of deviation is slim, and they are sufficiently normally distributed and reasonably symmetric because of the sample size. This ANOVA is considered robust for moderate violations of normality. The analysis revealed a significant large difference in self-assessed competency improvement between trainings (F(2, 84) = 17.3, p < 0.001, $\eta^2 = 0.292$) and a significant large main effect of the pre- / post-measures for the trainings (F(1, 84) = 26.54, p < 0.001, $\eta^2 = 0.446$). Moreover, there was a significant medium interaction effect between the two factors, F(2, 84) = 4.74, p = 0.011, $\eta^2 = 0.102$.

Given the significant main effect for the pre- / post-measures, we deployed Bonferroni corrected (p = 0.0167) paired t-tests to investigate which trainings significantly improved the competency of the student when retrospectively self-assessed. There was a significant large improvement for training 1 (t(21) = 6.4, p < 0.001, d = 1.36), training 2 (t(21) = 4.1, p < .001, d = 0.87), and also training 3 (t(21) = 7, p < 0.001, d = 1.49).

As the main effect of the difference between trainings was also significant, Bonferroni corrected (p = 0.0167) paired t-tests were used to compare all significant competency improvements between the three trainings. There was a significant medium improvement difference between Training 1 (M = 29.5, SD = 21.7) and Training 2 (M = 12.9, SD = 14.7), t(21) = 3.2, p = 0.005, d = 0.67. Furthermore, there was a significant large difference between Training 2 and Training 3 (M = 36.1, SD = 24.2), t(21) = 4.5, p < 0.001, d = 0.96. Finally, the small difference between Training 1 and Training 3 (t(21) = 1.4, p = .163, d = 0.31) was not statistically significant.

Competency Transfer Into Practice

The midwifery researchers [35] then asked students, *"In general, how would you rate the impact of AR simulations on your perceived competency to act during the practical study phase?"* as a qualitative question.

Here, 13 students provided feedback indicating, that they do perceive an impact of the app on their practical study phase competency. For their reasoning, they primarily state that they feel "more confident" and "consolidated" in their knowledge because of the "systematic" procedural nature of the training (10 students), but also state that the possibility to "refresh knowledge" at home has helped them (3 students), and that they liked the visualizations (2 students). Another 2 participants were undecided and stated that there was "somewhat" of an impact, but actual practical trainings should be trained more often instead. Finally, 6 students provided feedback that indicates that they perceived the impact as "low"/"little" (4 students), or that they think it had "no impact" (2 students). Out of those, 2 students stated that they think there would have been a higher impact if the trainings had been introduced earlier into the curriculum and if there were more "practical implementations".

Assessment of the Individual Elements Support Towards Learning Goal Attainment

To provide a way for students to self-assess how well specific technical/didactic expressions of the two available interaction concepts, across trainings, helped them attain their learning goals, *the medical didactics researchers [217]* asked the students, *"How effective do you consider these AR implementations to be for your learning success?"* for purely virtual trainings (TrainAR) vs. trainings with motor components (Decide-Freeze-Imitate) and purely virtual object interactions (TrainAR) vs. decision tree interactions (Decide-Freeze-Imitate) on 4-point Likert scales, ranging from "not effective" to "very effective", without a "neutral". Here, students found the purely virtual training form of TrainAR less effective (m = 2.48, SD = 0.59) but the interaction concept of TrainAR with virtual objects effective (m = 2.67, SD = 0.73). Students found the training including motor components (Decide-Freeze-Imitate) very effective (m = 3.5, SD = 0.51) and they found its interaction through decision trees effective (m = 3.11, SD = 0.75). Their answers are visualized in Figure 4.40. Spearman correlation analysis indicates that there is a non-significant medium correlation of answers between "purely virtual trainings" and "training flow with virtual object interactions" (r(22) = 0.38, p = .063) and a statistically significant high correlation between the "training with motor components" and "training flow with decision trees" (r(25) = 0,56, p = 0,002). But there were no correlations between "purely virtual training" and "training with motor components" (r(23) = -0.08, p = 0.719) and "training flow with virtual object interactions" and "training flow with decision trees" (r(25) = 0.06, p = 0.782), indicating the answers were not caused by individual preferences for one of the AR interaction concepts.

The medical didactics researchers [217] also asked students to retrospectively access *"To what extent did the following didactic delivery methods within the AR app help you to achieve your learning success?"* for several of the didactic elements of the Heb@AR Apps' trainings on a 5-point Likert scale, ranging from *"not supported at all"* to *"fully supported"*. Their answers are visualized in Fig-

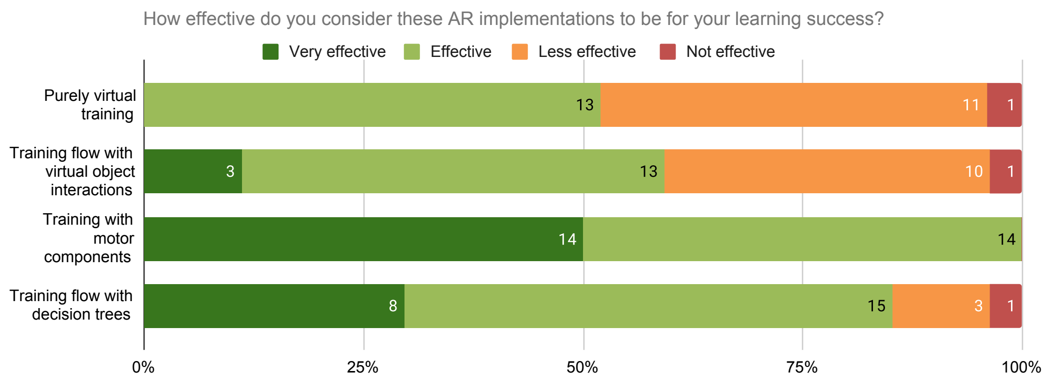


Figure 4.40: Students' assessment of how well specific technical/didactic expressions across trainings helped them for their learning success.

ure 4.41. Students reported that both the Audio (m = 3.71, SD = 0.76) and textual (m = 3,86, SD = 0.85) expert hints supported them, with a slight trend toward the audio hints. They felt like the textual instructions (m = 4.07, SD = 0.72) and the error feedback (m = 4.18, SD = 0.82 for textual and 3.71, SD = 1.01 for audio) supported them. When asked how much the instruction, training, and free modes (primarily present in Training 1) helped them, they stated that the instruction and training mode fully supported them (m = 4.35, SD = 0.69, and 4.42, SD = 0.70) and that the free mode supported them (4.00, SD = 1.10). Finally, regarding the summaries, they reported that the error counter (3.61, SD = 0.99) but also the detailed error summaries (3.70, SD = 0.87) helped them. They were neutral regarding the timer of the training summary (3.44, SD = 1.01)

4.7 Discussion

The interdisciplinary efforts of Project Heb@AR offer many interesting perspectives to discuss. For the purpose of this thesis, we will only discuss the Heb@AR App, its development, and its evaluation from the software development and HCI perspective. In this, as described in Section 4.6.1, perspectives described here do not necessarily represent interpretations or opinions of the project partners. Further discussions of more specific aspects can be found in our already published project publications [52, 53, 54, 60, 274, 275, 293, 481] and forthcoming works. Throughout the discussion, first the Heb@AR App is discussed as a learning construct and how it is a first endeavor toward our vision of ARBTs, then the development process of the app is discussed with its challenges and opportunities. Afterward, we discuss the preliminary insights from the current evaluation data of implementing the app into a midwifery curriculum, and finally address practical implications and potential future work.

Implications of the development of the Heb@AR App and its evaluation from the perspective of the exploration of the design space of AR authoring tools are discussed in the following chapter.

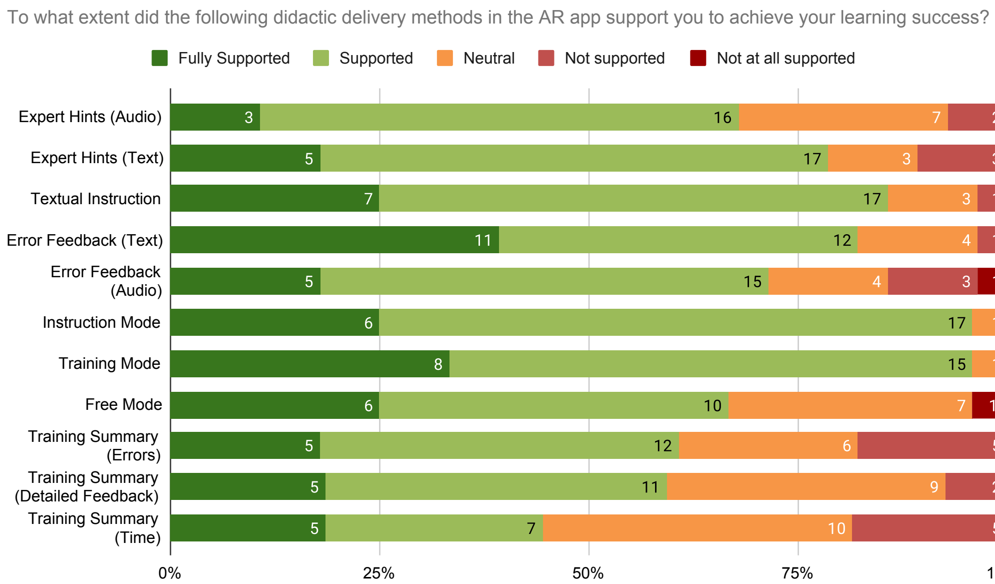


Figure 4.41: Students' assessment of how well components and didactic delivery methods of the Heb@AR app supported them in attaining their learning goals.

4.7.1 The Heb@AR App and Our Vision of ARBTs

While we initially developed the AR trainings of the app as prototypes, after several iterative improvements and refinements, the final combination of them in the Heb@AR App is not a prototype. It is a fully functional app that is available as an OER in both Android [23] and iOS [24] app stores and can be used by institutions and midwifery students today. With its 5 AR trainings, totaling over 4 hours worth of raw AR training content (without the supporting structures like WBT-based preparations for the AR trainings), supplementary material, and supporting structures like feedback forms for feature requests, it is, to our knowledge, the most comprehensive and largest handheld AR training application to date. In terms of included training content, which has been professionally refurbished and transferred towards AR trainings from the didactic perspective, it might also be the most comprehensive AR training application in general, including other hardware platforms. Beyond its size in terms of training content, it also incorporates several applied HCI research considerations, like consistency of UIs, consistent interaction metaphors, consistency of feedback modalities, and generally deliberate design choices toward shifting the focus of trainees from the novel technology to the training content of the AR trainings.

[23] https://play.google.com/store/apps/details?id=de.Mixality.HebAR
[24] https://apps.apple.com/app/heb-ar/id1621822317

In this, the Heb@AR App is a first exploratory step toward our vision of scalable ARBTs, even beyond the scope of AR trainings in the context of academic midwifery education. Therefore, we believe, learnings are likely transferable to other contexts and hope for the app to serve as a reference of a successful implementation of AR trainings for complex procedural training content in a structured and scalable way.

Referencing back to our identified factors for scalability in Section 4.1, firstly, the implementation of the app is exclusively done on ubiquitously available devices and specific efforts were made to not subsequently narrow this availability during development. For example, no additional technical hardware is required, like tracking aids or external hardware components, and no software packages outside the broadly covered AR tracking libraries of Vuforia, ARCore for Android, and ARKit for iOS, were used to implement the functionality. While this does limit more granular content-specific implementation opportunities, it ensures that the app realistically scaled to the full spectrum of the BYOD methodology and can currently be used on 48 iOS and 707 Android models. Secondly, place, and time independent learning is incorporated in the app as a central concept. While the larger SkillsLab exercises require the usage of training dummies and medical consumables during the training to incorporate the motor components of the procedures, the Training@Home versions are fully place and time independent. The AR SkillsLab exercises, though not fully independent in terms of location and time, at least do encourage self-regulated and independent engagement with the procedural training content. Thirdly, through the usage of context, technical, and lecturer-lead onboarding, but also the implementation of low-level interaction concepts and content-focused UIs, the usability, and low entry threshold for new users were largely achieved, as proven by the evaluation efforts. Finally, the app is deliberate in its consistent interaction metaphors and didactic presentation of content, feedback and additional information. This should not only help students to quickly learn the technical expressions of the AR trainings, and therefore focus on the learning content instead of the technology, but also help teachers to correctly implement the existing and new learning material.

We perceive the Heb@AR App as a scalable learning tool, an applied research "puzzle piece", in an ever increasingly multi-medial training environment for students in the academic context. A puzzle piece that, contrary to HMD-based MR approaches, can be realistically scalable today. While we contribute an OER-based learning app for the academic midwifery education specifically, the implications go beyond the app itself. While, in the long run, HMD-based approaches will likely outperform handheld approaches in almost any dimension and make interaction concepts like the Decide-Freeze-Imitate concept, inherently designed to circumnavigate the limitation of having a smartphone in one hand, obsolete, this has several HCI-specific research contributions. Firstly, in this, students engaging with the handheld AR technology first, will familiarize with AR and likely, when introduced to more cognitively demanding HMD-based AR approaches, be more accepting of the novel technology, which might simplify one aspect of the adoption. Secondly, it provides already fully developed procedural AR training content, that can be easily transferred towards HMD-AR or even VR approaches. Once the stateflow with its didactic considerations is finalized and all the textual content of the training, but also its assets, are already available, the technical transfer from handheld to HMD-based AR is substantially simplified. This is especially true with the chosen implementation platform of Unity. Because of this, with the availability of HMD-based AR devices that are scalable in the future, the AR training content is almost immediately available to use on them.

Ultimately, the app is intended as a supplement to be utilized by lecturers or students independently, not a singular or standalone solution for a learning problem. This perspective has two practical implications. Firstly, the design of the app, concepts for interaction, and technical implementation decisions are often pragmatic in their design and more applied than fundamental research. While the two interaction concepts developed for the app are novel and add to the body of work in AR interaction concepts in terms of fundamental HCI research, they deliberately do not incorporate more complex, experimental features, which would have been possible to explore, like hand-tracking, gesture recognition, object-detection or automatic state-segmentation based on computer vision approaches. Secondly, with this, in our vision the AR trainings of the Heb@AR App in the midwifery context, but also ARBTs in general, might not necessarily have to outperform conventional methods, other multimedia elements like quizzes, or even VR training applications. They are not intended to replace them. We think, it would likely suffice to show that they are worth developing and that they themselves do elicit learning benefits.

4.7.2 Challenges & Opportunities of Developing the AR Trainings

We faced several challenges and opportunities during the development of the AR trainings as a team during Project Heb@AR. One inherent challenge of the interdisciplinary development approach was the incorporation of all perspectives, as previously described in Section 4.3. The AR trainings always had to be correct from the midwifery researchers perspective, in line with the standard-of-care guidelines, had to be didactically valuable from the didactics researchers perspective, and still not only be technically possible to implement but ideally fit into a consistent, scalable concept from the HCI perspective.

A Causality Dilemma of Developing AR Trainings

While this on its own is already a challenge, these are likely challenges to be expected in interdisciplinary training development, and we did, at least partially, expect this. This is why we initially chose the DBR methodology as the primary development approach in the research proposal of the project. Nonetheless, this challenge was amplified by a causality dilemma: From the perspective of the midwifery and medical didactics researchers, it was initially difficult to envision the technical implementation of the final AR trainings, especially regarding aspects such as user interaction metaphors. This made it challenging to precisely define the training content and to establish the didactic objectives and expressions before starting the technical development. In the end, technical feasibility also influences realistic learning goals or how specific medical aspects are depicted best. Conversely, for us as the HCI researchers, the development of AR training modules was challenging without an established understanding of the intended training content and the didactic considerations that needed to be integrated. This dilemma mainly originated from the challenging starting point that there were no existing reference works or even interaction concepts for procedural handheld AR trainings that would fit the initial drafts of our ideas and didactic requirements during the exploratory stages of development.

Development of AR Training Transfer Procedures

As we developed several AR trainings, we took this challenge as an opportunity to explore and contribute a more structured approach of transferring the complex procedural task trainings toward AR trainings as a side-contribution of our efforts. While our approach is based on existing frameworks like the Blooms Taxonomy [242], the MARE model [529], and task-process analyses, that are already broadly used in training formalization efforts, and therefore is a more practical application of existing components than an entirely novel framework, it proved to be an effective way of continuous communication between the disciplines for us. Even in the iterative stages, beyond the initial transfer of the procedural training content, it helped to discuss potential changes based on a shared formalization, the training stateflow, that is closer to the implementation than the actual task in reality, compared to the work-process analyses.

In line with our hope that other researchers perceive the Heb@AR App with its AR trainings as a successful implementation of AR trainings for complex procedural training tasks to use as a reference, in this, we do not only deliver the final app as a reference but also the interdisciplinary blueprint on how similar AR trainings can be achieved. We believe, this will help other researchers and developers to more effectively and efficiently transfer their ideas, using the combination of our app and development methodology as a starting point.

Development of Novel Interaction Concepts

As stated, the major contributing factor of the causality dilemma during the development of the first AR training of each of the training types, was the missing references to create a shared understanding of how the final AR training could look like. As can be derived from the development timelines of the AR trainings in Figure 4.7 (E.g., Training 2 and 3, which both use the Decide-Freeze-Imitate concept and Training 1 and 5, which use the TrainAR interaction concept), when the Interaction concepts were subsequently reused, both the conceptual but also technical development efforts appear to have been accelerated, which is in line with our own perception.

Interaction concepts that would cover the learning requirements for the trainings we wanted to develop during Project Heb@AR would likely not have been a major challenge in the context of VR or HMD-based AR, where at least the majority of interaction metaphors are already covered by established development frameworks. For handheld AR, they were entirely missing and for this, we conceptually developed them from scratch and contributed them to the literature [53, 55]. While the interaction concepts likely cannot compete with their HMD-based AR or VR counterparts in terms of interactivity, we aimed to position the concepts to maximize their usability, scalability, and accelerated onboarding of students. They were designed to address people with little to no media competency regarding immersive technologies.

In the Heb@AR App, the two interaction concepts supplement each other, with the TrainAR interaction concept being entirely independent of physical material, and time and location independent, but not training motor components of the procedures, and the Decide-Freeze-Imitate concept training motor components but only being usable as an exercise during practical SkillsLab sessions, limiting the self-regulated learning perspective. Nonetheless, we believe the interaction concepts could not only be applied to other procedural learning tasks and even contexts, but could also be used independently of each other. Which interaction concept would be appropriate to use is based on the importance of specific learning objectives, which would become apparent during the "establishment of learning goals" in our proposed development methodology. Finally, even if other researchers are not interested in directly applying our interaction concepts to their procedural training tasks, they could still serve as a reference point to partially aid the causality dilemma of interdisciplinary AR training development efforts, as they provide a first example of how others approached this challenge.

Technical Development Challenges

While the focus of this chapter is on the HCI perspective more than the pure software development perspective of the development of the AR trainings, we want to cover the technical development challenges at least briefly. The size and complexity of the Heb@AR App and the goal for it to be more than just a prototype self-evidently comes with many technical hurdles, challenges, and balancing of competing implementation decisions, but likely most importantly for other researchers and developers to consider are the following.

The hardware performance of handheld AR devices was a severely limiting factor in the development of all AR trainings. Especially because of the complexity and size of the AR trainings with several shaders, animations, and large objects like entire reanimation stations, special considerations for the performance of the trainings had to be made. We accomplished acceptable performance on most devices by, e.g., by lowering the complexity of models and textures, using only selected shader implementations, and not using resource intensive C# functionality like runtime inverse kinematic or mesh outlining features, but rather pre-computing them.

These performance limitations were only exaggerated by the usage of ARCore or ARKit in combination with Vuforia as a secondary tracking library for marker tracking (Training 2 and 3), which, without any functionality implemented by us, already used the entire available resources on many mid-range Android devices. We addressed this by using the AR-marker tracking by Vuforia only on demand and intermediately deactivating the libraries' functionality during the AR training, handing off the tracking to ARCore/ARKit after initial detection of AR-markers. This combination of tracking libraries and subsequent performance demands on the smartphones were also the reason for the initial problems with crashes, we experienced with Training 3 (see Section 4.6.4), that were fixed by this approach and stability updates to the libraries. Here, initially, a combination of long usage and overheating caused segmentation faults.

Finally, the scalability of the app, while being one of its selling points, is also a technical hurdle, as several hundred devices have to be supported. Therefore, performance considerations have to be made for the app to also work on mid-range devices and UI considerations have to be made to work on different screen sizes, aspect rations, and with screen notches.

4.7.3 Insights Based on the Evaluation Efforts of Project Heb@AR

Combining all results from the selective-variable analysis of the evaluation efforts of Project Heb@AR reported in Section 4.6, we can now discuss our findings. For this, we will first discuss the results for the individual trainings, grouped by their evaluation type, discuss some non-representative observations not visible in the evaluation data, and then combine the findings to discuss the research questions on the usability and utility of the AR trainings. Subsequently, we discuss the usefulness of the Heb@AR app as a learning tool in the midwifery context and beyond.

Curricular Evaluations of Training 1, 2, and 3

Looking at the utility data for the curricularly implemented and evaluated trainings 1, 2, and 3, in the form of the competency pre-post measures, they all provide promising insights. Training 1, the preparation of an emergency tocolysis based on the Training@Home TrainAR concept, showed a significant large increase in perceived competency in the students after completing the training. Interestingly, beside a smaller standard deviation in the post measures compared to the pre measures, a negative skewness can be observed, when plotting the data of the post-measures. This could be an indication of a ceiling effect, but could also be interpreted as a greater consistency in the perceived competency among students, with only a minor proportion of students not being properly addressed by the AR training.

The perceived competency differences for AR Training 2, the permanent catheter placement as a SkillsLab exercise with the Decide-Freeze-Imitate concept, also exhibit a significant large difference across both recorded cohorts, in line with the results for Training 1. Again, a negative skewness in the post-measures can be observed, though the difference appears to shrink with the larger sample size. In line with this exploratory analysis, especially interesting is the fact that the residuals (difference of the differences) between cohorts were significantly different, but the post measures were not. This could also be an indication, that the AR training helped to create more consistent perceived competencies after completing the trainings among students.

While no pre-post measures of perceived competency were selected for Training 3, the SkillsLab exercise of the reanimating a newborn, from the combination of the students' self-assessments after the training, but also retrospectively in their learning diaries, after completing their practical phase, it is clear that they not only strongly agree when asked to self-assess their competency increase directly after the training, but the large majority also retrospectively thinks that the AR training did help them to feel more prepared during their practical phase. Students who did not directly agree that it helped them to feel more prepared, stated that they did not encounter the situation during their practical phase, or that they wished for even more support and training opportunities.

In terms of the trainings' usability, they all achieved at least marginally acceptable usability scores. With such a novel technology for the students to engage with for the first time, and the development of, at that point exploratory, novel interaction concepts, these results are sufficient indications for the trainings' long-term usability, but they do provide some insights. Training 1 had the highest perceived usability, which would be described as "excellent", while Training 3 had the second-highest perceived usability, which would be considered "good", and Training 2 had only "marginally acceptable", "ok" usability. Firstly, the ranking of the perceived usability scores seems to correlate with the training content complexity and especially with the length of the AR components of the training. Additionally, the usage of medical consumables and training dummies for Training 2 and 3 are inherently more complex than the purely virtual interaction concept TrainAR, which is used for Training 1. The SUS results are displayed in Figure 4.42. Not plotted in this graph are the SUS results for the multi-user version of Training 3. Students of the cohort which used the multi-user version of Training 3 reported a significantly lower perceived usability of the training. This is, at least partially, caused by training-terminating crashes of the apps at the time, which were caused by not anticipating stability problems caused by the larger institutional Wi-Fi networks at the time and not incorporating appropriate reconnection functionality because of the exploratory nature of the feature. Those problems then effected the whole group during the training session. But even if the students who encountered the crashes are removed, the perceived usability only marginally improved. Therefore, there are likely other factors influencing these results, which will be discussed in the non-representative observations section in detail.

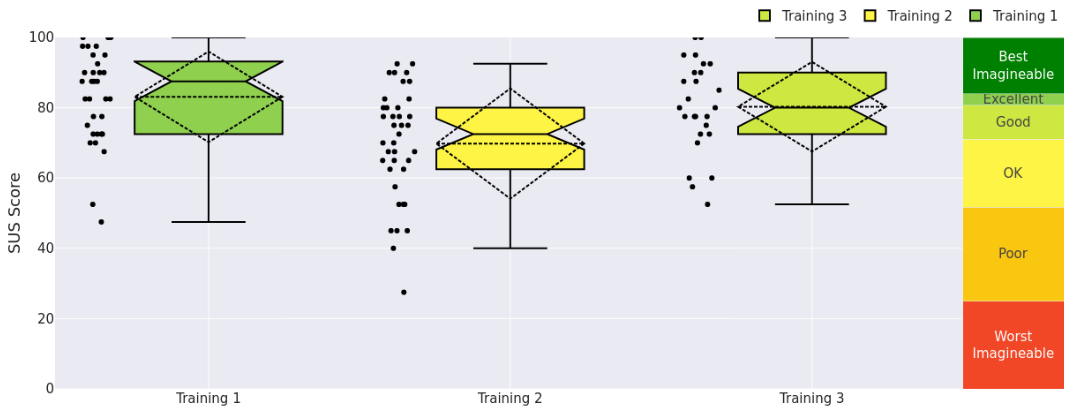


Figure 4.42: The SUS results for Training 1, 2, and 3 of the Heb@AR App, contextualized with the adjective rating scale. Plotted with the SUS Analysis Toolkit [50]

Evaluation Studies for Training 4, and 5

As expected, Training 5 increased the perceived competence (H1.1) and intrinsic motivation (H1.3) of the students compared to their previous experience with traditional methods. This is in line with previous research [6, 9, 153]. Perceived competence is theorized to support the overall intrinsic motivation [407], which could even elicit the measured result. Interestingly, hypothesis 1.2 had to be dismissed, since perceived pressure/tension increased after using the training. There are multiple potential explanations for this result. On the one hand, this could have been a novelty effect, as the students were very inexperienced with AR in general and used an AR training for the first time (see Figure 4.7). But arguably this should also have then been visible in the perceived usability, which was not evident in the results of the SUS. On the other hand, and the most likely explanation in our opinion, this could have been caused by the training actually requiring one to contextualize all pieces successfully to complete all levels. This likely not only creates the immediate pressure of connecting every piece correctly to complete the game, but also the pressure of questioning how well the knowledge was actually previously acquired, as there was no validation method for the conventional learning method. There could also have been pressures created by this being part of a lecture with other students present, or the time constraints. Moreover, compared to the other two subscales, the difference in perceived pressure was smaller, although still significant. In general, the results indicate Training 5 increased the overall motivation of students through their perceived competence and intrinsic motivation scores significantly compared to the conventional learning methods they previously used (Q1).

For Training 4, as expected, the intrinsic motivation significantly increased after completing the AR training, compared to their previous experience with traditional methods. Contrary to our expectations, no significant changes were found for the perceived competency and pressure/tension. In contrast to our expectations as well, observed tendencies even indicate that the perceived competency and pressure/tension decreased after the AR training. To analyze these surprising results further, we investigated the correlation of all IMI [406] subscales. It is interesting, that the perceived competency pre-post measures were highly negatively correlated for Training 4, while they were highly positively correlated for Training 5. For Training 4, this would mean that students who felt more confident before the AR training, decreased in their perceived competence even more after the training. While there are potentially interesting implications in these findings to answer Q1, the sample size of $n = 9$ for Training 4 would not allow an appropriate discussion and this might have been caused by external factors or the sample size.

Training 5 was not only usable by the midwifery students, but achieved exceptionally high SUS scores (see Figure 4.43, right), answering Q2 with a clear yes for this training. This is partially explained by the technical maturity of the app at the time of implementing this training, but also by the fact that, compared to more complex procedural trainings with several interactions, the game could be completed with the same interaction of "combining" two objects that had to be repeated multiple times. This SUS study score was achieved despite the fact that the participants had virtually no previous AR experience (see Figure 4.7). Additionally, pre-existing Latin certification, or the lack thereof, did not influence the usability of Training 5. While this is to be expected, this indicates the learning content was successfully separated from the interaction metaphors of the training, and it is usable regardless of Latin knowledge, as desired.

Moreover, Training 4 was usable by the midwifery students as well, though the evaluation was considerably smaller in sample size and therefore neither conclusive according to Tullies et al. [463], nor appears to be symmetrically distributed around the average (see the lower end of the notched box plot for Training 4 in Figure 4.43 on the left). As it achieved "good" usability scores, in line with usability scores recorded for its SkillsLab version (Training 3), we argue that Q2 can therefore also be answered with a yes for Training 4. Nonetheless, it should be noted that the SUS study scores for the SkillsLab version (M = 80.29, SD = 12.76) were higher than the virtual version (m = 71.25, SD = 15.6) of the training. Combining the occurrence of crashes during the evaluation of Training 3 with the fact that Training 5 is, for the most part, the same AR training, just without the motor components, this result is surprising. As analyzed in Section 4.6.6, this might have been caused by students rating their confidence with the contents of the AR training, rather than the usage of the app. Especially as "competency" was already eliciting surprising results as a subscale of the IMI [406]. The evaluation was contextualized right before an optional on-site SkillsLab exercise in preparation for an upcoming exam and is the second surprising result from this dataset. Therefore, we argue that this might have generally influenced the results for Training 4, and they should be interpreted with special caution or even disregarded entirely.

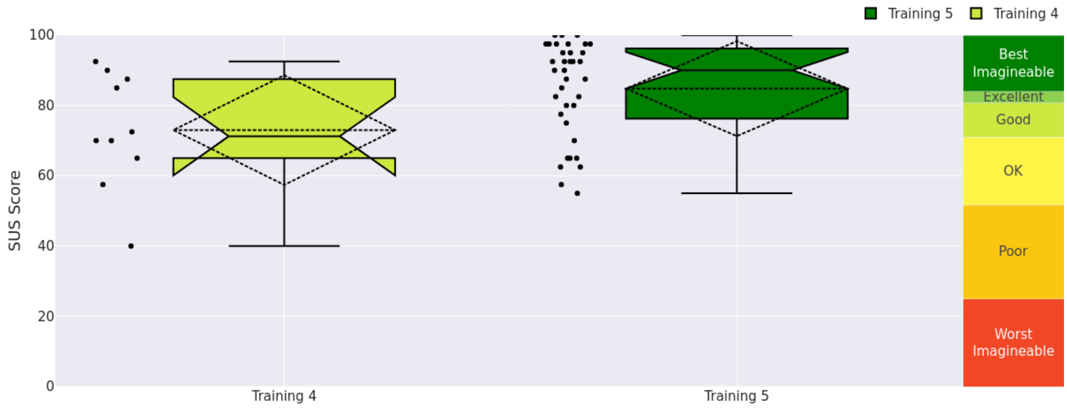


Figure 4.43: The SUS results for Training 4 and 5 of the Heb@AR App, contextualized with the adjective rating scale. Plotted with the SUS Analysis Toolkit [50]

Answering Q3 for Training 5, the qualitative feedback indicated that students perceived the interactive handheld AR Training as a valuable addition to existing learning approaches, which was in line with our expectations (H3). When looking at the Likert scale value alone, participants seemed to be less critical of the Pelvis AR Training that replaces conventional learning methods than expected. Looking at the qualitative feedback though, they did indicate that they would prefer a mix of both the conventional method and the Pelvis AR Training, indicating that our statement might not have been clear enough and the results of the second qualitative statement have to be interpreted with caution. Notably, participants also focused on the "visualization" benefits of the game more often than on the "interactivity" and "fun" aspects in their qualitative feedback, which we perceived as the more important benefit in incorporating cognitive, affective, and psychomotor learning holistically. This, however, could also be a novelty effect as the visual aspects of AR are deviating the most from conventional methods and known forms of interactions with learning content, e.g., through smartphone apps or web-based trainings. Finally, while the engagement with the optional qualitative questions was relatively low, we believe this was no indication of little involvement in the study or disinterest in Training 5, but rather caused by time-pressures during the testing in actual curricular usage and some participants filling out the questionnaire forms on their smartphones.

Answering Q3 for Training 4, the results are entirely in line with our expectations. Students agreed and stated that Training 4 would be a good opportunity for retention/"follow-up" to them, and they agreed that they would rather use the AR training instead of their conventional method. Again, in line with the answers provided for Training 5, most participants state that they would ideally want to "combine" approaches. As expected, when asked if they think Training 5 could entirely replace Training 3, they strongly disagreed, stating the importance of the motor components as their primary reason.

Lecturers' Perspective on the Heb@AR App

As lecturers are ultimately the decision makers of the usage of the Heb@AR App as a component of their teaching curriculum, their perspective on the utility and usability of the app is at least as important as the students' perspective.

In terms of utility, the lecturers not only strongly agreed that the didactic approaches chosen in the AR trainings 1, 2, and 3 were well done, but also agreed, even strongly agreed for Training 2, that the AR trainings open up new perspectives for their teaching methods. We interpret the combination of these results as an indicating that they do see utility in the app, which is further supported by the qualitative feedback, where lecturers already mentioned wishes for technical support and funding, and our non-representative discussions during the workshops.

In terms of usability, the recorded SUS results for Training 2 and 3, the two trainings that are SkillsLab exercises and would therefore have to be actively supported by them during the practical lectures, show no statistically significant differences in usability for them, compared to the students' perceived usability. While the sample size was small, this would be the desired result.

Summative Evaluations of the Heb@AR App

Combining the summative results of the OSCE comparisons to a control cohort with the subjective retrospective feedback, students of the first intervention cohort provided in a concluding questionnaire, the impressions of the students themselves during the curricular implementation (but also the lecturers expert assessment) are only further supported.

Though limited and to be interpreted as exploratory, the OSCE results of the AR intervention cohort were significantly better than the control cohort across included stations that were addressed by one of the trainings, while the control station was not significantly different.

When the students were asked to retrospectively assess their perceived competency before and after completing the AR trainings, they reported that they perceived competency increases across all three trainings, including interaction effects. These results suggest that the largest increases in perceived competency can be found after completing Training 3, followed by Training 1 and then Training 2. When asked, how they perceived the impact of the AR trainings on their competency to act during the practical study phase, the majority did perceive a positive impact, while some only perceived somewhat of a positive impact. Nobody reported a negative impact, and the infrequent students who reported no impact stated feedback that suggested they would have wished for more training opportunities (see Section 4.6.9) rather than stating feedback that would be directly addressing the AR trainings, which was already the case for the qualitative feedback in their learning diary for Training 3 (see Section 4.6.4).

Arguably, this is in line with the impressions which can be gathered from their assessment during the curricular implementation. Additionally, it is also partially in line with the assessment of the didactic approaches by the lecturers, where they were slightly more satisfied with the approach for Training 3, closely followed by Training 2, and then Training 1.

Interestingly, while students retrospectively self-assessed the largest perceived competency increase after completing Training 3, this is also the training that, during the analysis of the OSCE results, had to be excluded because of abnormalities in the results of the control cohort, where the cohort performed significantly worse than the control cohort, which raised suspicions. If we visually compare the OSCE results and the students' retrospective competency self-assessments for each of the three trainings (see Figure 4.37 and Figure 4.39), they consistently are the highest for Training/Station 1, followed by Training/Station 2 and the lowest for Training/Station 3. That self-assessments of perceived competency on cohort level correlate with students' actual performance in the exam would be expected. Looking at the OSCE results of the control cohort, this pattern is consistent for Training 1 and 2 again and only deviated from for Station 3, where the control cohort performed substantially better and significantly more consistently. We believe, this provides indications that the included OSCE results do show valid differences between cohorts, and the exclusion of Training 3 was the correct choice for the analysis, as the difference was likely caused by rater variability. Nonetheless, it could still be true, that the cohort perceived their competency increase as the highest for Training 3 and subsequently performed in line with their expectations in the OSCE, but the control cohort was still significantly better in Station 3 compared to them and substantially better and more consistent compared to their relative own performance in the other stations.

When the students were asked to retrospectively assess the effectiveness of specific AR implementation choices and didactic delivery methods during the AR trainings 1, 2 and 3, they reported a slight preference for the trainings incorporating motor components and the interaction through decision trees, but there appears to be no correlation between their assessments that indicate individual preferences for one of the two AR interaction concepts. Their retrospective assessment of the effectiveness of specific didactic delivery methods revealed the error feedback and the less instructive "training mode" to be especially supporting for them. They assessed the error counter and timer of the training summary after each training to be least supporting.

Non-Representative Observations Not Visible in the Analysis

Besides all these results across evaluations, there are also observations we made during the evaluation stages, which partially influence our assessment of the usefulness of the Heb@AR App as a learning resource. While explicitly not representative, they are the following.

While this impression can partially be derived from the feedback for Training 5 and the slight preference for the Trainings with motor components in the summative feedback questionnaires, observations and non-representative feedback discussions after the curricular implementations of the AR trainings indicated that students perceive slightly more value in the trainings with decision trees and motor components, compared to the purely virtual Training@Home scenarios. Our interpretation of non-representative discussions with lecturers, which are not covered by the analysis of this thesis, would support this perspective. In this, it is especially unfortunate that the sample size of the evaluation of Training 5, the virtual version of the reanimation, is small, as this could have provided exploratory insights into whether this observed tendency is mainly due to the decision-tree-based interaction concepts or the incorporation of motor components.

During these non-representative feedback discussions, there was also important feedback specifically regarding Training 2, where students perceived the decision-tree interactions and putting the smartphone aside to perform the motor action as interrupting. In the feedback for Training 3, this discussion feedback was opposite. While the interaction metaphors slightly differed, e.g., Training 2 also incorporating video-based instructions for motor actions, we believe this perspective primarily caused by the perceived level of emergency. As the instructional design of the Decide-Freeze-Imitate interaction concept is comparatively behaviorist in its approach, students seem to appreciate this when the consequences are severe, as would be the case of the reanimation of a newborn, but perceive this as disruptive in training situations with less severe consequences for errors. Specifically regarding Training 3, the observations provide the impression that students did not appreciate the additional benefits of having to cooperatively make decisions on their own smartphones in a multi-user setting, compared to the added cost of having to wait for each trainee to make a decision, having less opportunity to train the motor components, and having increased setup times. It was additionally also observed that students used the individual training version of Training 3 cooperatively in teams already.

In general, we did not operationalize setup and preparation times for the trainings, but observations indicated that preparation for SkillsLab sessions was substantially more time- and resource consuming. As one of the midwifery researchers implemented Training 1 into a new curriculum when the Heb@AR App was already published as an OER, we have first insights on the BYOD methodology, that indicated that it was surprisingly quick and simple to deploy during the training session and demanded almost no support by the professor.

While theoretical models like the TAM [101], but also the accepted user experience principles of *usability + utility = usefulness*, e.g., as described by Nielson [347], theorize utility (in TAM called usefulness) and usability to be independent constructs, our non-representative observations in combination with the usability and utility results suggest that the perceived usability was at least partially influenced by the perceived utility. This was, e.g., visible in the usability score for Training 2, and Training 3 in the multi-user version. We do not believe this to be invalid measurements of the perceived usability construct, as the SUS was sensitive to perceived usability changes in formative and summative evaluations and consistent with the supplementary feedback provided.

4.7.4 The Usefulness of the Heb@AR App as a Learning Resource

Combining the results from all the evaluations and our observations, we can now address our exploratory research questions stated in the evaluation section: *Are there indications, that the five trainings of the Heb@AR elicited learning benefits? Were the Heb@AR Trainings usable by the target group of midwifery students?* and then subsequently discuss: **Is the Heb@AR App a useful learning resource?**

The results of the evaluations during the curricular implementations of Training 1, 2, and 3, clearly indicated utility for all three trainings in the form of perceived competency gains contextualized in self-efficacy theory. In retrospective self-assessments of these competency gains, they furthermore persisted, and the main effects suggest a general perceived competency increase. Main effects were also found, when comparing the OSCE results of the entire AR intervention cohort to a control cohort. Moreover, results for Training 4 and 5 both indicated an increase in intrinsic motivation to engage with the training content, with Training 5 also pointing toward a perceived competency gain contextualized in self-determination theory. As the qualitative feedback indicated that the majority of students furthermore not only perceived the app as preferable to their conventional learning methods but also an impact of the AR trainings on their practical study phase and the lecturers perceived utility in the app, we think there are clear indications that the Heb@AR elicited learning benefits. Consequently, we think it is appropriate to argue that the Heb@AR App provides perceived but also objective utility as a learning resource. Especially, as this is achieved while providing added practical utility, such as self-regulated learning, place and time independence for students, or reduced resource consumption in terms of rooms and tutors for practical training.

Any learning technology, however, can only prevail, if the use of the technology does not interfere with the learning. In terms of perceived usability, we think it is appropriate to assess, that the Heb@AR App was usable by the target group and did not negatively influence the delivery of the utility. Students were able to independently complete all the AR trainings, and they reported SUS scores that across all five trainings indicated at least marginally acceptable perceived usability. The only AR trainings, which elicited perceived usability which would not be described as "good", "excellent", or "best Imaginable" when contextualized on the adjective scales by Bangor et al. [28], are Training 2 and the multi-user version of Training 3 (see Figure 4.42 & Figure 4.43). They are both SkillsLab exercises based on the Decide-Freeze-Imitate concept, are the most comprehensive trainings of the App, and we suspect that the perceived usefulness influenced the perceived usability for these trainings at least partially. Interestingly, while the SUS samples for Training 2 and 3 (Decide-Freeze-Imitate concept) appear to elicit symmetric distributions of scores, Training 1 and 5 (TrainAR concept) SUS data is skewed toward higher perceived usability scores. We interpret this as a ceiling effect where the students were fully satisfied with the perceived usability, which is especially pronounced in Training 5. Finally, as desired, the perceived usability of the app for the lecturers did not significantly differ.

While the results on which this assessment was made are first explorations of the potential utility of the included AR trainings, they indicate the Heb@AR App to be an overall useful learning resource in the midwifery context. Based on studies on the acceptance of AR trainings, using TAM [101], which rely on intend to use or perceived usefulness, there was a high expectation that this would be the outcome. Although we do not directly apply the TAM by Davis et al. [101], these expectations could be met after actual use. Consequently, students and institutions should form a behavioral intention to utilize it when the usefulness is adequately conveyed. Through our perspective of the app being a useful "puzzle piece" in the vision of ARBTs, we believe it is useful regardless of comparative learning benefits to conventional methods or competing technologies.

Beyond the usefulness of the Heb@AR App as a learning resource in the midwifery context, we can also use the findings to discuss further practical implications, beyond the scope of midwifery. Most importantly, in the context of our vision of ARBTs, the indications of the Heb@AR App being a useful learning resource also serve as first indications that AR trainings based on immediately scalable interaction concepts and design decisions also elicit learning benefits in general. From our perspective, this is not self-evident, as implementation decisions and interaction metaphors specifically designed at task-level can be tailored towards the individual training's specific needs. In this, they might address specific aspects of the training more in-depth, but learnings and resulting trainings are more challenging to scale and transfer to new tasks or contexts.

4.7.5 Limitations of the Results

Despite these results being very promising regarding the usefulness of the Heb@AR App as a learning resource and contributing toward our understanding of the didactic benefits of realistically scalable handheld AR apps, they should be considered exploratory in nature. Thus, the results are considered preliminary, and we encourage replication. Generalizations should be made with caution and larger scale studies are needed.

Firstly, the selective variable analysis, where we retrospectively chose scale from larger evaluations during Project Heb@AR to answer the research questions from the perspective of the usefulness of the Heb@AR App, increases the risk of selection bias. Though, no scales were removed after the analysis, preliminary learnings and discussions during the project could have influenced the initial selection. Moreover, the research questions themselves were formulated in an exploratory, pragmatic manner to address the accepted user experience principle of ensuring an AR learning application's utility and usability to argue it to be a useful learning resource. While, this perspective is also, for example, found in the technology acceptance model by Davis [101], in this no hypothesis are generated post-hoc and the statistical analysis is therefore exploratory as well.

Furthermore, several measures, like the single-item competency assessments during curricular implementation, but also the retrospective pre-post designs are not considered universally valid measures, as they often correctly identify but not always correctly quantify differences of the constructs [46, 106]. While these methods are considered acceptable in exploratory work as they allow to "map out the main effects [...] rather than to identify detailed aspects of constructs and their interrelations" [106], they should not be interpreted individually or as generalizable effects and have to be subsequently proven with proper operationalization to substantialize our findings.

Moreover, the comparisons between cohorts only compare two cohorts and the retrospective self-assessment was only provided by a single cohort. The sample sizes for the curricular implementations are, in the context of educational science, still considered to be small to find actual differences and the sample sizes for the lecturers' assessment of the app and the evaluation study for Training 5 would not be considered sufficient to draw conclusions in the context of educational science. Further limitations of the results of the OSCE, like potential rater variability, team members rating OSCE stations and the statistical limitations of having to mix parametric and non-parametric tests, were already discussed in Section 4.6.8.

Furthermore, while we were able to show motivational benefits and increases in perceived competency consistently across evaluations, we applied the AR interventions to first-time users and only in a timeframe of about 2 years. While this is substantially longer than usually reported on in the literature, there could still be an uncontrolled novelty effect on motivation and perceived competency increases, and the benefits could decrease with familiarity during actual curricular usage over time [6].

While we did not operationalize any insights on gender effects, and it would not be possible to control for them in the context of midwifery education, it should be noted that results of the self-assessments, objective performances, but also the assessment of the Heb@AR App with its underlying concepts, could potentially be influenced by gender effects. Both, the students, but also the lecturers, were almost exclusively female. Generally, the male representation among midwifes is negligible. For illustration, in 2019, out of 10,005 midwifes working in clinical settings, only 52, therefore 0.005%, were male [120]. While the almost exclusive representation of female students in our evaluations at least contributes towards a more equal representation of genders in AR evaluation studies, as Merino [319] et al. for example found, only 28.10% of study participants to be female across 248 evaluations with 5,761 participants, the potential for gender-effects in our findings is a possibility. Besides the participants' gender, students inherently also had a homogenous educational level and general age group, which could have influenced the results.

4.7.6 Current & Future Work

Besides the obvious usage of the context, interaction concepts and learnings from the project to create an authoring tool that enables domain experts to create their own AR trainings, which will be discussed in the following chapters of this thesis, there are several aspects which should be addressed in future work.

Most importantly, we argue that motivational benefits and perceived competency gains alone justify the usefulness and effectiveness of the learning interventions. Additionally, based on previous work, we theorize and strongly expect comparative learning benefits to conventional methods from the interactivity of the AR trainings based on the more holistic learning approaches. In the end, the qualitative feedback provided by the students indicated that they also self-assess these comparative benefits. Nonetheless, this still has to be shown empirically. We support these efforts by providing Analytics functionality (see Section 4.4.9), which will be released with an update to the Heb@AR App that also includes a complete English language version.

Furthermore, we actively support the implementation of the app into new curricular. The app was, for example, already implemented at the Bielefeld University of Applied Sciences and Arts and University of Cologne, beyond the scope of the Heb@AR Project [381]. We believe these efforts will provide unique opportunities to gather "in-the-wild" usage data for AR training applications, that are currently missing in the literature. As we already have sophisticated utility to analyze perceived usability data in the SUS Analysis Toolkit, it is also possible to implement optional self-reporting features into the analytics functionality, e.g., to gather more usability data, or even qualitative feedback. This longer-term evaluation perspective is something we are currently preparing to explore from a technical perspective, but also by recruiting lectures for further collaborative evaluation efforts.

Finally, it would also be interesting to explore how well AR trainings, implemented with the two proposed interaction concepts and through our proposed methodology, perform compared to each other but also competing immersive technologies, like HMD-based AR or VR, in different contexts. Because of the interactivity and immersion, competing immersive technologies would likely outperform handheld AR concepts in terms of learning benefits. But when also taking scalability, self-determination, cost-factors, and setup-times into consideration, we expect there to be practical considerations, where there are learning benefits that are "sufficiently" elicited through the handheld AR concepts and the peripheral factors would make them the most suitable choice. This could lead to empirically developed recommendations on what immersive technology to utilize for individual scenario-specific or contextual expressions. In our perspective, these are practical considerations that have to be taken into account when trying to scale MR technology.

4.8 Summary

This chapter reported on the recent academization of the midwifery education and Project Heb@AR, where we tried to, at least partially, address the challenges for the practical training components of this new educational path, through the usage of scalable handheld AR trainings. This vision of ARBTs as one potential successor of WBTs is discussed beyond the scope of academic midwifery education, and a first step toward achieving it is made by contributing the Heb@AR App as an OER. Not only is this contributed app, to our knowledge, the largest handheld AR training app to date, but the development process is also transparently reported for other researchers and developers to learn from as a secondary contribution. Thirdly, two entirely novel interaction concepts for procedural handheld AR trainings are contributed, that make the transfer of procedural training tasks possible for the first time and are likely transferable to other trainings, or even contexts. Finally, exploratory evaluations through selective-variable analysis of data recorded during Project Heb@ARs overarching evaluation efforts and two supplementary evaluation studies for Training 4 and 5, provide clear indications that the Heb@AR App itself is a useful learning resource. Beyond that, these results are also a first indication that didactic benefits, generally observed in the literature for more scenario/context-specific prototypes, likely persist when moving toward realistically scalable hardware and concepts.

5 Design Space Exploration: Creating a Heb@AR Authoring Tool?

"Experience without theory is blind, but theory without experience is mere intellectual play" – Immanuel Kant

As can be seen in Figure 4.10 at the start of the Heb@AR chapter, the main menu of the Heb@AR App does have a "Create new AR training" button. Ideally, from a purely theoretical HCI perspective and in line with striving for consistent low-level interaction metaphors which we discussed throughout the last chapter, this button could lead to an interface, where midwifery lecturers and professors can simply create their own AR trainings in-situ, which can then be uploaded and instantly used by the midwifery students. But in actuality, it leads to a page explaining with its first sentence: "To start with: unfortunately, it's not quite that simple at the moment!", and links to an external GitHub repository for a Unity plugin. How we made this decision, and why an internal authoring solution directly in the Heb@AR App would be a naive design decision, will be discussed in this chapter, where we combine the learnings from the theoretical design space with the practical learnings of developing five AR trainings in the context of academic midwifery education the conventional way. In this, we explore the design space backwards, starting with the AR construct and the usage of it, and then designing and developing an authoring tool, which can create similar AR constructs to the ones we developed the conventional way. Some aspects of exploring the design space are therefore already implicitly discussed in detail in Chapter 4 and are only briefly readdressed here. We additionally also do not retrospectively reference specific literature from the design space in this chapter for the decisions, as we inherently explored the entire design space of AR authoring tools to construct it. But, as stated in Section 3.3.2, it should be noted that if others want to explore the design space of AR authoring tools, they should actively consult the literature map, to inform their decisions.

5.1 The 5w-1H-Guided Reflection

Where: In which context are Authoring Tool and AR Constructs used?

While the Heb@AR App itself is clearly contextualized in the academic midwifery education, with only some of its trainings, like the training regarding the German/Latin denomination of the female pelvis (see Section 4.4.7), arguably also having at least some value outside the midwifery context [52], is midwifery education actually the ideal context to address for the authoring tool?

With about 26000 active midwifes overall, as of 2019 [469], the midwifery context is generally comparatively narrow to directly address with the development of an AR authoring tool at this stage of the field of AR authoring tools. Additionally, many of the learnings during the development, but also the created ideas and concepts, appear to be transferable to other contexts, where procedural task learning is also of importance. For this, we broaden the scope and therefore also the context for the AR authoring tool we develop toward **procedural task training in educational contexts** in general.

Why: For what purpose are the AR constructs authored?

The purpose of the authored AR constructs, as described throughout the Heb@AR Chapter, is to provide scalable (see Section 4.1), self-regulated (see Section 4.1.3) opportunities for students to prepare for, or rehearse, practical procedures (see Section 4.4). Therefore, the purpose of the AR construct is to enhance the students' **learning** experience through added AR interactivity on ubiquity available hardware.

What is "Augmented" by the AR Constructs?

What would be augmented by the AR constructs, in our case, is dependent on the task, or rather which perspective we want to cover: The training@home perspective, or the supplementation of SkillsLab exercises, and therefore, inherently also the chosen interaction concept, as we developed two AR interaction concepts for procedural trainings in the context of project Heb@AR for these two perspectives: TrainAR and Decide-Freeze-Imitate, which are introduced in Section 4.3.5 and Section 4.3.5. For the TrainAR concept, using purely virtual AR interactions with objects and deliberately no physical objects for scalability purposes, the environment would be augmented in general, but there would be **no specific contextualized AR content** in the AR constructs, and they could be freely placed. For the Decide-Freeze-Imitate interaction concept, the answer would be highly task-dependent. In our cases in the midwifery context, **e.g., training dummies** would have to be augmented with instructional material by the AR constructs.

Who is the user of the created AR constructs?

The envisioned users of the created AR constructs are **students**, which use the AR constructs before, during, or after practical trainings of procedural tasks in the educational context.

Who is the author of the AR constructs?

Then, we finally have to reflect on whom the author of the AR constructs, in our case AR trainings, would be. While the most influential factor for the design of the AR authoring tool itself, this factor was not discussed before, but there are some learnings from the conventional development of AR trainings in the context of Project Heb@AR.

Generally, stemming from the non-representative observations and the specific workshops where we presented and worked with prototypes of the authoring tool we developed, but also the feedback provided through the other lecturer workshops (see Section 4.6.7), it appears that lecturers themselves might not be the ideal target group for an AR authoring tool for procedural task trainings on their own. While in the literature, teachers are often the intended construct authors, those AR authoring tools mainly author static AR constructs or simple visualizations. While it appears realistic that lecturers would be able to author similar constructs independently, the nature of the complexity of procedural task training makes this more challenging for our context. Especially in our case in the context of Project Heb@AR, the trainings included comprehensive, non-linear flows of logic, animated 3D models often unique to the context and tasks, and several interdisciplinary perspectives to consider. Additionally, as illustrated in the GANTT chart of development efforts (Figure 4.7), which provides a rough visualization of the conventional development process we conducted, the potential time investment required is on the order of magnitude of months rather than days. And it appears that the lecturers in our context in Project Heb@AR also reflected on this consideration, where they appeared to be eager to use the trainings, were able to use them as is apparent through the perceived usability results, but, if asked what support they would need, ask for technical support or even explicitly somebody who is hired at the institution to support the creation and usage of AR trainings.

Ultimately, as was already discussed by Hampshire et al. [184], there is a direct relationship between the abstraction of the AR authoring interface and the complexity of the created AR construct. Therefore, we can either lower the complexity of the created AR construct for the lecturers themselves to be able to author them independently, or we do not specifically target the lecturers themselves as the author of the AR trainings. While previous work explored these design decisions starting with the authoring tool, we deliberately explore the design space backwards.

Therefore, while being the primary-targeted author in the literature 2.18 and a logical choice, we explicitly do not target lecturers as the primary author of the AR trainings, but will try to maximize towards lowering the entry barrier of whom a potential author could be, that can create AR trainings of similar complexity as we developed conventionally in Project Heb@AR. Our target based on expectations and therefore inspiration for the exploration would be a **media technologist**, who is not a programmer but is, e.g., already using software like Adobe Photoshop, Premier, or other media tools. Alternatively, it could also be a domain expert who has significant media competency inherently from the context they are teaching in. Likely, the authoring process, similar to the conventional development, always incorporates several authors. In this, **lecturers might still be involved as an author**, even though they might not be the main author and therefore user of the authoring tool. One potential collaborative usage of the tool could, for example, be that lecturers, experts for the procedural tasks and trained to incorporate didactic considerations, author the procedures of the training, while the media technologists gather, create, and optimize the assets necessary to virtualize the training.

5.2 Exploration of Potential Design Decisions

Having now reflected on these five guiding questions, we can address how the AR authoring tool should be designed by exploring the design space. As previously stated, we explore the design space backwards, inspired by the conventionally created AR constructs in the context of the Heb@AR project and want to create an AR authoring tool which can create trainings similar to the ones, which we developed the conventional way in the project. The Heb@AR App merges two distinct AR training approaches: Training@home and AR SkillsLab exercises, each of which already has its own specialized interaction concept. We are therefore faced with the decision of determining which interaction concepts should be used by the AR authoring tools AR constructs. The perspectives of the interaction concepts are likely too different, to incorporate both through the same authoring tool (see Section 4.4). But, this decision automatically also accomplishes part of the design space exploration in our case.

For the AR authoring tool reported in this thesis, we decided to work backwards from the TrainAR interaction concept instead of the Decide-Freeze-Imitate concept. There are several reasons for this decision. Firstly, the Decide-Freeze-Imitate interaction concept incorporated motor components in the training task and, as discussed (see Section 4.7), appears to be perceived as providing slightly more utility to the trainee, compared to the TrainAR interaction concept. But this incorporation of motor components comes at several costs. The trainings require more setup-times, require more lecturer support, have slightly lower perceived usability, and still consume physical material, like medical consumables and training dummies. Additionally, the concept is inherently an interim solution, as the actions which are instructed by the AR app have to be physically performed, the smartphones have to constantly be put aside (see Section 4.6). Though this makes the concept immediately scalable today while still incorporating motor components in the training, which was the goal, the emergence of scalable HMD-technology would arguably displace the utility of this concept considerably sooner than the TrainAR concept, which offers location-independence and training@home opportunities which would remain useful after scalable HMD-devices emerge at institution level. While TrainAR, when assessing all evaluation efforts of Heb@AR combined, appears to be perceived as providing slightly less utility to the trainee comparatively, it clearly did provide the utility we aimed for with the concept: E.g., motivation benefits, increased perceived competency, and the possibility for self-determined, location-independent learning (see Section 4.7). It also generally elicited a slightly better perceived usability. Furthermore, TrainAR offers several advantages when considering peripheral aspects of realistic scalability, which we discussed in Section 4.1 and Section 4.7.1, especially with widening our scope of the context and potential users of the AR constructs/trainings. The TrainAR concept is arguably more easily applicable to a wider range of potential procedural tasks, it is more scalable as it does not rely on any external material or location beside the student's own smartphone, it requires less support and onboarding because of its comparative simplicity, and it has a more well-defined scope. This scope has several advantages when inquiring to develop and AR authoring tool for it. For one, it is arguably easier to develop or acquire the content for potential authors, as the content in the Decide-Freeze-Imitate concept heavily relies on complex contextualized 3D-animations, but secondly, it also sets a realistic, comprehensible scope of the authoring process. This, ultimately, also provides a realistic scope for the challenging endeavor of developing this AR authoring tool itself.

5.2.1 Content Type, Content Sequentiality, Construct User Hardware, User Interaction Concept, and Tracking Type

With choosing **TrainAR as the interaction concept** through which the AR constructs are created, several design decisions were already implicitly made. The content is therefore **sequential** to be able to implement procedural trainings, and incorporates **3D models, 2D sprites, textual content, and audio assets**. Because of the intended purpose of deploying the AR constructs as scalable learning opportunities for students at scale, **handheld AR devices** remain the envisioned user hardware and therefore also **markerless tracking** (ARCore [170] & ARKit [14]) the chosen tracking type (see Sections 4.3.5 & 4.4.1).

5.2.2 Authoring Hardware & Internal/External Authoring

For the authoring hardware, we decided to utilize **Desktop PCs**, which also automatically makes TrainAR an **external authoring tool**. For one, the trainings developed in the TrainAR interaction concept have their complexity in the flow of state, therefore the procedural component, rather than the spatial component. As this is not AR-specific, desktop approaches are the logical choice as they have more well-known forms of interaction to represent this type of complexity, compared to mobile approaches like HMD or handheld devices. Additionally, as can be seen in Figure 4.7, the development efforts were interdisciplinary and often took several months. Even if the development is substantially accelerated through the authoring tool, the authoring is likely a lengthy process rather than something that can be completed impromptu. This would make the authoring in mobile hardware likely quite challenging. Especially the asset generation and incorporation would be a challenge and the incorporation of textual content would be time-consuming. Moreover, we envision the authoring process to also incorporate several authors at times, and desktop approaches provide the best pre-existing collaboration utility.

5.2.3 Authoring Tool Modularity & Authoring Interaction Concept

While only sparsely explored for AR authoring tools in recent years (see Section 2.21), we want to implement the AR authoring tool as a **plugin solution**, rather than a standalone (or even web-based) solution. We are aware that this substantially increases the entrance barrier of using the authoring tool, but with a media technologist being our primary envisioned author and therefore user of the authoring tool, this offers two advantages.

Firstly, with our decision of using Unity as the host software of our plugin, one of the primary development platforms for AR and VR applications as of today, if implemented correctly, the scope provided by the plugin/framework is always just an initially enabling factor, but never limits the author in their realization of ideas. Therefore, if the authors have requirements, which are not covered by TrainAR itself, they can still use the full Unity functionality by programming parts themselves, if they chose so. This would be substantially harder to technically implement in standalone solutions and even then, would have to be separately documented and explained to potential authors.

Secondly, this decision allows us to implement our primary envisioned interaction concept for the authoring process: **traditional UI Interactions in combination with visual scripting**. Through this interaction concept in the authoring tool of TrainAR, authors are enabled to create complex logic trees and non-linear stateflows, while still not having to program them. The visual scripting implementation we develop is therefore also deliberately inspired by the stateflow descriptions which we developed during Project Heb@AR (see Section 4.3) for the transfer of task-process analyses of training procedures towards AR trainings, to simplify the transfer and provide a description of the trainings flow that is comprehensible to both; the media technologist but also other stakeholders or collaborators like teachers/lecturers who might be involved in the authoring.

5.2.4 Authoring Contextualization & Authoring Preview

The complexity of authoring the TrainAR trainings is primarily creating the flow of the states, therefore the training's logic, and acquiring and converting the training's assets rather than the contextualization of AR content. Therefore, we decide that content can be authored **decontextualized**, though a **3D preview** should be provided to the author.

5.2.5 Markup Notation & Distribution

While we decided that the first implementation of the TrainAR authoring tool **does not use a markup notation**, but distributes the trainings **locally as binaries**, this is more a shift of focus rather than a deliberate decision because of the context, author, or user. As will be discussed in Chapter 6, while the initial open-source version of TrainAR does not include them, we are currently actively working on distribution aspects and through the usage of Unity as the host software, many potential technical hurdles are already taken into account (see Section 6.6.6).

5.2.6 Additional Decisions & Novel Perspectives

Finally, we also want to specifically address several gaps in the literature with additional design decisions in line with the gaps that were identified in the scoping review (see Section 2.7). As can be seen in the literature map, only very few tools evaluate both perspectives (authoring and the usage), very few tools are published as open-source, and not a single tool is both, fully evaluated, and published as open-source. Additionally, many publications on AR authoring tools do not explicitly state or evaluate who the author and the user of the AR constructs might be and not a single tool provides a full documentation for authors to utilize it. Therefore, beside the design decisions stated previously, we deliberately address all these aspects throughout the development of our AR authoring tool: TrainAR, which is described and evaluated in the following chapters.

6 The TrainAR Framework

"Consistency is one of the most powerful usability principles: when things always behave the same, users don't have to worry about what will happen. Instead, they know what will happen based on earlier experience." — Jakob Nielsen

Figure 6.1: TrainAR Framework: The TrainAR authoring tool (left), that is used to create TrainAR trainings (right) using the TrainAR interaction concept and didactic framework.

Through the contextualization of digital content and information directly into physical reality, Augmented Reality (AR) provides a powerful set of possibilities for training and learning purposes. The added benefits of using AR in education are generally well known, and studies indicate that the application of AR as an additional "multimedia source" in existing curricula can already lead to improved retention, attention, and satisfaction [413]. A meta-analysis shows increased academic achievement with AR compared to traditional learning methods, along with increased concentration, and it also indicates that it enables teachers to convey concepts faster and with greater clarity through the demonstration of connections between concepts and principles [359]. Additionally, secondary literature furthermore points towards a consistently positive impact of AR tools used in educational settings [438], especially through interaction, catching the learner's attention, and increasing motivation [387]. In particular, although significant differences can be observed for all levels of education, the largest effect size of learning benefits can generally be observed for students at the undergraduate level [359]. Therefore, while user attitudes towards AR are influenced by the perceived usefulness and perceived enjoyment of the user, the findings indicate that perceived enjoyment is a more significant factor than perceived usefulness regarding the intention of using AR as a learning source [507].

Current Challenges in Educational AR

If those benefits are already well known, why is there no widespread adoption of AR in educational settings? While the answer may at least partially be found in the generally hesitant adoption of information and communication technology in educational contexts because of rigid structures, restrictive curricula, and teachers (including university lecturers) lacking relevant pedagogical training [68, 510], there is also the problem of the availability of a suitable, scalable AR that can be directly applied to aid their educational goals. While AR hardware in the form of head-mounted devices (HMD), which is increasingly available at lower costs, and commodity handheld/mobile AR hardware, which already comes with solid marker-less tracking techniques, would, in theory, allow for a realistic deployment of AR into teaching curricula today, this currently is not true for the AR software complements. Currently, most AR applications are rather narrow in their scope, often focusing on specific subjects and showing learning benefits with predefined prototypes [510], while teachers would primarily need approaches with common concepts, where they would like to be involved in the development process [466]. Therefore, as teachers rarely have relevant AR programming expertise [408], the development of new AR content should be made as easy as possible for them and ideally not involve any scripting or programming [96, 408].

The Challenge of Creating AR Content

This process of AR content creation is generally referred to as "authoring," and it is one of the biggest general challenges in AR today. Even outside of education 10 years ago, Schmalstieg, Langlotz, and Billinghurst [418] already recognized authoring to be one of the five big challenges holding back the widespread adoption of AR. While most of their proposed challenges, such as low-cost platforms, mobility, and suitable back-end infrastructure, are already solved on mobile AR devices utilizing Android and iOS, even today, AR content authoring remains a challenge. There are multiple reasons why this is the case. Firstly, there is an inherently complex need for a tradeoff between the fidelity of possibly created AR applications and the required technical expertise [184]; in other words, either you can create powerful AR scenarios with complex interactions or it can be used with very little technical expertise. Finding the "sweet spot" where non-experts in the technical domain can create complex AR scenarios on their own is still a research question to be answered, so today's AR applications are either built by programmers with significant expertise or authored AR content by non-technical domain experts offering little-to-no interactions [340]. Secondly, AR App development is difficult. A study conducted by Ashtari et al. [20] found that even programmers, ranging from hobbyists to professionals, face consistent challenges with AR development. They report a current lack of concrete design guidelines and examples and incorporation of novel interaction metaphors such as physical aspects to be challenging, and they complain about many unknowns in terms of development, testing, debugging, and user evaluations.

Lastly, developing authoring tools as a research topic is a thankless endeavor, or as Nebeling [339] described it for toolkit research in general, a "tricky game". Toolkits are hard to develop, tackle multiple challenges from different disciplines at once, and, in the end, are often hard to publish based on their perceived lack of novelty compared to the time investment necessary to make them realistically applicable for further use in research applications or even curricular usage. This lack of recognition of software artifacts is also a topic of ongoing debate in the educational technology research area [229].

As we therefore believe that only usable, scalable, and self-sufficient comprehensive concepts for AR content can realistically enable AR usage in education, we propose TrainAR, a threefold combination of an interaction concept for procedural task training on handheld AR devices, a didactic framework discussing its implementation as a multimedia source in existing curricula; an AR authoring tool (see Figure 6.2) enabling both technical but also nontechnical domain experts to create AR trainings. The TrainAR authoring tool is an open-source extension for Unity that uses visual scripting as its main authoring interaction, while still allowing the usage of the full C# functionality on demand. Therefore, TrainAR would be classified as a high-level programming tool but also a low-level content design tool based on the digital media authoring taxonomy proposed by Hampshire et al. [184] and allows non-programmers with significant media competency and programmers to utilize it.

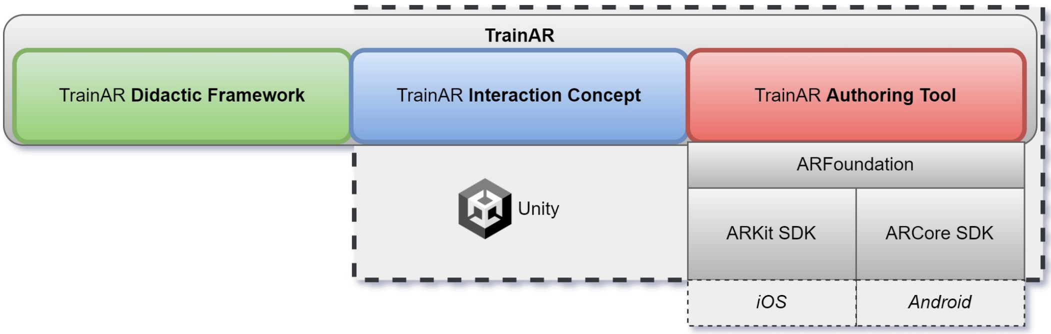


Figure 6.2: TrainAR: Didactic framework, interaction concept, and authoring tool. This chapter discusses the Unity-based authoring tool that is built based on AR Foundation, allowing the deployment to both Android and iOS devices through the ARCore and ARKit SDKs, respectively.

This chapter is structured as follows: First, the TrainAR interaction concept is defined in Section 6.1. Then, the concept is described from the perspective of an already developed TrainAR training in the project Heb@AR context (the preparation of a tocolytic injection) in Section 6.2. Afterward, Section 6.3 describes formative evaluations of this implemented example and iterative design improvements that lead to the final TrainAR interaction concept. Section 6.4 provides a didactic framework of TrainAR and supporting considerations to transfer procedural training tasks towards TrainAR trainings. Finally, Section 6.5 describes TrainARs components from the technical perspective before the TrainAR authoring tool is introduced in Section 6.6. Finally, the current state of TrainAR is discussed in Section 6.7, before Section 6.8 concludes the chapter.

6.1 The TrainAR Interaction Concept

The goal of the proposed handheld AR interaction concept is to be scalable, stable, and easy to use in the context of procedural training tasks by users with varying levels of media competency. Therefore, it improves upon traditional AR interaction metaphors from both the literature and common non-AR applications and combines them for non-linear procedural interaction chains, creating a more holistic interaction concept. The proposed concept is mainly targeted at currently available consumer-grade Android & iOS smartphones but, with little changes, can also be adopted to tablets.

While traditional AR augments physical objects or structures with virtual computer-generated content, this interaction concept targets purely virtual procedural training through handheld AR devices. While it is true that, for example, in assistance scenarios a direct in-situ contextualization of instructions is beneficial [59], studies have shown that for training scenarios, tangibility has no significant effect on learning outcomes [236] but introduces limitations for the scalability and prohibit the possibility for training-at-home usage.

Therefore, the concept is deliberately purely virtual and pragmatic in both, its interaction metaphors and its design. On an abstract level, it consists of 5 interlined ideas: A **Virtual Training Assembly** representing the training setup and objects that are used for the trained tasks, **Adaptive Instructions** that are provided to the trainee, **User Actions** that are triggered by the trainee and **Layered Feedback** that provides feedback to the trainee by matching actions with instructions. Furthermore, **Insights** provide supplementary declarative knowledge and contextual framing in relation to objects or procedures found in the training task (see Figure 6.3).

The first two procedural AR trainings derived from this interaction concept and affiliated didactic framework are shown in Figure 6.4, one was developed in the context of the BMBF project Heb@AR and the other in the context of the EU project CHARMING [114]. The AR training of preparing a tocolytic injection, using the TrainAR interaction concept, is used to describe a TrainAR training from the perspective of an authored training in detail in Section 6.2 and to report the formative evaluation efforts of the TrainAR interaction concept in Section 6.3.

6.1.1 Virtual Training Assembly

The virtual training assembly consists of virtual 3D models of all objects relevant during the training. Besides the objects needed to complete a procedural task, this may also cover so-called distractors, objects that are needed to construct situations for decisions, as well as hidden objects in the form of, e.g., trigger areas, that can be used to check if an object was placed at a specific location.

When starting a training application based on this interaction concept, users are explained the context of the training (contextual onboarding) and explained how to use the application and conduct the training (technical onboarding). They are then guided through a setup onboarding

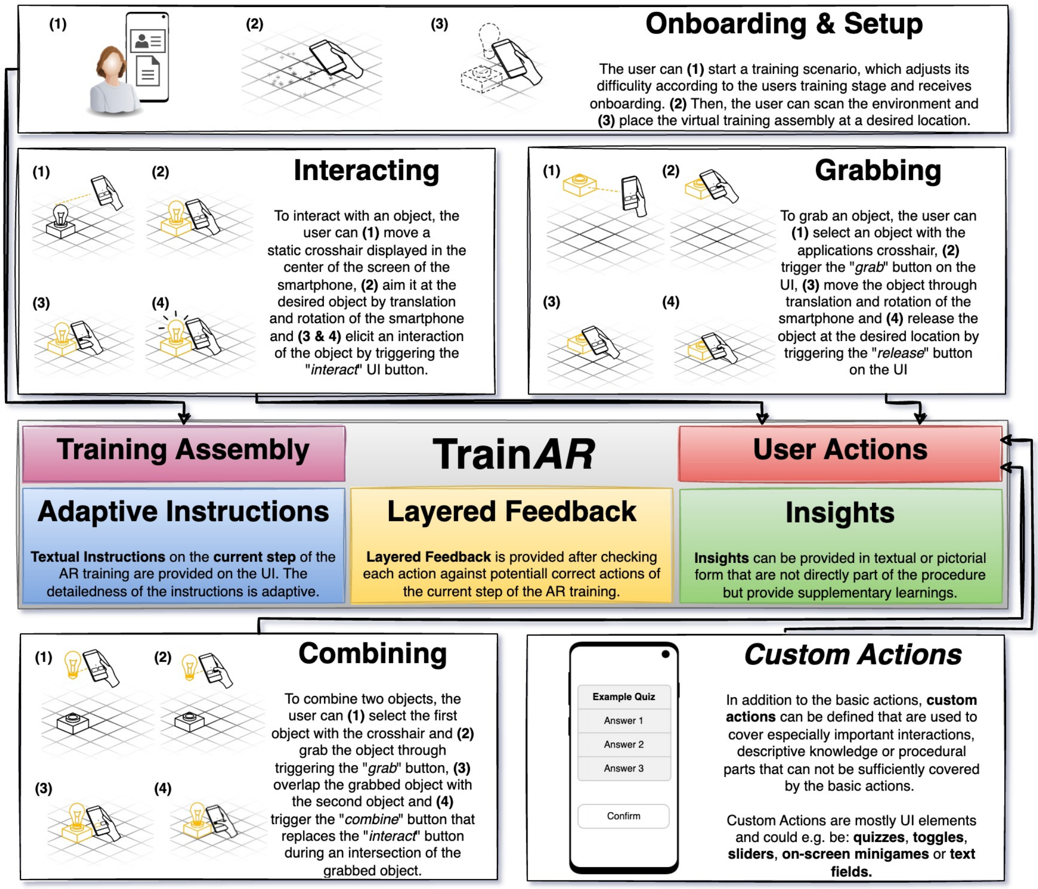


Figure 6.3: The TrainAR interaction concept: After the **Onboarding & Setup** of the **Training Assembly**, the internal state process model provides **Adaptive Instructions** to the user of the application. Additionally, it can provide **Insights** and hints if needed. The user can trigger **User Actions** with virtual objects **Interacting** with them, **Grabbing** them, **Combining** objects or trigger **Custom Actions** (e.g., quizes or mini-games). The internal state process model checks those actions against an internal state catalog, advances the procedural task, and provides **Layered Feedback** to the user.

process that explains the process to establish a frame of reference by scanning for visual feature points. When completed, users can place the virtual training assembly and the training starts (see Figure 6.3, Setup).

6.1.2 Adaptive Instructions

During the training, the state process model provides continuous instructions to the trainee, detailing the next steps to complete the training tasks (see Figure 6.3, Adaptive Instructions). These

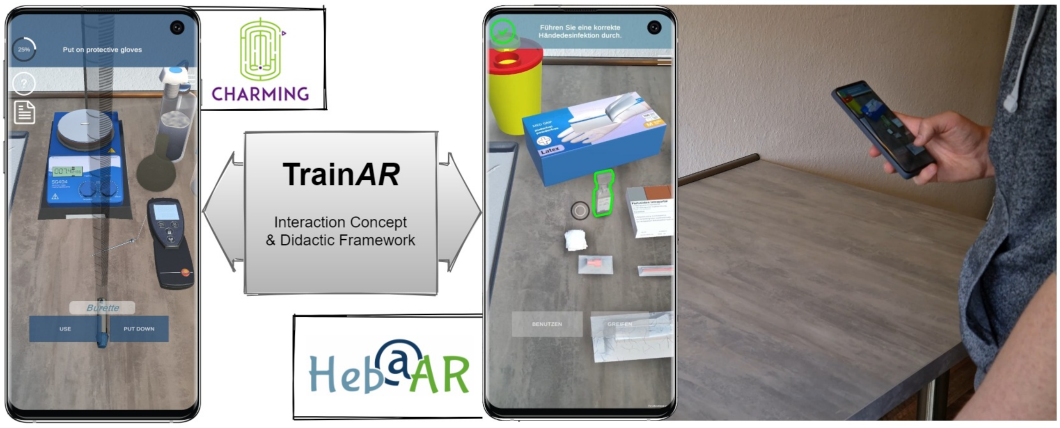


Figure 6.4: TrainARs first two trainings, which are fully evaluated: The conduction of a titration experiment in the context of chemical education in project CHARMING (**left**) and the preparing of a tocolytic injection in the midwifery context of project Heb@AR (**right**).

instructions are provided through the UI, e.g., in the form of text at the top of the smartphone screen. These instructions are adaptive regarding two orthogonal perspectives: Firstly, different sets of instructions can be created to support distinctive levels of difficulty, relevant to support multiple didactic contextualization stages as well as to increase replayability. Empty sets of instructions are also supported to be used for summative training assessments or exams. Secondly, instructions can be adaptive regarding the sequence of actions chosen by the users, creating a non-linear training experience. The concept works for both, strictly linear procedures which would display instructions with specific solutions to a current step of the linear procedure, but also rule-based instructions, where more than one linear path would be correct and specific actions trigger state changes and the state process model checks those against a necessary procedure list.

6.1.3 User Actions

To trigger actions during the training, the user of the application can use four basic actions provided by the interaction concept. Additionally, quizzes, sliders, or toggles on the UI based on implementations of a “custom action” can be utilized to implement actions that are especially important and should be highlighted or cannot be sufficiently covered by the four basic actions (see Figure 6.3).

The user can **select** and **deselect** an item by using a crosshair in the middle of the smartphone screen and aiming it directly at a virtual object. The user can then **interact** with selected objects by clicking the interaction button (see Figure 6.3, Interacting), triggering a state check with the state process model and corresponding feedback. Alternatively, the user can **grab** a selected object, which automatically lerps the virtual object to a position relative to the front of the smartphone while retaining a static vertical rotation towards the training assembly. This allows the user to manipulate the object's position and rotation by then releasing the object at a different location (see Figure 6.3, Grabbing). Positional changes can but do not necessarily have to be checked against the internal state catalog of the state process model. Grabbed objects can also be **combined** with secondary objects by overlapping the grabbed object with the secondary one and triggering the combine button, which replaces the interact button when two objects overlap. This combining of objects is then validated against an internal state catalog of the state process model and if allowed, the objects are combined into a single object (see Figure 6.3, Combining). Additionally, grabbed objects can also directly be interacted with using the interaction button.

6.1.4 Layered Feedback

When the user triggers actions to follow the provided instructions by the state process model, those actions, when checked against an internal catalog of potential actions, can either be ignored, correct or incorrect. Ignored actions are not processed by the state process model at all and do not elicit any feedback. This can for example be used if selection/deselection of objects or grabbing/releasing objects to move them around does not have implications in the training task the interaction concept is used in. Correct actions always trigger visual feedback, e.g., in the form of a green blinking outline of the object and auditory feedback either representing the sound of the interaction itself or, if not applicable, a short sound that implies positive feedback/success. While correct interactions additionally trigger their internal event (e.g., an animation or additional visual information) and correct combinations combine the two overlapping objects, wrong interaction potentially need to elicit feedback to the user beyond simple visual and auditory error feedback. As incorrect actions can vary in severity and too much feedback could be annoying to the user, a layered feedback system is used. Here, basic interactions that are not severe only elicit the normal short error feedback comparable to the feedback of correct actions in the form of a visual error symbol on the UI, a blinking red outline of the virtual object and an error sound. If an error is detrimental and should always immediately be corrected or repeated errors of the same step are triggered, more intrusive feedback is given by overlaying the whole screen of the application with textual and pictorial explanations, containing hints for the user to complete the task they are struggling with.

In line with the provided instructions, feedback can be adaptive, therefore both, very behaviorist approaches are possible where wrong actions are immediately corrected but also constructivist approaches can be deployed where the user is incentivized to explore and incorrect interactions as parts of overarching procedures are not immediately prohibited, but only the result is checked with the state process model.

6.1.5 Insights

Besides the procedure itself that is trained, there might be insights that a trainer wants to provide, that cannot be contextualized in the procedure itself or are not part of the procedural component but are rather supplementary insights or information which cannot be visualized in context in physical training but could add extra learnings for the trainee. Those could for example be contextualized visualizations of declarative knowledge bits which were learned from theory, or additional hints and insights from experienced trainers from practice.

6.2 Exemplary Implementation: A TrainAR Midwifery Training

To date, several trainings using the proposed interaction concept are either fully developed and evaluated or are currently in development. At the time of writing this thesis, the implementation of the titration experiment in the context of chemical engineering education (MAR Lab) [114] and the implementation of preparing a tocolytic injection as part of the practical training of academic midwifes in the context of project Heb@AR [54] are both fully developed and evaluated. As the tocolytic injection training is the not only the oldest TrainAR training but also the training used in the formative evaluation phases to establish the interaction concept, it is chosen to demonstrate the interaction concept from the view of an exemplary implementation in this section.

The TrainAR training, reported here, describes the iteration after the conduction and subsequent improvements of all formative usability studies that are reported in Section 6.3. Notably, the PhD thesis of Jessica Lizeth Domínguez Alfaro [113], the main developer of the MAR Lab application using the TrainAR interaction concept, also includes exemplary descriptions of her TrainAR implementation and formative evaluations and summative evaluations of the learning success, that might be interesting to the reader for a more independent perspective and descriptions.

As already described in more detail in Chapter 4, midwifery education is currently transitioning towards a full academization in Germany, where midwives will soon be exclusively qualified at universities, rather than by vocational training through the dual education system. While this is an important step towards increasing the status of midwives in the medical context, it also leads to new challenges. The practical component of the training still has a high priority and this naturally leads to bottlenecks regarding available practical tutors, training space and scheduling restrictions for trainees. Preparing a tocolytic injection, an injection used for inhibiting labor contractions that is for example administered in preparation of a C-section, is a relatively basic procedural task that every midwife has to be proficient in, which made it an ideal candidate for the first implementation of TrainAR.

6.2.1 General Considerations and Design Decisions

The implementation combines clean UI elements in healthcare-inspired color palettes with realistic high-resolution 3D models and comic/drawn stylized visualizations for conceptual contextualization and feedback mechanisms. The clean UI elements and healthcare-inspired color palettes were chosen, so they provide (mostly textual) instructions and feedback with as little distractions as possible, while eliciting a sense of trust and familiarity in the user. The high-resolution 3D models were designed as realistic as possible on mobile devices to be visually recognizable as their physical counterparts. Finally, the conceptual contextualization and feedback mechanisms like the summative assessment of the training or additional practical insights were implemented in stylized comic/drawn form to elicit some sense of play and gamification in the user while retaining the seriousness of the context and purpose of the procedure.

The procedure of preparing a tocolytic injection starts with strict hygiene procedures, preparing the workspace according to protocol and then starting the preparation of the tocolytic injection by selecting and opening all necessary material. Then, a syringe has to be connected with a needle and a carrier solution and tocolytic medication has to be drawn up in correct order and quantity. Afterward, the needle has to be disposed of according to procedure and the syringe has to be sealed using a luer lock and labelled. Afterward, all remaining utensils have to be disposed of. The virtual assembly for the training contains all the objects necessary to perform this procedure and additionally several distractors, like medication that is out of date and a needle which would not be used to draw the syringe with a solution in this context.

6.2.2 Conceptual & Technical Onboarding

When starting the AR training, the users first receive conceptual onboarding. Specifically, in this case, they are told that they start a shift in a midwifery ward and during a routine examination realize that the prepared tocolytic injection is expired, and a new one has to be prepared (see Figure 6.5a). Trainees can then decide to receive technical onboarding, explaining how to use the application, before starting with the scenario. They can also opt in to receive insights in the form of practical know-how by an experienced midwife called "Agneta Reuter" during the training (see Figure 6.5b). Trainees are then shown 3 sets of animations and textual instructions on how to interact with virtual objects during the training (see Figure 6.5c–e), before they are transitioned into the AR context and instructed to scan the environment (see Figure 6.5f). When a sufficient amount of feature points are detected by the tracking algorithm, users can position the virtual assembly setup into the physical environment through translation and movement of the smartphone and confirm the position by an on-screen touch (see Figure 6.5g).

6.2.3 Instructions

Besides the instructions given during the onboarding (see Figure 6.5a–g), instructions are continuously provided in textual form on top of the smartphone screen. Additionally, a progression circle with a percentage number is displayed in the top-left corner, showing the Trainees' progress through the procedure of preparing the injection. In the current implementation, 3 levels of instructions are implemented. The first one are step-by-step instructions, which guide the user through the training with explicit instructions on what to do for each step. The second one only guides the user through stages of the training, such as starting with the hygiene requirements, the preparation phase and the actual preparation of the injection itself (see e.g., Figure 6.5h, top). The third level is to provide no initial instructions on the top UI element at all, though the progress circle, error feedback and reinforcement of correct interactions are still provided.

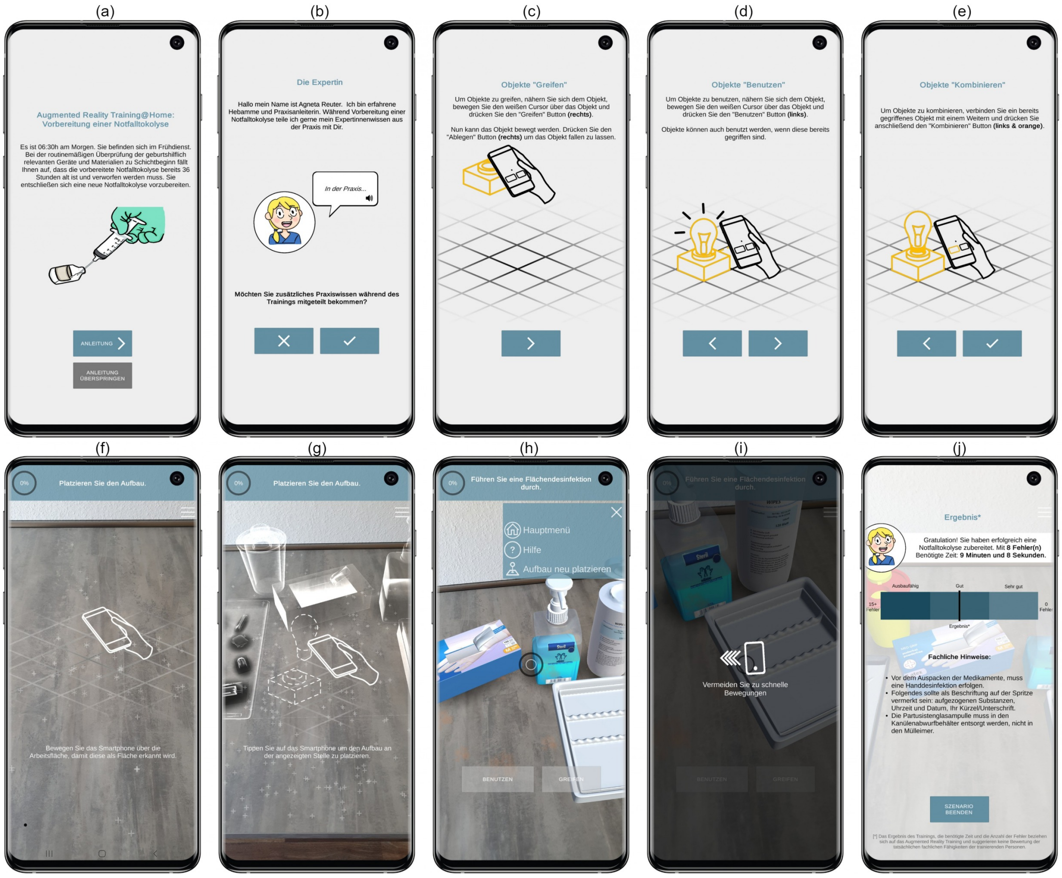


Figure 6.5: UI components implemented for midwifery AR training scenario. **Top, a-e:** The scenario and interaction onboarding. **Bottom, f-j:** The onboarding of placing the training assembly into the room, menus for rewatching onboarding tutorials, replacing the assembly or exiting the scenario, exemplary warning for AR tracking problems (Others warn of problems with too little illumination or insufficient feature) and the End-screen of the scenario providing contextualized performance feedback and an additional training assessment with professional feedback.

6.2.4 User Actions

The basic user actions were implemented closely following the proposed interaction concept in Section 6.1. A crosshair is used for the selection of objects, with visual feedback if a target object is in range (see Figure 6.6a,b). Selected objects have an orange outline and subtle shading to visualize the selection (see Figure 6.6a). Selected objects can be interacted with, grabbed, released, and combined with other objects by pressing the corresponding buttons. Those buttons either display the name of the generic action, like “interact” or “grab”, as well as object-specific derivations of that action, e.g., displaying “open” instead of “interaction” for opening packaging (see Figure 6.6g). If no interaction is currently possible, the buttons are greyed out (see Figure 6.6b). Grabbed objects are no longer outlined and shaded (see Figure 6.6c). If the user of the application overlaps the grabbed object with a second object, this object is outlined and shaded while the grabbed object is made transparent. Additionally, the interaction button changes to a combining button and changes its color to visualize the combining state as explicit as possible (see Figure 6.6d).

Figure 6.6: Interactions with virtual objects in the midwifery AR training scenario (Displayed actions do not represent a sequence of events. They visualize specific functionalities described in Section 6.2). **Top, a-e:** Selection of an object with context triggered insights, no selection, a grabbed object, an object before "combining" and a scenario-specific custom action of drawing the syringe. **Bottom, f-j:** Positive feedback for an interaction, positive feedback with additional feedback, negative feedback for an error, an overlay for severe or repeated errors and an example of a custom action in the form of a quiz.

Two custom actions were utilized in this AR training scenario. One custom action of using a UI Slider to conceptually imitate drawing up a syringe was used twice in the AR training, once for drawing up the carrier solution and then successively the medication (see Figure 6.6e). The other custom action was used for the labeling of the prepared injection so that users of the application do not have to type out the full label with name, date, time, carrier solution, medication, and signature but the knowledge of what inscriptions are necessary can still be quizzed. (see Figure 6.6j)

6.2.5 Guidance & Feedback

Grabbing and releasing objects, combining them, or triggering their internal interaction always triggers visual feedback in the form of the animated blinking outlines in either green or red on the virtual AR object itself. It also displays a success or error icon on the UI, momentarily replacing the progress bar. Additionally, all actions either play an object-specific sound, such as the ripping of packages or liquid sounds for drawing up the medication, or can play a default success sounds as feedback for correct actions. Error sounds are hereby always played on incorrect actions, regardless of sounds set by the author of the training. Protruding green and red colors, not in line with the utilized color palette, were chosen to make the feedback prominent to the user (see Figure 6.6f,h).

Some errors in the medical context, like actions that endanger sterility or switching up the sequence of drawn up solutions, have severe implications. Subsequently, for some steps, a standard error is not sufficient and the severe layer of error feedback is provided instantly by displaying a white UI overlay, temporarily taking the users out of the scenario and focusing them on this specific feedback. This modality is also used to provide specific feedback with additional guidance if users repeatedly trigger incorrect actions, implying they need additional help (see Figure 6.6i).

Furthermore, some interaction, like disinfecting the hands or putting on gloves, are not exhaustively covered through basic interactions, as they would not have been implementable in a satisfactory manner and would have distracted from the core learning goal of the AR training. Therefore, they are only covered by a basic interaction on their object and a UI element that informs the user that this action happened (see Figure 6.6g).

In the event of tracking problems, the current AR training is paused, and a black screen overlay is displayed guiding users through possible steps to resolve the tracking problems, e.g., instructing them to move the handheld device more slowly, ensure sufficient light in the environment or trying to track a different surface as not enough feature points could be detected (see Figure 6.5i). If tracking problems persist, users can also re-position the virtual training assembly entirely, restarting the placement onboarding (see Figure 6.5h,f, g).

6.2.6 Professional Midwife Insights

Insights in the training scenario of preparing a tocolytic injection were implemented in the form of a professional midwife called “Agneta Reuter”, which provides anecdotal knowledge from practice as well as hints and contextualized advice at specific moments in the training procedure (see Figure 6.6a,e). When triggered, an audio file is played and a short version of the insight is displayed on the UI right under the instructions. Users can decide, if they want to use these supplementary insights, at the start of the training (see Figure 6.5b).

6.2.7 Training Assessment

After the AR training is concluded, a training assessment screen is shown to users (see Figure 6.5j). Here, AR training specific feedback and measurements are provided, such as how fast the training was concluded and how many incorrect actions were triggered. The amount of incorrect triggered actions is also contextualized on a feedback graph to make results comparable. This graph deliberately does not use traffic light colors but rather shades of blue, to not discourage trainees, e.g., if they were in the yellow or red in early iterations. In line with this endeavor, users are also informed that the assessment measures are AR training specific and do not imply assessment of their professional performance. Additionally, "professional notes" are provided that are displayed when specific actions were triggered which suggest that users were not following the correct procedure. E.g., this could be trying to use a carrier solution which is out of date, placing a used syringe onto the work area, or trying to throw away the medication before the syringe is labelled.

6.3 Formative Evaluation & Iterative Improvements

To evaluate the usage of the interaction concept, three formative usability studies were conducted iteratively, using the Android version on the Samsung Galaxy S10 (SM-G973F). Here, the TrainAR training described in the preceding section was used, and the evaluations were done within the midwifery context.

The focus of the first study was on gesture-based interactions, as suggested by related work, and textual as well as pictorial onboarding. Other elements, such as instructions and error feedback (see Section 6.1) were also realized but not in focus. The second study improved upon the onboarding and introduced the training assessment at the end of the training. Additionally, it implemented an alternative, more explicit interaction concept based on buttons, subsequently referred to as the "explicit" interaction concept. The differences between the two types of interaction are visualized in Figure 6.7 for the combination of two objects. The third study only provided the explicit interaction concept with further refinements to the explicitness of the interaction feedback. Furthermore, improvements to the training assessment, the technical handling of AR tracking and feedback thereof, and user actions were made.

For all studies, a task-based research methodology was used, where participants were given a context in which the training task would have to be completed and were encouraged to "think aloud" during the experiment. Participants did not receive external help during the experiment. After completing the task, participants were asked to fill out a System Usability Scale (SUS) questionnaire [65] and a user experience questionnaire (UEQ) [257], so pragmatic and hedonic qualities were evaluated. Finally, participants were asked to fill out a qualitative questionnaire, asking what they liked or did not like about the application, what they had problems with and additionally provided the opportunity for further feedback or remarks.

6.3.1 Participants

Overall, 24 participants (16 unique participants), aged between 21 and 46, with an average age of 28.75 (SD = 6.16), took part in the evaluations. All participants were either midwifery or nursing students that were familiar with the preparation of a tocolytic injection. 15 out of 16 participants were female. To gather both iterative feedback across the developed versions, including increased familiarity with the application, as well as fresh feedback and "first impressions", some participants were deliberately invited to multiple studies, while others only conducted the experiment once. In the first study 6 students participated, in the second 10 participated with 5 per condition for the explicit and implicit interaction concept and in the third study 7 students participated. Across the studies, 9 participants took part in one, 5 participants in two (3 participated in 1 & 2, the other 2 in 1 & 3) and the remaining two participants took part in all 3 studies. While primarily aimed to be a qualitative study and therefore marginally sufficient in sample-size, it should be noted that these numbers were impacted by the pandemic situation at the time of conducting the formative evaluations in 2021.

6.3.2 Results: Usability

Regarding usability, participants reported an average SUS score of 64.58 (SD = 7.81) in the first study. In the second study, participants reported an average SUS score of 63 (SD = 8.91, Mdn 67.5) for the implicit interaction concept and an average SUS score of 81 (SD = 3.35, Mdn = 82.5) for the explicit interaction concept. This difference is highly significant according to a Mann-Whitney U-Test (U=0, p=0,008, r= 0,83). For the final study, improving upon the explicit interaction concept, participants reported an average SUS score of 80 (SD = 7.91). According to Bangor et al. [27], SUS scores of 63 and 64 would be considered "ok" and thus represent a (low) marginally acceptable usability, while SUS sores of 80 and 81 are considered to be "good" to almost "excellent" and imply acceptable usability. Additionally, achieving a SUS study score of above 80 is commonly used as a non-empirical "Industry benchmark" [278] to surpass.

It has to be noted that those SUS scores are not conclusive with the number of participants used in the separate usability studies according to Tullis et al. [463], though the sample of participants was fairly homogeneous, and the observable variance is small (see Figure 6.8).

Notably, in line with the average SUS scores, the two participants who took part in all three usability studies both reported a SUS score of 72.5 in the first study. The participant in the "implicit" interaction concept condition in the second study reported a SUS score of 70, the participant in the "explicit" interaction concept condition a SUS score of 82.5. In the third usability study, the participants reported SUS scores of 82.5 and 90 respectively. All results analyzed with the SUS Analysis Toolkit are included as tables in Appendix 28.

6.3.3 Results: User Experience

For the user experience, the reported UEQ results were analyzed for the 6 measures: attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty using the UEQ benchmark, which contextualized the measured scale means in relation to a benchmark data set of over 450

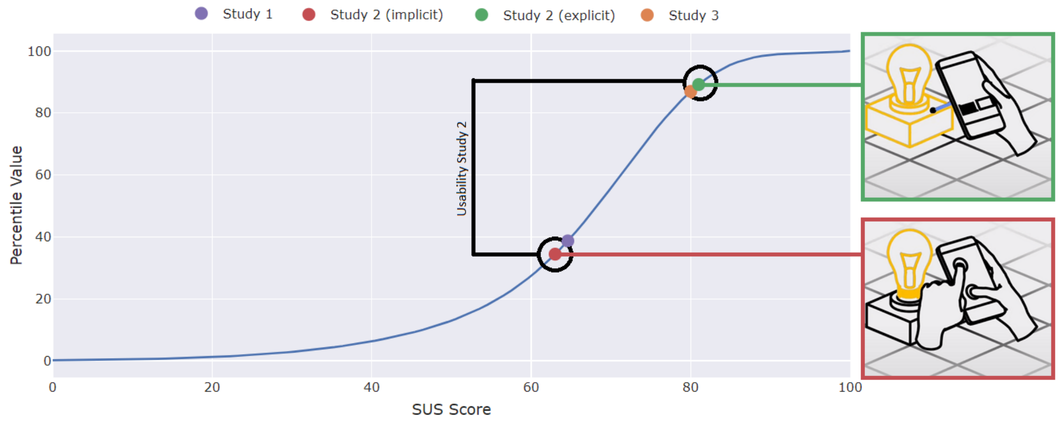


Figure 6.7: The SUS results for both the explicit and implicit interaction concept from usability study 2 contextualized on the percentile curve of SUS scores according to Kortum et al. [240] (left) and an example of combining a grabbed object with another object through both interaction concepts (right).

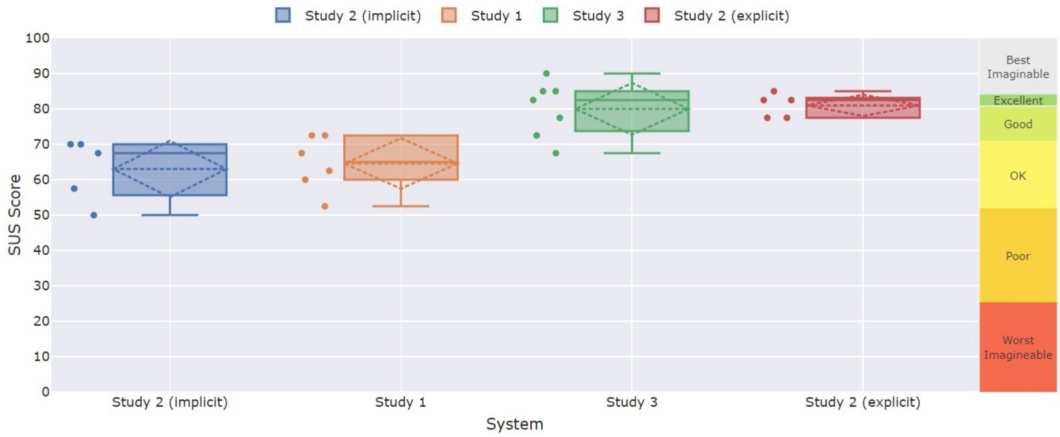


Figure 6.8: System Usability Scale results contextualized on the adjective interpretation scale according to Kortum et al. [240], sorted from lowest to highest average score. Datapoints are represented as dots, mean values as solid lines and average values with their standard deviation as dotted lines.

UEQ studies [419]. In the first study, participants reported an average attractiveness score of 1.25. This would be considered "Above average" in the UEQ benchmark. Regarding perspicuity, they reported a score of -0.13, which would be considered "Bad". The average efficiency score of 0.75 indicated a "Below Average" perceived efficiency of the tool, and a dependability score of 0.96 a "Below Average" dependence. The stimulation score of 2.04 and a novelty score of 2.38 would both be considered "Excellent" compared to existing values of the benchmark data set. In the second study, participants reported an average attractiveness score of 2.40 (Excellent) for the explicit interaction concept, 1.20 (Above average) for the implicit interaction concept and an average perspicuity score of 1.55 (Above Average) and perspicuity 1.10 (Below Average) respectively. In terms of efficiency, participants reported an average score of 1.65 (Good) for the explicit interaction concept and an average score of 1.25 (Above Average) for the implicit interaction concept. For the dependability, participants reported an average score of 1.80 (Excellent) for the explicit and 0.82 (Below Average) for the implicit interaction concept. Participants reported an average stimulation score of 2.50 (Excellent) for the explicit interaction concept and an average stimulation score of 1.70 (Good) for the implicit interaction concept. For both conditions, an Excellent average novelty score was reported, with 2.35 for the explicit interaction concept and 2.10 for the implicit interaction concept. In the third study, participants reported an average attractiveness score of 1.93 (Excellent), perspicuity score of 1.36 (Above Average), efficiency score of 1.21 (Above Average), dependability score of 1.14 (Below Average), stimulation score of 2.04 (Excellent) and a novelty score 2.39 (Excellent) (see Figure 6.9).

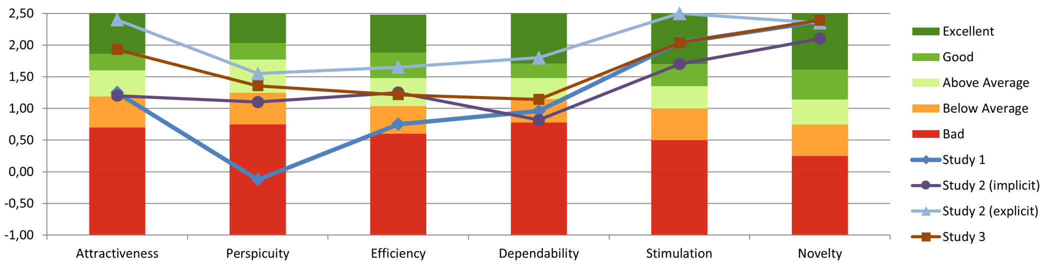


Figure 6.9: User Experience Questionnaire scores of all 3 usability studies in relation to the UEQ benchmark data set.

6.3.4 Results: Qualitative Feedback

Qualitative feedback provided through the qualitative questionnaires, observations, verbally during the experiment or implicitly provided through the "think aloud" methodology, were transcribed, prepared and inductively coded according to Linnenberg et al. [287]. The qualitative questionnaires consisted of 4 questions: What participants liked about the application, what they did not like, what they had problems with during the training and what additional feedback or remarks they wanted to provide. While the combined qualitative feedback was fully utilized for the design-based research process and iterative improvements to the application and TrainAR, they are only reported in very condensed form here and filtered for feedback targeting the interaction concept.

Across studies, participants noted that they liked the *"comprehensible" "step-by-step instructions"*, the continuous feedback provided after actions and the verifiable progress of the training task. They noted that they liked the color scheme and clean design, especially the *"details"* and *"realistic graphics"*, underlining the fact that the virtual objects are *"recognizable"* as their physical counterparts. Additionally, they perceived the application as a *"promising new type of learning"* and enjoyed the gamification aspects of training in AR. Some participants also noted across studies that they sometimes had problems with the tracking and that the virtual assembly sometimes shifted out of place or was temporarily not visible, though this feedback decreased in later studies. Participants also noted that text was sometimes too small for them to read.

In the first study, participants noted that the provided onboarding based on textual instructions and pictures was not sufficient and should be repeatable. They perceived the interaction with objects as *"cumbersome"*, especially for the feedback regarding the process of combining two objects, with all participants providing qualitative feedback indicating they struggled with this interaction. Moreover, some participants struggled to understand the spatial component and distances of objects.

For the implicit interaction concept in the second study, participants who also participated in the first usability study provided feedback indicating that the interaction concept and especially the onboarding somewhat improved. In contrast, the qualitative feedback provided by new participants indicated similar perceptions to the first usability study, still describing the interactions as *"complicated"*, *"abstract"* and *"frustrating"* and especially again noting combining objects as an obstacle. Some suggested a *"trial"* scenario where the interaction could be tested. For the explicit interaction concept, all participants who took part in the first usability study provided feedback that the interaction *"drastically improved"* and that the usage was *"less frustrating"* as it provided *"more feedback"*. This sentiment was shared by the participants who used the application for the first time, describing the instruction handling of the application as *"clear"*. In both conditions, participants noted that they liked the training assessment at the end of the scenario, though noting that they believe that high error counts might be discouraging for some users.

In the third study, participants especially liked the improved training assessment, now also explicitly stating what kind of professional errors were made. Participants who only conducted the first study or the implicit interaction condition in the second study provided feedback similar to the explicit condition in the second study, indicating *"improved"* handling and onboarding. Moreover, some participants stated that they think the application is somewhat *"strict"* concerning what procedures would be correct.

6.3.5 Subsequent Improvements

Besides a substantial amount of midwifery context-specific adjustments and changes regarding the state flow of the training across the three formative usability studies, the most important implications and subsequent improvements to the TrainAR interaction concept are as follows:

In the first study, an interaction concept based on on-screen gestures for all basic actions described in Section 6.1 was developed, e.g., using a short press for an interaction, a long press for grabbing & releasing objects and a combined long press with a short press while overlapping two virtual objects for combining objects. Contrary to the results suggested by previous work and our expectations, at least in the context of academic midwifes, those prior findings could not be replicated. Even with the improved onboarding based on textual instructions combined with explanatory animations in the second study, participants struggled to utilize the interaction concept effectively. While the perceived perspicuity did drastically improve in the second study, most likely due to the improved onboarding, it was still below average and lower than the perceived perspicuity of the explicit interaction concept. Additionally, the overall usability of this condition in the second study did not improve compared to the first study, but the usability of the newly introduced explicit interaction concept was significantly higher. When contextualizing all three studies on a percentile curve of SUS scores gathered in a meta analysis by Kortum et al. [240], this difference becomes even more apparent, clearly visualizing two groups of usability scores for both interaction concepts across the usability studies (see Figure 6.7). This was further affirmed by singling out the participants who took part in all studies, the qualitative feedback by all participants indicating that they would need more onboarding or even a trial scenario using the implicit concept, before starting the actual training scenario and the repeatedly noted frustration. Neither was similar qualitative feedback reported in the questionnaires, nor observable for the explicit interaction concept during the second or the third study.

During the first usability study, it was possible to select objects from any distance. It was observable that participants did not utilize the translation and rotation of the device itself effectively, some even voicing the need for "zooming" to better red displayed text in the context. Subsequently, a maximum range at which objects would be selected was introduced, and the crosshair was improved, so the two circles would converge when close to the distance at which an interaction would be possible (see Figure 6.6a,b). This improved the observable utilization of the device translation/rotation as part of the interaction in the subsequent studies.

Partially independent of the interaction conditions (implicit or explicit), two additional trends emerged throughout the studies. Explicitness and deliberate redundancies of interaction visualizations and feedback mechanisms improved the users' perceived attractiveness and efficiency of the AR application, and was particularly reported as positive through qualitative feedback. Subsequent improvements, especially for the improvement of explicitness in the third study therefore comprised of not only outlining a selected object, but also slightly coloring it in the selection color using a shader (see Figure 6.6a), no longer outlining objects when they are grabbed (see Figure 6.6c) and, for the state of combining, making the grabbed object transparent while outlining the object to be combined with (see Figure 6.6d). Additionally, the buttons used in the explicit interaction concept were only displayed when an object is selected, grabbed or in a combining state when they are usable and also depicted the specific interaction that would be triggered. The redundancy of feedback mechanisms was perceived positively, therefore correct or incorrect interactions elicit a visual feedback on the UI, visual feedback through blinking outlines in the AR context itself and auditory feedback.

In the third study, spatially contextualized speech bubbles were introduced to communicate implicitly triggered interactions that are not actually performed in the AR training, like the disinfecting of the hands or the insights provided by the professional midwife. As observations, qualitative feedback and the higher variance of reported usability scores indicated this could potentially be overwhelming for some users, those speech bubbles were subsequently also transitioned into UI elements (see Figure 6.5a,e,g).

6.4 Didactic Framework & Utility

In modern education theories, the focus is on problem-based and therefore learner-centered learning settings that enable both individual and self-determined, but also collaborative learning [535]. The aim is to promote the development of complex technical and practical knowledge as well as professional competence. Action and work process orientation, which represent central concepts of vocational pedagogy [397], are suitable for this purpose and are well compatible with the interaction concept provided through TrainAR. Action and work-process orientation find methodological expression in the complete action [448]. The acquisition of competences takes place through repeated runs of application-oriented phases: behavior of the learner/actor, feedback and evaluation of the actions with renewed goal setting [227], which corresponds to the phases of a complete action: 1. informing, 2. planning, 3. deciding, 4. executing, 5. controlling, 6. evaluating [448]. For the specification, conception, and development of work process-oriented AR teaching/learning scenarios, correspondingly detailed descriptions of the work processes including necessary decisions and information flows are required. Referring to Howe et al. [110], subject-specific methods for collecting and describing information flows are developed. Based on this, authentic, complex problems are used as a starting point for work process-oriented knowledge acquisition from the above-mentioned subject areas and diverse AR learning scenarios are derived according to the competence goals. For learning and transfer effects, one of the central concepts is to create suitable occasions for reflection and to support them with learning guides. In the practical design of TrainAR, the minimalism dimension according to Drljević et al. [116] is taken into account, so that only necessary information is provided. This avoids stimulus overload and supports focusing on the procedural flow.

The intention is to systematically put knowledge into practice. For this, the assumption that a person is enabled to act independently and responsibly is pursued. TrainAR's training scenarios are therefore based on work process descriptions and competence-oriented learning objectives, where the learners' learning conditions, preexisting experiences, and knowledge should be considered by authors of TrainAR trainings [100].

6.4.1 Training Contextualisation & Structure

TrainAR training applications focus on the teaching of intellectual skills and cognitive strategies, according to instructional design theory as proposed by Gagne [145]. They are also aligned with the principles of the ID/4C model by van Merrienboer [471] for learner-centered competency-based curricula instructions in the context of complex learning tasks. During TrainAR trainings, first verbal information and declarative knowledge is taught through traditional class-based teaching or in self study. Afterward, the procedural knowledge, combining intellectual skills and cognitive strategies, can be trained using TrainAR, but motor skills are not reinforced at this point. Through this contextualization, TrainAR serves as a pre-training and motor skills will be trained with physical material in the practical training settings (e.g., SkillsLabs in the clinical setting). Before applying learned procedures in practice or as a reinforcement of best practices and attitudes, TrainAR can also be applied as a retention training after the physical on-site trainings, to enable rehearse procedures (see Figure 6.10).

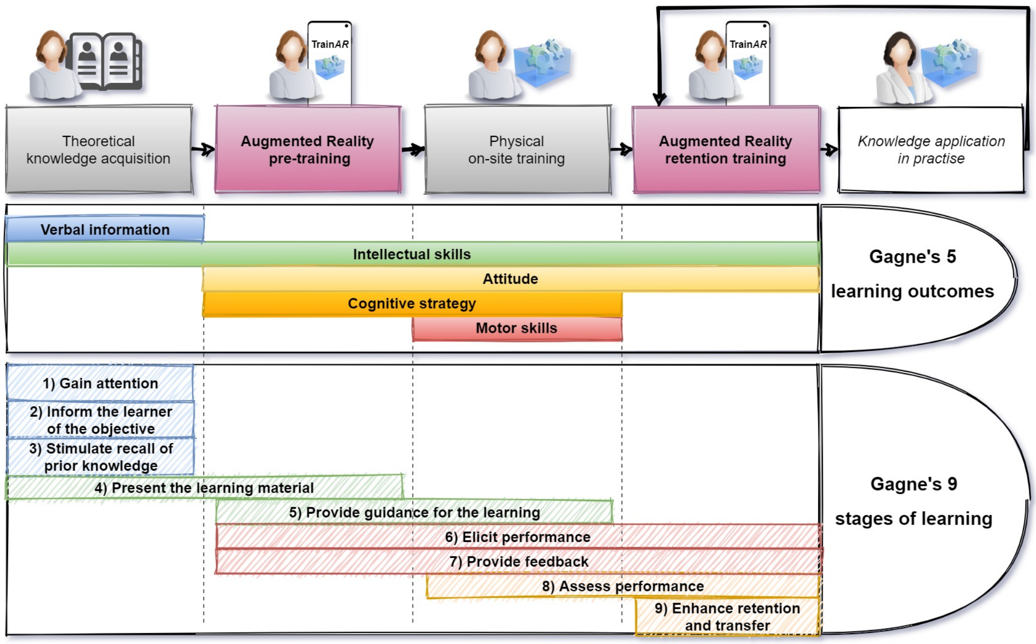


Figure 6.10: Possibilities for the curricular embedding of TrainAR as pre-trainings to practical on-site trainings, as retention opportunities after practical trainings, or a combination of both. Contextualized with the instructional design theory of Gagne's 5 learning outcomes and 9 events of learning [145, 146].

During each TrainAR session, the training starts with a short case description according to the principle of problem-based/learner-centered learning [535]. The trainings always run based on a specific case, therefore contextualizing the procedural knowledge taught, as described in Sections 6.1.1 and 6.2.2, including declarative knowledge peripheral to the task. The aim of this is to link academic theory and practical competences, to support the transfer of complex learning tasks from theory into practice [471]. During the training, expert knowledge is available in contextualized form (see Sections 6.1.5 and 6.2.6) and after completing the training, trainees will receive an assessment of their training performance and professional feedback (see Section 6.2.7).

In this, when contextualizing TrainAR trainings on the ID/4C model [471], it covers all four components (learning tasks, part-task practice, procedural information, and supportive information), but to different degrees. It can formalize, conceptually arrange and explain learning tasks in the form of one or more part task practices with guidance that can be used in the initial stages, but it has to be noted that even in those cases, there are always practical "part-task practices" necessary to cover the motor components of training tasks in real environments and some learning tasks are likely rather introduced through conventional methods. Primarily, TrainAR trainings are aimed to be utilized to train the procedural information component during the part- and whole-task practices, therefore providing step-by-step instructions and then the possibility to perform them and iteratively self-assess this performance. Finally, while supportive information is partially integrated as insights/hints into the trainings, TrainAR is only peripherally focused on this component and relies on teaching concepts surrounding the TrainAR trainings to provide this component in the curriculum.

6.4.2 Integration in Curricular Teaching

Utilizing TrainAR in the course of a curriculum, the teacher transitions from a lecturer to a tutor and (at least partially) gives up control and steering of the learners' learning activities. Instead, they offer support, guidance and the theoretical (supportive) framework. This is intended, among other things, to support the empowerment of the learners [116]. The AR training can be used at different stages in the course of study, even iteratively. This is achieved through the usage of adaptive instructions providing difficulty settings for the same training procedure (see Section 6.1.2). The first mode, known as guidance, should not require any prior experience. Above all, the intellectual skills and cognitive strategy associated with a procedural task are trained here. The trainees are introduced step-by-step to the procedure, following a primarily behaviorist approach, as described in Sections 6.1.4 and 6.2.5. The second mode is the training mode, in which prior knowledge of the subject is required. The aim is to consolidate the process and elicit reinforcement of prior knowledge. Different courses of action can be followed. Here, more cognitivism approaches are followed, taking into account the cause and effect mechanisms, including the learning process. Therefore, the focus is primarily on methods of knowledge transfer in the first place, the competence transfer of procedural learning.

Expert knowledge is integrated in both modes and linked to actions or objects (see Sections 6.1.5 and 6.2.6). This knowledge is reproduced auditory and visually. Trainees receive real-time feedback after each session, as described in Section 6.2.7. In the form of a point scale, the trainees can rank/rate their performance and feedback is also given in written form. Hence, both positive and negative aspects are highlighted according to the mastery principle so that the trainees receive confirmation of their success, but also information about their mistakes or suggestions for improvement. The provision of real-time feedback has a positive influence on the motivation of the learner, as it can support the comparison with the individual learning success [4]. The choice of learning environment is very open, so that the AR application can be used anywhere, e.g., at home, in a SkillsLab, or in the classroom, especially enabling BYOD approaches where trainees can use their own smartphones for the AR training. There is generally a need for flexibility in the educational process; The chosen flexibility dimension also makes it possible to carry out the training outside the curricular integration, regardless of location and time, to consolidate the procedural flow. For example, before or during a practical study phase [116].

Figure 6.10 shows an exemplary curricular integration envisioned for the curricular integration of TrainAR, contextualized with the five learning outcomes and nine stages of learning proposed by Gagne [145, 146]. In the first step, the theoretical framework is dealt with in the context of classical forms of teaching, such as lectures and seminars. Here, the learners' attention is gained, the learner is informed of the objective, the learning is contextualized in prior learning and the procedural task is presented. In this stage, primarily verbal information, therefore declarative knowledge of the procedural task, with some intellectual skills, such as broad concepts, are introduced. As a second step, the AR-supported procedural training using TrainAR takes place as a pre-training. Therefore, the guidance mode offers the behaviorist support during this training. The trainees have the opportunity to understand the process at their own pace but are strictly guided. Trainees are presented the learning material, are provided guidance, and are provided the opportunity to elicit the performance and receive feedback from the application. In this stage, the intellectual skills are trained in combination with their corresponding cognitive strategies to solve the procedural component. In the third step, motor skills are practiced and consolidated in practical on-site trainings (e.g., SkillsLabs in the medical setting, or laboratory experiments in the chemical engineering context) and the performance of the learner can be assessed. Here, learners already know the entire sequence and developed a cognitive strategy to solve it, that can then be linked to the motor actions required to perform it in reality. Finally, the AR retention training is envisioned as a training mode, that helps learners in consolidating and rehearsing the sequence of their actions. The trainees can carry out the action more freely, compared to the pre-training, and also consider new action alternatives with AR support. Additionally, it can be used for self-directed knowledge verification, not only assessing performance but also enhancing retention and transfer.

6.4.3 Applying TrainAR to Procedural Training Tasks

In many vocational settings, it is important to train procedural courses of action as precisely as possible, as errors in the procedure can have devastating effects. Especially in medical and health science, where standardized procedural trainings are taught regularly, and their correct application is especially important, methods were developed to transform procedural knowledge from practice into controllable and verifiable training settings. Derived from those methods and applied for the exemplary implementation of TrainAR described in Section 6.2, but also applicable outside the medical scope, we propose that scenarios utilizing TrainAR should be developed by: *Identifying & observing the procedure*, *analyzing & deriving the work-process-description*, *defining the competency-based learning objectives*, and *transforming the didactic considerations towards an AR application* utilizing TrainAR's interaction concepts. (see Figure 6.11).

Hereby, this procedure is envisioned as systematic and strictly sequential, condensing but still largely following the classic instructional design model by Dick et al. [109] that defined the necessary steps for the development of training instructions as a 10-step process: First, the teaching objectives have to be determined (1). Following this, teaching material and learning processes (2) as well as previous knowledge should be analyzed and determined (3). Then, criteria for learning success (4) and test items (5) have to be developed. Afterward, the instruction strategy is defined (6), which includes the didactic method, exercises, and feedback. The teaching material can then be selected and produced (7) and formative evaluations can be planned and carried out (8). Finally, the learning offer is revised (9) and summative evaluations are planned and carried out (10).

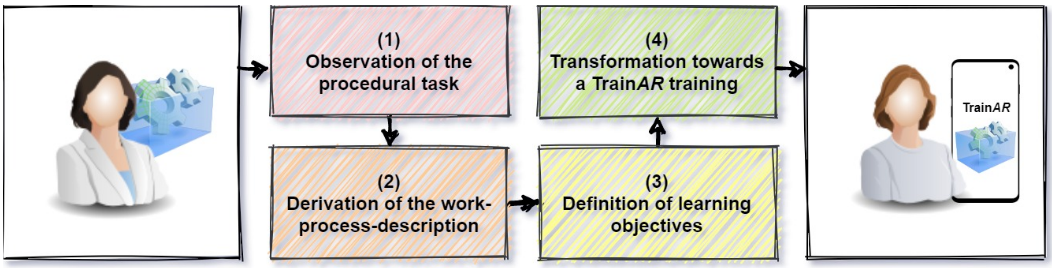


Figure 6.11: The transformation of a procedural action sequence into a TrainAR training scenario.

Identifying & Observing the Procedural Task

As a central concept of design-oriented media didactics according to Kerres [226], media sources should be utilized as a contribution towards solving an educational problem and not applied without specific cause. While new media sources fundamentally open up new opportunities and have potential for different types of learning, this is not based on an inherent effect of increased learning success. They require dedicated planning and conception to be able to induce benefits [226]. This includes AR training scenarios. TrainAR scenarios should be therefore carefully identified based on their suitability for training in AR. What procedural AR trainings are suitable is dependent on the complexity and contingency of the educational field, but generally procedures that combine declarative knowledge with complex cognitive strategies are ideal. While procedures with significant amounts of motor skills are possible, as shown in Section 6.2, motor-learning components of the procedure itself have to be training in physical on-site trainings and cannot be trained using TrainAR autonomously (see Figure 6.10).

After a suitable procedural training task is identified, the training task, demonstrated by a domain expert, should be systematically observed and ideally videographed. Recording does not only allow preservation of the initial observation and expert input, but also serves as a basis for the development of the work-process-description.

Analysing & Deriving the Work-Process-Description

When the selected procedure is observed and documented, it should be converted into a work-process-description as described in [54, 70]. This should be developed towards a work-process-model, describing each possible step and action of the procedure and their interconnections. Therefore, while the work-process-description only describes the procedure as observed, the work-process-model also forces a decision about which measures have to be taken after each step. In Section 6.1, this is refereed to as the *state process model* from a technical perspective. In such work-process-models, a distinction is traditionally made between input, work sequences and output. Here, task instructions are the input, which are given to the trainee, including distractors and deliberate disturbances and interruptions in the course of action. The model should be derived by starting with an initially stringent, linear, idealistic action sequence and then alternative, further sequences can be added. The results are then included in the output. This means that all the necessary information from the documented work process descriptions is in the process model and can be used for further design developments.

Definition of Competency-Based Learning Objectives

After the work process has been described, the definition of the competency-based learning objectives can be carried out. For this purpose, the cognitive and psychomotor learning goals are derived from both the work-process-description and the work-process-model. Those should primarily be based on taxonomy levels according to Bloom [11] and clinical competence levels according to Miller's pyramid of clinical assessment [323]. These established educational frameworks include learning objectives as well as assessment measures. Bloom's taxonomy is well established for lesson planning, design, assessment, and evaluation. Bloom divided the learning levels into cognitive, psychomotor and effective areas, which are independent but mutually influence one another. In the Miller pyramid, the learning process is divided into four levels. Knowledge is the basis and routine application, especially in clinical environments, is the top priority.

To achieve this, first target group analysis should be carried out, e.g., in the form of Personas [226]. This includes framing conditions such as the intended curricular integration, localization of the application and previous knowledge of the learners [109]. The previous knowledge of the learners in particular gives an important and decisive direction both in the formulation of learning objectives and in the later technical application development. The work process model should then be divided into sections and formulated in constant comparison with the prior knowledge of the learner's learning content. To be able to formulate learning objectives, cognitive and psychometric taxonomy levels are assigned to the learning content [242]. Here, verbs should be assigned to each taxonomy level, to formulate learning objectives precisely. Based on these taxonomy levels, the assignment to the Miller [323] pyramid levels can be made.

Transformation Towards a TrainAR Training

When completing the classification of the learning objectives and competence levels, the transformation towards a TrainAR training scenario can be carried out utilizing the "mobile augmented reality education design frameworks" (MARE) [529]. The MARE-Model is a developed outcome layer that combines the Miller pyramid and the Bloom taxonomy levels. It contains these differentiated dimensions of learning and enables a transfer to AR learning activities via these classifications. The general requirements for AR learning activities described by Zhu et al. [529] are predefined by the usage of the TrainAR features described in Section 6.1. Based on those general requirements, scenario-specific AR requirements should be formulated. Depending on the taxonomy level, different approaches can be utilized: Should trainees be given an explanation of the procedure, should they carry them out independently or is a combination necessary (see Figure 6.10)? In addition, scenario- and location-specific AR implementation recommendations could be worked out on the basis of an AR property overview [133]. Since the scenarios are usually very complex and detailed analyses have taken place in advance, it might be helpful to take a step back and look objectively at the combination of the state-process-model and learning objectives and go through the scenario step by step and consider which AR properties were utilized effectively.

The MARE design framework is a learning theory that serves as a guide for developing AR apps for educational purposes. Primarily aimed at educational AR apps in the medical context but arguably applicable beyond that scope, it was constructed using a conceptual framework analysis method in which Zhu et al. [529] identify interconnected key concepts. In an iterative process, they discovered three main elements: (1) Foundation, (2) Function and (3) Outcome. Learning theories form the basis (1), as they are elementary for the form of teaching content. Zhu et al. [529] selected situated-, experiential- and transformative learning theories for the foundation. The situated learning offers learners a real-life-environment of learning and interaction. Experiential learning combines experience and behavior, e.g., in a virtual learning environment in which feeling, thinking, observing and acting are the focus. Transformative learning involves critical reflection and transformation in meaning and perspective. The focus here is on changing problematic frames of reference. The Foundation (1) and the Outcome (3) layer support the design aim. The Outcome layer comprises learning objectives as well as expected skills of the learner and assessment of the learning. These elements are helpful in finding out which skills may be achieved utilizing MARE. For the transfer of learning objectives into AR trainings, the outcome layer offers a basis that provides orientation for implementation. This also includes Bloom's taxonomy levels, which are well known for conventional lesson planning. If there is not yet routine in the definition of learning objectives, it might be challenging to derive them. In this case, we suggest including the outcome layer in the definition of learning objectives, as it is immediately visible which levels contain which activity, making it more practical. The Function (2) layer includes how learning can be achieved with the following levels: learner's personal paradigm, learning activities, learning environment and also learning assets [529].

TrainAR is primarily developed with the theory of experiential learning as one of the central concepts. The learning theories and the procedure for the application of TrainAR for a training task presented in this section do not necessarily have to be selected. Alternatively, also more constructivist planning models like the R2D2 model by Willis [503] would be conceivable as a basis for further scenario development. However, the learning and instructional design theories largely determined the design of the interaction concept and the presented application procedure provides a clear, didactically reasoned approach for the development of additional AR trainings using TrainAR.

6.5 TrainAR Components from a Technical Perspective

Forms of interaction with virtual objects and especially procedural interactions in the form of chains of actions are well studied in Virtual Reality (VR) settings and comprehensive toolkits, and frameworks exist to implement them. For example, pre-implemented interaction metaphors and presets delivered with frameworks such as the SteamVR Toolkit, XR Interaction Toolkit SDK, VRTK, OpenVR, or Microsoft MRTK allow developers to focus on the content of their training application rather than worrying about the basic interaction principles with their usability and learnability considerations. For AR, this becomes more challenging, but for HMDs, there are at least gestural interactions, external controllers, and frameworks such as the Microsoft MRTK to provide interaction concepts and basic principles to expand on.

Arguably, for handheld/mobile AR, the case is even more complicated. While interaction concepts are sparsely explored in the literature and some interaction toolkits do exist, they are currently neither evaluated, nor do they provide the same "out-of-the-box" application utility for developers to directly apply them the same way, compared to VR development. As mobile AR interaction concepts are mostly visual/viewing experiences or ray-casting-based approaches utilizing direct screen touch or UI button approaches, this challenge is only exaggerated in the context of procedural chains of actions necessary for task training. Combining this research gap with the current general challenges faced in Mixed Reality research of finding out how to onboard users to this novel type of application and type of interaction with the uncertainty of how, when, and how much feedback to provide to the user during trainings, this leads to a substantial amount of time spent by developers on designing interaction concepts from scratch and technical aspects of AR trainings instead of focusing on the content of the training itself. Furthermore, this always requires iterative feedback loops with didactic experts on the interaction and feedback mechanisms. It creates a causality dilemma of not being able to develop a fitting interaction concept and feedback mechanisms without knowing the training task in detail, but also not being able to transfer practical trainings towards a technically implementable flow of states before having a reference for what such a handheld AR training could look like. This dilemma not only makes AR training development particularly time-consuming for interdisciplinary teams of experts, but makes development of AR scenarios by non-programmers or programmers without AR-specific expertise (e.g., technical-domain experts or designers) impossible.

6.5.1 Combining TrainAR Components Towards an AR Authoring Tool

To address these challenges holistically, TrainAR is the threefold combination of (1) an interaction and feedback concept for procedural trainings on mobile AR devices (Android and iOS) that is realistically scalable today, (2) a didactic framework explaining the instructional design theory behind the concepts and how an author should conceptually transfer procedural training tasks into TrainAR trainings, and (3) an authoring tool allowing authors without programming expertise to create TrainAR trainings through visual scripting, based on the interaction concept and didactic considerations (see Figure 6.2).

While the interaction concept and didactic framework were already elaborated and evaluated through exemplary implementations in the previous sections, we now focus on the Unity-based visual scripting authoring tool of TrainAR that allows for the creation of such trainings in accordance with these didactic considerations, utilizing the proposed interaction concept and feedback modalities. To understand the technical implementation, the following subsection, after already having described it from the theoretical and practical perspective, describes the components of the interaction concept and didactic framework, combined from the technical perspective, before describing how to author them and how the authoring tool was developed.

6.5.2 Components of TrainAR from a Technical Perspective

Combining the interaction concept and didactic framework, TrainAR includes concepts and technical solutions for onboarding the user on how to use TrainAR (Section 6.5.2), automatic technical tracking and assembly placement utility, instructing the user on what action to perform next (Section 6.5.2), letting the user perform a procedural non-linear chain of actions (Section 6.5.2), and providing contextualized feedback, insights, and final training assessments aligned with the didactic considerations described in Blattgerste et al. [55] (Section 6.5.2). Those components are included in the authoring tool and automatically included in every training.

Onboarding & Assembly Placement

When an authored TrainAR training is started, trainees are first shown onboarding animations with textual explanations describing how to interact with objects in AR and how to trigger actions on them (see Figure 6.12a–c). This onboarding utility is included with the framework and automatically deployed by the authoring tool when building the training for a target device. Therefore, no considerations regarding onboarding and concept explanation have to be made by the authors of trainings. When using a training, TrainAR automatically lets the trainee scan (see Figure 6.12d) a surface area until a sufficiently large free area is recognized by the underlying tracking library and places the TrainAR training assembly onto the surface to start the training (see Figure 6.12e). This technical tracking and placement utility with its associated onboarding animations is also automatically included in each authored TrainAR training.

TrainAR Objects and Procedural Chains of Actions

After the placement of the training assembly, the trainee can complete a procedural chain of actions defined by the author of the training. Those trainings consist of the basic actions of selecting (see Figure 6.13f), grabbing (see Figure 6.13g), interacting with (see Figure 6.13h), and combining (see Figure 6.13i) virtual AR objects called “TrainAR Objects”.

These TrainAR Objects are virtual AR 3D models that were converted by the TrainAR authoring tool and enriched by scripts, providing them with consistent basic interactive functionality. These automatically inherited responses are as follows. When selecting a TrainAR Object, subtle shading and outlining of the selected object is applied (see Figure 6.13f). When grabbing a TrainAR Object, it leaps and attaches itself into a static position in front of the handheld device, where it is always rotated into an upward position from the assembly ground and keeps a defined distance from the device to make the object stay visible on the screen throughout the interaction. Users can then displace it, interact with it, or combine it with another stationary object. When interacting with and combining objects, outlines visualize the current state-change of the object and

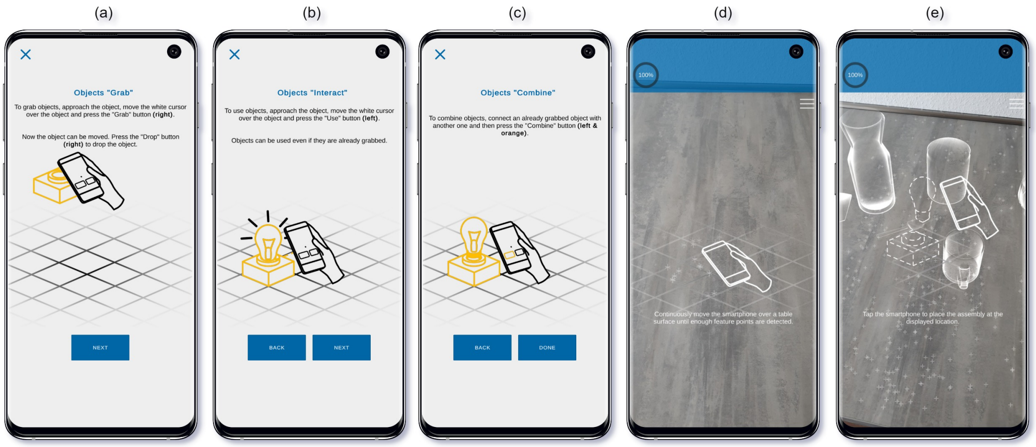


Figure 6.12: TrainAR automatically includes onboarding screens and technical utility for (**a**) grabbing objects, (**b**) interacting with objects, (**c**) combining objects, (**d**) scanning the training area, and (**e**) placing the training assembly.

whether the action was accepted (valid) or not, while object-specific interactions that are triggered are defined by the author of the training.

Alongside those basic actions, trainings can have Custom Actions (see Figure 6.14m) that serve as customizable action triggers defined by the author. This allows authors to implement independent concepts outside the interaction scope provided by TrainAR. Furthermore, trainings can utilize predefined UI components such as input fields (see Figure 6.14j), questionnaires (see Figure 6.14k), or list selections (see Figure 6.14l) to realize in-procedure quizzes or material selection, or to check for decision procedures that could not otherwise be sufficiently covered by just the basic actions of TrainAR.

Instructions, Insights, and Feedback

Besides the actions, which serve as input from the trainee, several types of output modalities are delivered with TrainAR. They are used to elicit instructions, feedback, and insights or to indicate technical problems. Firstly, the technical feedback screens are always included when deploying a training. They automatically trigger when technical problems are detected to provide feedback to trainees, e.g. if there is insufficient light, not enough feature points for tracking, or the smartphone moves too fast (see Figure 6.15n). Authors do not have to develop any technical instructions or problem feedback themselves.

To instruct the trainee on what action or bundle of actions should be performed next, textual instructions are displayed on the UI panel on top of the device screen, including a progress bar showing the current completion percentage of the training to the trainee (see Figure 6.15o). After triggering one of the actions, the trainee is always provided with feedback in the form of a blinking

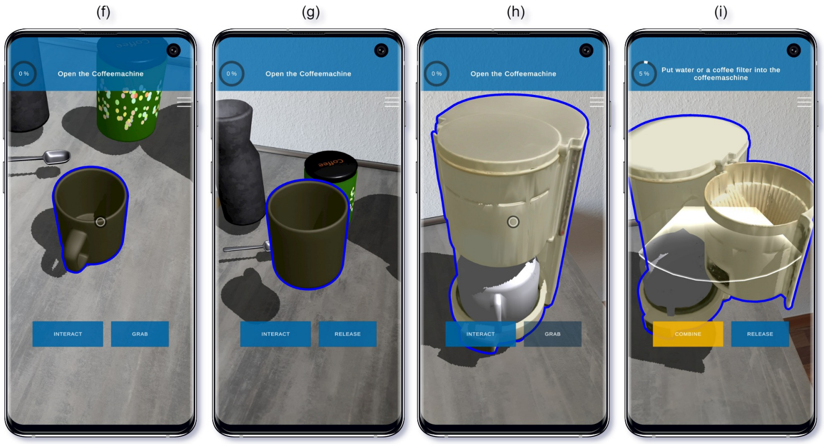


Figure 6.13: The basic AR actions of the TrainAR Interaction Concept that allow trainees to (**f**) select and (**g**) grab TrainAR objects. Selected or grabbed objects can be (**h**) interacted with. Grabbed objects can be (**i**) combined with another TrainAR object by overlapping them.

outline and a sound effect. For some errors, it might be necessary to communicate a message. In this case, error overlays take the trainee out of the training context into the UI and can show textual feedback to the trainee which has to be dismissed manually (see Figure 6.15q). Occasionally, there might be information that is neither instruction nor feedback on an action of the user but still important as part of the training, e.g., expert insights that provide additional tips from practice. In this case, insights can be used that display textual tips as a speech bubble UI element at the top of the screen, optionally also including auditory tips (see Figure 6.15p). After the training is concluded, a training summary is displayed to the trainee showing the training time, number of errors, and the errors contextualized on a performance scale (see Figure 6.15r).

6.6 The TrainAR Authoring Tool

The TrainAR Authoring Tool is a Unity-based authoring environment that is built upon the Unity Editor interface and utilizes the ARFoundation, ARKit, ARCore, and Visual Scripting packages. Its layout inside Unity is displayed in Figure 6.16. It allows authors to create TrainAR procedural trainings out of the components described from the technical perspective in Subsection 6.5.2, utilizing the interaction concepts and didactic perspective proposed in [55]. The authoring tool thereby delivers the interaction concept and all action implementations, feedback mechanisms, and technical solutions for onboarding, tracking aid, and training assembly placement. Furthermore, it provides tools to convert 3D objects into TrainAR Objects that automatically inherit all TrainAR behaviors necessary to work within the flow of states of the training. The author of a TrainAR training only has to import 3D models, convert them, and then refer-

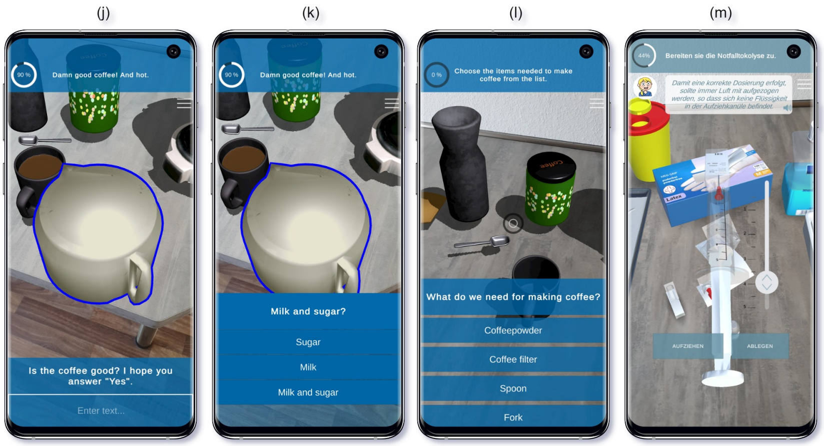


Figure 6.14: Custom Actions include UI-based "quiz" actions such as **(j)** text input fields, **(k)** questionnaire elements, and **(l)** list selection elements. Authors can also create their own UI overlays that trigger custom actions, for example, **(m)** a slider to pull up a syringe (see Section 7.1.1).

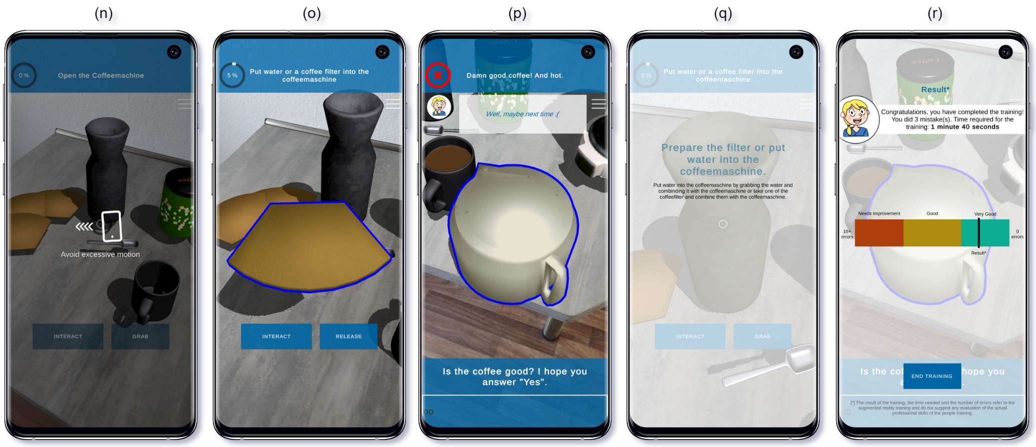


Figure 6.15: The output modalities of the TrainAR Interaction Concept consist of **(n)** feedback to aid technical problems, **(o)** instructions and progress indicators, **(p)** expert tips and insights, **(q)** error feedback overlays, and **(r)** a training summary at the end of each training.

ence them in a procedural visual scripting flow to specify their state changes during the training based on user actions. Authors can then optionally implement additional guiding instructions, feedback modalities, or quizzes. These two central concepts and the remaining tasks for the authors are referred to as the "TrainAR Objects" in the "Training Assembly" and the "TrainAR Stateflow" in the "TrainAR Statemachine" (see Figure 6.16). To enable authors to accomplish those tasks, the layout of the authoring tool is split into several regions: the "Unity Project folder" that shows all imported Assets in the project, a simplified "Unity Inspector" with a list of Objects currently displayed in the scene, the "TrainAR Assembly Scene", allowing authors to view the TrainAR Objects of their training contextualized on a reference setup, and the "TrainAR Visual Statemachine", which allows authors to determine the state flow during the training, based on the users' actions. The "Device Preview" allows authors to preview the training assembly from the perspective of the users' smartphone.

At the time of publication, the TrainAR Stateflow encompasses 10 types of visual-scripting nodes that can be used by referencing the corresponding TrainAR Object by name. All included nodes are visualized in Figure 6.17 in relation to the interaction concept [55]. They are described in more detail in the TrainAR online documentation. The *TrainAR: Onboarding completed, and training assembly placed* node indicates the start of the TrainAR training and automatically starts the flow of states after the Training Assembly was placed in AR by the trainee. The *TrainAR: Object Helper* node is a collection of tools that help to change the state of TrainAR Objects when reached during the flow of states, e.g., changing their visibility, possible actions this object responds to, or replacing them with other objects during the flow. Additionally, four of the nodes are action nodes. If the TrainAR Statemachine reaches one of these nodes during a TrainAR training, it waits for an action by the trainee. These actions can be grabbing, interaction with or combining TrainAR Objects. This can either be exactly one specific action to continue (*TrainAR: Action*), n multiple actions in no particular order (*TrainAR: Action (Multi)*), n actions that lead to $m <= n$ different flows of actions as a consequence (*TrainAR: Action (Fork)*), or the requirement for the user to complete a quiz such as a questionnaire, list-selection task, or text input (*TrainAR: Action (UI)*). The four remaining nodes are output and feedback nodes. The *TrainAR: Instructions* node allows the author to provide adaptive textual instruction to the user of the training. If the user should be provided with specific feedback during the training when performing an incorrect action, the *TrainAR: Feedback* node can be used. Sometimes, information has to be conveyed that is neither direct instruction on what action to perform next, nor feedback based on a performed action. In this case, *TrainAR: Insights* can be used to, e.g., provide additional tips or insights from practice to the trainee.

6.6.1 Design Considerations for the TrainAR Authoring Tool

According to Hampshire et al. [184], AR authoring tools can generally be classified into low-level programming tools, high-level programming tools, low-level content design tools, and high-level design tools. With the increasing abstraction of concepts, authoring tools can also use higher-level interface abstractions, which makes them easier to use. However, as a consequence, this also increasingly limits the AR scenarios that can be created with the tool. Although it would seem plausible to try to target teachers themselves and, therefore, design a high-level content design tool, we deliberately designed and developed a low-level content design tool with TrainAR. While higher-level standalone approaches were considered during the conceptualization, this decision was made for two reasons.

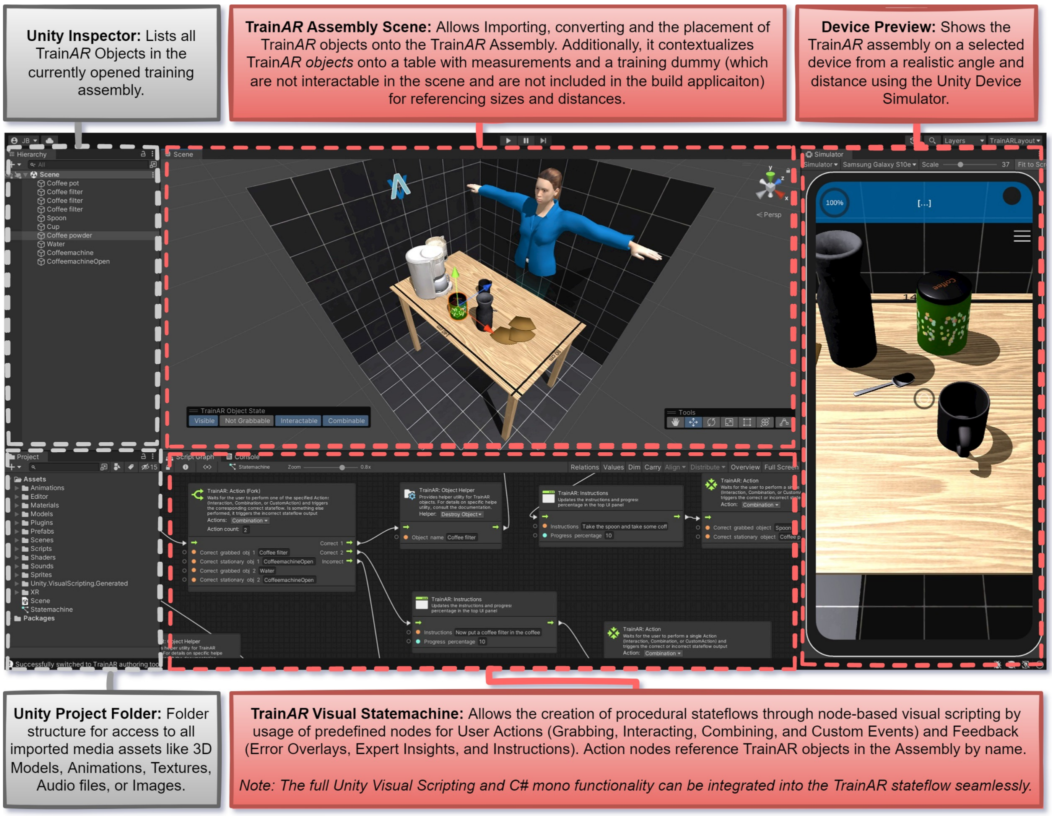
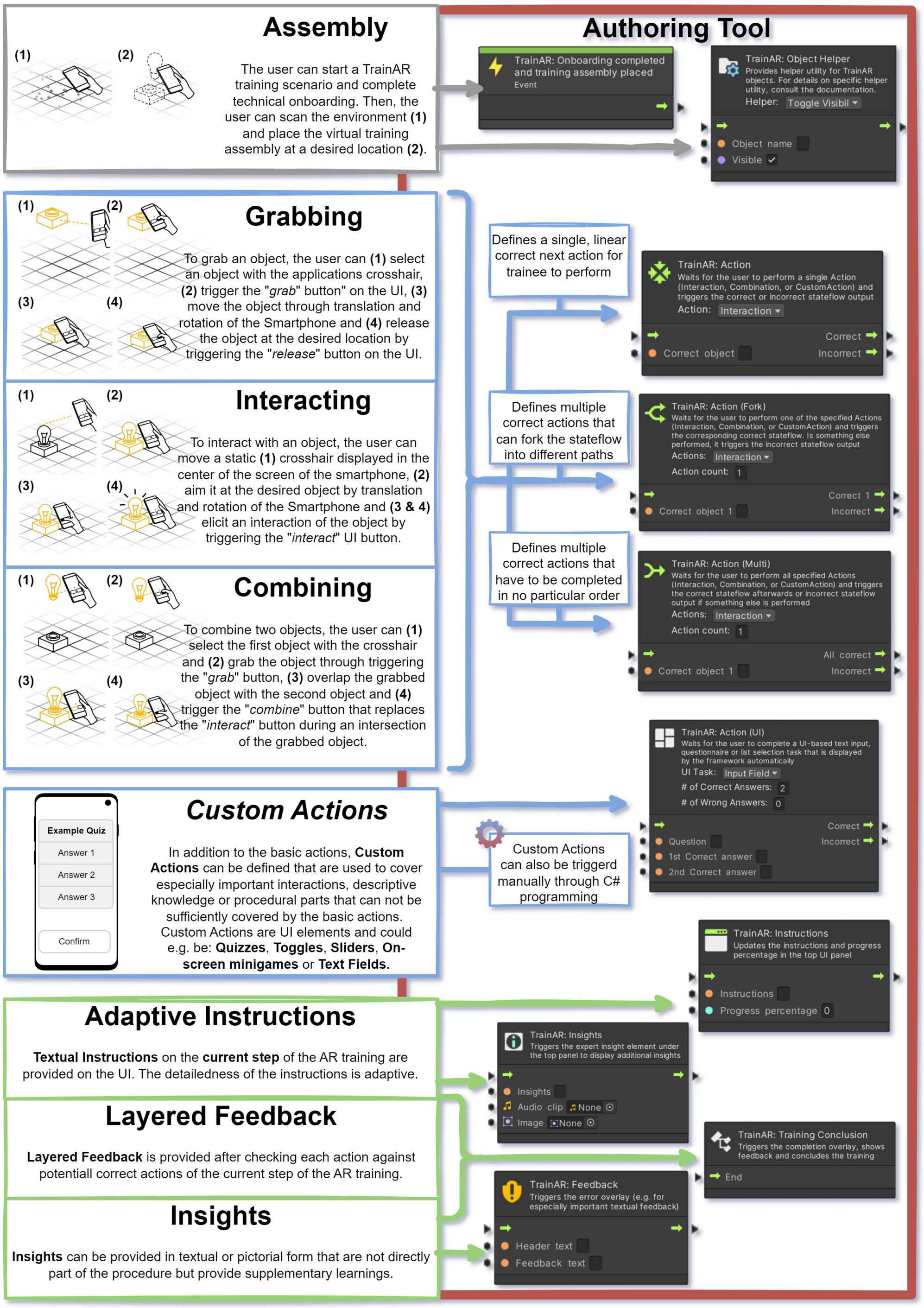


Figure 6.16: The TrainAR Authoring Tool layout combines the Unity Inspector and projects folder with the TrainAR Training Assembly, TrainAR Visual Statemachine, and a Scene preview, allowing authors to create procedural TrainAR trainings.

Firstly, even if the authoring tool itself was designed as a high-level content design tool, the generation, or acquisition of 3D assets to use in the authoring tool would still likely require significant media competency (described in more detail in Section 6.6.4), possibly nullifying the gained advantages from the higher interface abstractions, and could even be too time-consuming to be realistically performed by educators/teachers themselves.

Secondly, this approach allows us to implement the TrainAR authoring tool as a Unity extension, which provides several advantages but inherently comes with increased complexity of the user interface of the tool. As such, TrainAR is an abstraction layer, which allows creating and deploying AR Trainings without any programming expertise (a low-level content design tool), but as an extension, it is also fully integrated into the C# environment and Unity's own Visual Scripting approach. With this approach, programmers can also use TrainAR as a starting point or high-level programming tool and expand it where necessary (described in more detail in Section 6.6.7).



Assembly
(1)
(2)
The user can start a TrainAR training scenario and complete technical onboarding. Then, the user can scan the environment (1) and place the virtual training assembly at a desired location (2).
Authoring Tool
TrainAR: Onboarding completed and training assembly placed
Event
TrainAR: Object Helper
Helper: Toggle Visibil
Object name
Visible
Grabbing
To grab an object, the user can (1) select an object with the applications crosshair, (2) trigger the "grab" button" on the UI, (3) move the object through translation and rotation of the Smartphone and (4) release the object at the desired location by triggering the "release" button on the UI.
Defines a single, linear correct next action for trainee to perform
TrainAR: Action
Action: Interaction
Correct
Correct object
Incorrect
Interacting
To interact with an object, the user can move a static (1) crosshair displayed in the center of the screen of the smartphone, (2) aim it at the desired object by translation and rotation of the Smartphone and (3 & 4) elicit an interaction of the object by triggering the "interact" UI button.
Defines multiple correct actions that can fork the stateflow into different paths
TrainAR: Action (Fork)
Actions: Interaction
Action count: 1
Correct 1
Correct object 1
Incorrect
Combining
To combine two objects, the user can (1) select the first object with the crosshair and (2) grab the object through triggering the "grab" button, (3) overlap the grabbed object with the second object and (4) trigger the "combine" button that replaces the "interact" button during an intersection of the grabbed object.
Defines multiple correct actions that have to be completed in no particular order
TrainAR: Action (Multi)
Actions: Interaction
Action count: 1
All correct
Correct object 1
Incorrect
TrainAR: Action (UI)
UI Task: Input Field
# of Correct Answers: 2
# of Wrong Answers: 0
Correct
Incorrect
Question
1st Correct answer
2nd Correct answer
Custom Actions
Example Quiz
Answer 1
Answer 2
Answer 3
Confirm
In addition to the basic actions, Custom Actions can be defined that are used to cover especially important interactions, descriptive knowledge or procedural parts that can not be sufficiently covered by the basic actions. Custom Actions are UI elements and could e.g. be: Quizzes, Toggles, Sliders, On-screen minigames or Text Fields.
Custom Actions can also be triggerd manually through C# programming
TrainAR: Instructions
Updates the instructions and progress percentage in the top UI panel
Instructions
Progress percentage 0
Adaptive Instructions
Textual Instructions on the current step of the AR training are provided on the UI. The detailedness of the instructions is adaptive.
TrainAR: Insights
Insights
Audio clip None
Image None
Layered Feedback
Layered Feedback is provided after checking each action against potentiall correct actions of the current step of the AR training.
TrainAR: Training Conclusion
End
TrainAR: Feedback
Triggers the error overlay (e.g. for especially important textual feedback)
Header text
Feedback text
Insights
Insights can be provided in textual or pictorial form that are not directly part of the procedure but provide supplementary learnings.

Figure 6.17: The concepts introduced in the TrainAR interaction concept and didactic framework (**left**) and their corresponding Statemachine nodes in the TrainAR Authoring Tool, which can be used to author the TrainAR training.

6.6.2 Open-Source Availability and Documentation

The complete source code of the TrainAR authoring tool is available as a Git repository at `https://github.com/jblattgerste/TrainAR/` (Accessed: 31.03.2023) under the MIT License. Besides the full source code for the authoring tool as a Unity Editor extension, this includes the complete source code for the TrainAR interaction concept, a full documentation of the code, API references (see `https://jblattgerste.github.io/TrainAR`, accessed: 31.03.2023), and a "Getting Started Guide" (see `https://jblattgerste.github.io/TrainAR/manual/GettingStarted.html`, accessed: 31.03.2023) that helps authors of TrainAR trainings to quick-start their AR training development. Additionally, it helps programmers to expand TrainAR towards context-specific needs in a dedicated section to expanding TrainAR.

6.6.3 Envisioned Workflow for Authoring TrainAR Trainings

The designed workflow of using the TrainAR authoring tool is described in detail in the "Getting Started Guide" for one example scenario. Abstractly, it is envisioned as follows.

First, the user downloads the Unity Editor and installs it on a Windows, Linux, or macOS computer. Afterward, the user can download the TrainAR project from GitHub either as a .zip folder or by cloning it via git. Opening the project in the specified Unity version allows the author to then switch Unity to the TrainAR authoring tool mode through a context menu, providing the author with the authoring tool setup shown in Figure 6.16.

The user can then start with the authoring of a TrainAR scenario by importing 3D models into the Unity Project Folder, placing them into the TrainAR Assembly Scene, and converting them into TrainAR Objects through a simple click on a button that starts TrainAR's model conversion process visualized in Figure 6.18. Afterward, through overlays in the TrainAR Assembly scene, the author can then translate, rotate, and scale models and define the TrainAR objects' initial set of interaction abilities (Visible, Grabbable, Interactable, Combinable). The author can thereby compare sizes and distances based on the reference preview scene provided with the authoring tool. After the conversion of all models and the arrangement of the training assembly scene, the author can create the flow of states in the TrainAR Visual Statemachine through visual scripting. This is carried out by adding the TrainAR logic nodes specified in Figure 6.17 and referencing TrainAR Objects in those nodes by name.

After completing the authoring process of both the TrainAR Stateflow and assembly, the author can connect an Android or iOS device to the computer and press the Play button at the top of the editor to install the TrainAR training app to a smartphone. This deploys the training to the device and, besides the authors' objects and stateflow, automatically includes the TrainAR interaction concept, onboarding animations, technical tracking, and assembly placement utility.

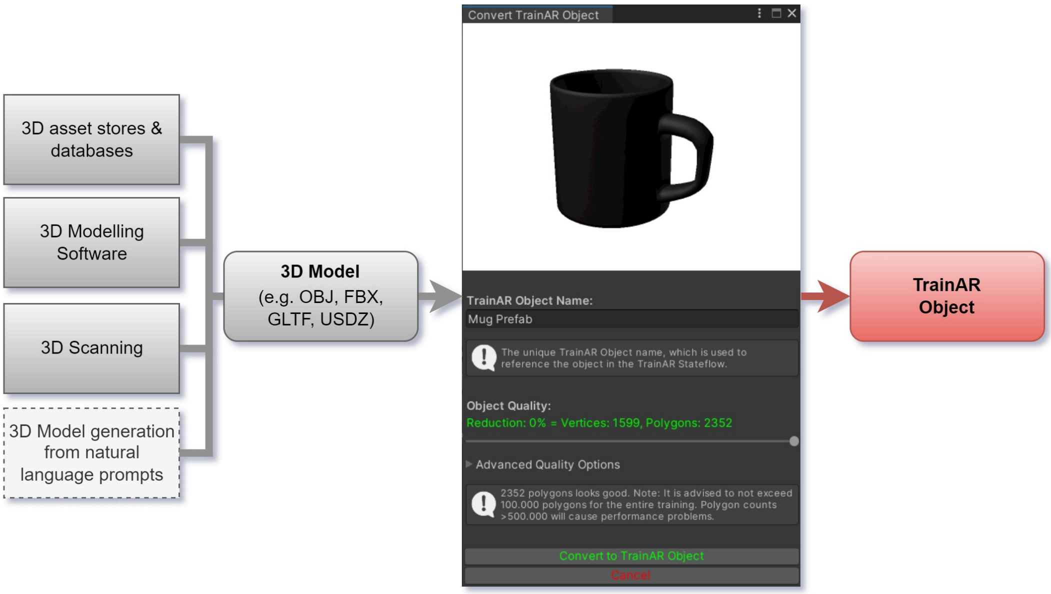


Figure 6.18: The modal window for the conversion of 3D objects from a variety of sources and in different formats into a consistent TrainAR Object, which is simplified, compatible with TrainAR, automatically inherits all intended behaviors, and can be referenced in the TrainAR Visual Statemachine.

6.6.4 Content Generation through 3D Scanning and Natural Language Prompts

With this authoring workflow, the most challenging technical aspect remaining for authors is likely the generation of the 3D content for the AR trainings. While there is an increasing availability of educational 3D content on the web [174], available models might not always be fitting or there might be no models available for specific training contexts. Therefore, besides the creation of models through 3D modeling software such as Blender, or the processing of CAD models, a central consideration for TrainAR is the generation of 3D content through 3D scanning. Through pre-checks and mesh conversions during the conversion from normal Unity GameObjects with attached 3D meshes in various formats to TrainAR Objects, models are created automatically that are compliant with the TrainAR framework, independent of their source and initial structure (see Figure 6.18). As this includes mesh reparation, simplification, and merging, this is not only helpful for 3D scanned objects but also paves the way for the inclusion of meshes from other sources that will emerge in the near future, e.g., 3D models generated through natural language prompt-based approaches, as is currently being researched by Google Research [206].

6.6.5 Key Aspects of the Technical Implementation

As the source code for all technical components of the framework is fully open source and in accordance with common coding and commenting conventions for C# [320], including API references in the documentation, technical aspects of the framework are not described in detail here again. Nonetheless, to provide an overarching explanation of how the framework is technically structured and how key aspects of the framework were implemented, we concisely describe the architecture of an authored TrainAR training first, and then based on this frame of reference, illustrate how a TrainAR training is deployed by the tool. Then, we show how the two main authoring tasks are implemented: the TrainAR Object conversion process and the TrainAR Statemachine.

Architecture of a TrainAR Training

As visualized in Figure 6.19 conceptually, an authored TrainAR training, which is started on a handheld device and completed the technical onboarding screens, initially starts with the **Prefab Spawning Controller**. This controller handles the trainees' placement of the training assembly in the physical environment and signals the **Onboarding Controller** when and where sufficiently large surfaces were detected for the placement of the training assembly. After the assembly is placed, the **Interaction Controller** is triggered and starts to listen to the **Interaction Button Controller**. When a Button (e.g., interaction or combination) is pressed, while a **TrainAR Object** is selected, the TrainAR Object sends a request to change its state in the form of a **State Information** struct to the **Statemachine Connector**. The Statemachine Connector then hands this State Information to the **Visual Statemachine**, which checks this information against the desired states set by the author of the training and answers to the Statemachine Connector if this request was a valid or invalid request. On the one hand, the Statemachine Connector then hands this information back to the TrainAR Object, which can then trigger object-level consequences based on this decision (e.g., shading, audio playbacks, outlining, animations, Custom events or physics) using the TrainAR Objects controllers (**Material Controller**, **Collision Controller**, **Audio controller**, **Rigidbody Controller**) or custom behavior attached to the objects through the event system. On the other hand, the Statemachine Connector also triggers the flow-level consequences, e.g. using the **Questionnaire Controller** to display the quiz elements of the TrainAR interaction concept, the **Top Panel Controller** to display new instructions, the **Error Overlay Controller** for the layered feedback, or the **Direct Info Controller** to display insights, when requested by the author of the training in the Visual Statemachine as a consequence of the trainees actions.

This architecture is automatically deployed with every TrainAR training and is initially static in its structure. Therefore, as long as no custom behaviors, custom action, or C#-level changes to the framework were made, it is the same for every deployed TrainAR training, which was authored with the authoring tool. On the one hand, this enables distribution considerations for TrainAR trainings (which will be discussed in the following Section 6.6.6). On the other hand, this necessitates only two remaining central technical challenges for the authoring tool to solve: Enabling the author to import TrainAR Objects and then allowing the author to create a visual

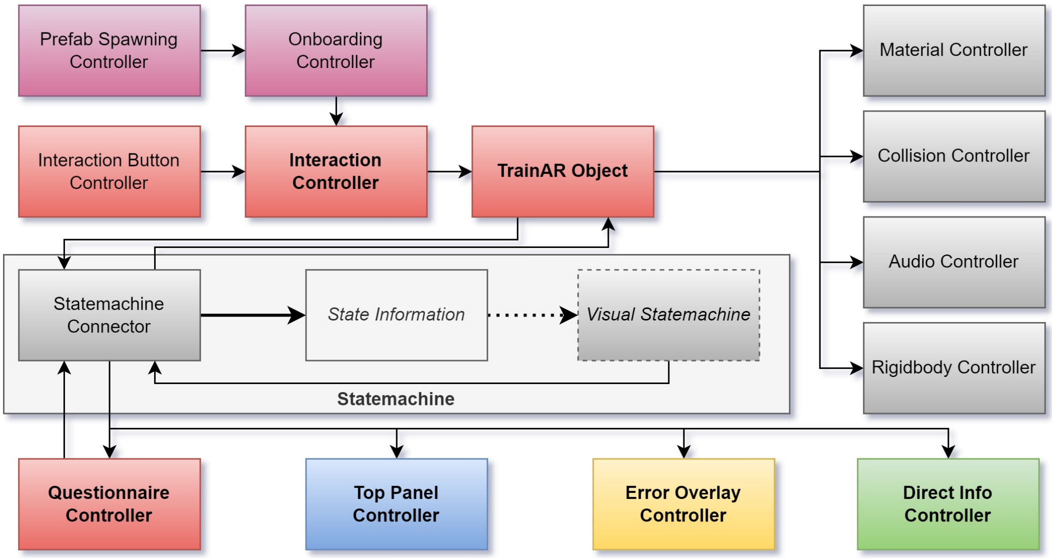


Figure 6.19: An abstract conceptual visualization of the architecture of an authored TrainAR training.

flow of states, formulating a Visual Statemachine. Afterward, these two components from the author of the training are bundled and deployed with the predetermined technical architecture of a TrainAR training.

TrainAR Object Conversion Process from a Technical Perspective

When the user imports 3D models and then clicks the "Convert to TrainAR Object" button in the Authoring Tool, from a technical perspective, a series of steps are performed. First, the authoring tool ensures that the object to be converted actually is in a format that could be converted (e.g., having a Transform component, a MeshRenderer, a MeshFilter, and actually having at least one mesh attached), and it checks if the object was not already converted into a TrainAR Object. Furthermore, some Unity-specific considerations are taken into account, e.g., if objects are currently packed into so-called Prefabs (object bundles), they are unpacked. Then it collapses the entire substructure of the GameObjects hierarchy and combines all meshes [457] and sub-meshes into a single GameObject with a single mesh, but multiple Materials attached to it. Afterward, the conversion process opens the modal window (see Figure 6.18), where the author can use a slider to, based on vertices and polygon count and a live preview of the resulting object, simplify the Mesh. With every update to the sliders' position, the combined mesh is simplified based on the Fast-Quadric-Mesh-Simplification algorithm [122] and a textual assessment of the mesh is provided. When the author specifies a name, descides on how much mesh simplification they wanted, and clickes "Convert to TrainAR Object", the conversion is finalized. In the finalization stages, the mesh is renamed, tagged to be included in TrainAR builds, and all standard TrainAR object-level behaviors are attached to the GameObject as C# scripts. These are needed for the object to elicit the standard TrainAR behaviors mentioned in Section 6.5.2. Afterward, the complete performed chain of functionality is registered as a single performed action in the Undo class of the Unity Editor, to enable authors to revert changes.

The Visual Statemachine of the Authoring Tool

While we call TrainARs Visual Statemachine a "statemachine", which is correct from the theoretical perspective, from a technical perspective, it is not a statemachine but actually a combination of a statemachine and a so-called scriptmachine [468]. A scriptmachine normally does not store states, is not concerned about the transition period between states, and would by default simply execute the sequential order of all scripts specified in its flow, until the scriptmachine reaches terminating states for each of the active flows. As we wanted to keep the design of the "TrainAR Statemachine" as close as possible to known concepts, we are also not concerned with the transition period between states and do not want to manually trigger each of the transitions. E.g., if an action is performed, effects like an update on object-level and the textual instructions on the top UI element should happen sequentially, but automatically, in the authored training. On the other hand, we do need to wait for actions (inputs of the trainee) and therefore need a hybrid of the script and statemachine. We realized this by adapting Unitys scriptmachine as a basis, and implementing custom visual scripting nodes (see Figure 6.17) for all the input and output behavior. The output behavior (like instructions, object-level changes, helper utility, or displaying hints) largely functions in line with the design of the existing scriptmachine concepts, with the only exception being that reaching an empty connection does not terminate the flow but simply waits for another action on the last input node. The input nodes themselves (therefore, nodes that wait for the trainee to perform an action and continue the flow based on their decision into a specified direction), on the other hand, only execute partial functionality initially and then register their position and state inside the flow to the Statemachine Connector. Technically, they now terminate the current flow. When an action is performed by the trainee, the flow is restarted through the Statemachine Connector to check the requested state change against the values input into the node by the author. The Visual Statemachine then executes the remainder of its internal functionality and afterward triggers one of the connections to continue the flow, based on this information.

Importantly, all underlying functionality and behavior of the TrainAR Statemachine is stored in C# scripts for each of the 10 visual scripting nodes (see Figure 6.17) that are always deployed with every training. In this, the graph of the visual statemachine (see Figure 6.16) does not hold any actual C# functionality, but holds the logical sequence to execute and the parameters to execute the C# specified functionality with. Similar to Unity's scenes serialization, the visual scripting graphs are serialized as YAML[41] descriptions on the chain of functionalities.

6.6.6 Distribution of TrainAR Trainings

Combining the fact that the TrainAR visual statemachine is simply a YAML serialized description of which C# functionality (that is already delivered with each TrainAR training) to execute in which order and with which variables, with the fact that TrainAR objects on an abstract level are meshes with materials/textures, and some additional state information with consistent behaviors across objects, considerations towards the distribution of trainings, that do not implement functionality beyond the functionality scope of TrainAR, are fairly simple. Besides the obvious choice of using the Unity-specfic implementations of "addressables" to share content, this also allows for more generic, simplistic approaches that work without servers, making them ideal for sustainable open-source publication. This architecture is primarily possible, as no C# functionality has to be actually deployed, but rather the sequence of small modular C# nodes to execute can be specified in the serialized stateflow. As a first step in currently ongoing technical efforts, we are simply trying to serialize each TrainAR Objects mesh data and all its relevant state information (position, rotation, scale, grabbability, interactability, combinability, visibility, and name of the object) into a single XML file and supplement the XML file with all materials as ".png" files. Then we combine all objects serialized this way with the statemachine file, which is already in YML format, and compress them into a .zip file, which is subsequently transmuted into a ".trainAR" format. In a next step, this could then, for example, enable the possibility to not only deploy the build trainings onto a connected device directly as binaries in Unity, but also alternatively allow building a ".trainAR" file and uploading it to a server. In this minimal viable implementation of server distribution functionality, trainees could then download serialized trainings from within a "host" TrainAR app, which is publicly available through the app stores, by providing URLs. While only a first step and limited in several ways, this would already substantially simplify distribution.

6.6.7 Beyond TrainARs Statemachine: Expanding on the TrainAR Framework

Besides these distribution aspects, in anticipation that the TrainAR Statemachine, while being the factor that enables domain experts without programming expertise to utilize it, would also be the most limiting factor for more experienced users and programmers trying to implement more context-specific requirements, we deliberately chose to develop TrainAR's authoring tool in the form of a Unity extension and expanded upon the Unity Visual Scripting Package [467] for the visual TrainAR Statemachine.

This approach allows for an expansion of the TrainAR authoring tool in several directions. Foremost, the custom Action node allows for triggering state changes with a parameter from a MonoBehaviour manually, allowing for user actions besides grabbing, interacting, and combining out of the box. Additionally, nodes provided by Unity's Visual Scripting package are completely compatible with all TrainAR nodes, making it possible to integrate them into stateflows for more complex behaviors in terms of the flow of actions, therefore providing stateflow-level expansion possibilities for TrainAR. For the expansion of TrainAR on the object level, when switching into the Unity Editor layout, all MonoBehaviours of converted TrainAR objects are exposed, and object-level Events, e.g., for this specific object being selected, interacted with, grabbed, or combined, are exposed as UnityEvents and can be used to implement more complex object-level behaviors such as animation triggers or object-specific MonoBehaviour C# scripts that trigger event-specific custom behaviors. Finally, if authors want to use the interaction concept and technical onboarding utility of the framework in non-procedural training contexts, e.g., for conceptual training games, rule-based stateflows, or simply want to program stateflows themselves, the visual statemachine can be switched off entirely by simply commenting out a single line of code in the Statemachine connector (see `https://github.com/jblattgerste/TrainAR/blob/main/Assets/Scripts/Static/StatemachineConnector.cs`) and handling the requests of the state change function manually through C# scripting.

6.7 Discussion

From both, the conceptual but also the technical perspective, we designed TrainAR from a practical case study towards a generalizable framework. Therefore, we first developed a TrainAR training, the preparation of a tocolytic injection, which uses the TrainAR interaction concept and formulated didactic ideas around it in the midwifery Heb@AR context. Then we refined the interaction concept through iterative usability studies. Afterward, when we were happy with the framework after subsequent improvements, we generalized the didactic ideas and deployed TrainAR trainings into additional contexts with the help of external researchers. When we were happy with the results of those TrainAR trainings as well, we developed the TrainAR authoring tool, which as open as possible, as pragmatically as possible, and ideally in a form that is already familiar to potential users, allows for the creation of such TrainAR trainings.

This authoring tool of the TrainAR framework allows domain experts to create their own procedural AR trainings by utilizing the TrainAR interaction concept, including its onboarding and technical utility, allowing authors to focus on the content of the training, and not having to worry about the technical aspects or AR-specific implementation challenges. Authors, if they choose to do so, can follow our didactic consideration framework but can also implement their own didactic ideas and approach the creation of their TrainAR trainings independently. With TrainAR being completely free, published as open-source under the MIT license, being fully documented, and using handheld AR devices as the target hardware, the created procedural AR trainings are realistically scalable today. This enables bring-your-own-device methodologies, self-directed learning, location-independent learning, and self-paced preparation or retention opportunities through interactive AR trainings, which are engaging and incorporate psychomotor learning components [52]. This pragmatic and, more importantly, holistic approach to the authoring of interactive AR trainings is novel in the literature and should not only help the efficiency of the authoring process but also contribute in other ways.

Foremost, it provides guidance for the created trainings to follow proven principles, others can use and expand upon. Secondly, we think that having established a first set of principles will also help to resolve the causality dilemma of not knowing what an AR training could look like, before even starting to author them, but also not being able to develop an AR training before having an extensive process description of the training. A learning we experienced first-hand in the Heb@AR context (see Chapter 4). Finally, while HMD-based AR hardware is still not readily available, the expert development, formalization of flows, didactic considerations and even the technical implementations of the TrainAR Statemachine with its TrainAR Objects, can easily be transferred to new hardware. In this, the trainings can be developed now and are scalable immediately with currently available hardware, but are also future-proof and can be reused down the line for HMD-based systems. At last, this is ensured through the open-source availability of the entire framework, and its licensing under the permissive MIT license, which allows distribution, modification, and commercial use.

As discussed in Section 6.6.1, we deliberately chose to implement the TrainAR authoring tool as a Unity extension and developed the TrainAR Statemachine and UI layout to be extendable and replaceable. Ultimately, programmers can utilize the full Unity functionality seamlessly with the framework. Where the TrainAR scope and documentation ends, there are a good number of documentation, tutorials, and getting-started utilities on how to use the Unity engine. In this context, TrainAR should enable and accelerate development, but never be a limiting factor. While the authoring tool does not require programming skills and is designed as a low-level content-design framework, we are aware of the inherent trade-off of this introducing interface complexity, compared to standalone authoring tool approaches but believe that this trade-off is merited.

This pragmatic, open perspective continues in the didactic consideration aspects of the framework. From our first explorations into different contexts and the work of the partners on their TrainAR trainings, we expect the utility (see Figure 7.1) and therefore inherently also the didactic perspective on the utilization of the TrainAR trainings to differ significantly based on the context-specific needs. Therefore, while we provide didactic considerations [55], they are more of a "didactic cookbook" for authors to use according to their own tastes than strict rules we envision for the framework to follow.

In its current stage, the gathering or generation of 3D assets to use in the training likely remains the biggest technical challenge for authors. On the one hand, there are multiple ways to obtain assets, ranging from online asset stores, over free online databases, to 3D scanning. This makes it challenging to evaluate this aspect systematically. On the other hand, and more importantly, we believe that the asset generation will be increasingly simplified with technological advancements, as already today, modern smartphones can create acceptable 3D assets through LiDAR scanning, and zero-shot 3D asset generation through AI (Artificial Intelligence) is showing promising recent scientific results [206]. We believe that, eventually, the 3D asset generation, and therefore the TrainAR Object aspect, will become decreasingly challenging, and we therefore explicitly designed it with this development in mind.

6.7.1 Current & Future Work

Besides our efforts to continuously evaluate and iteratively improve the state of the authoring tool's source code for the current scope of the TrainAR framework as stated in Section 7.5.5, we also plan to expand TrainARs ideas and perspectives in several directions and hope for future cooperations and third-party contributions to our open-source project.

As stated in Section 6.6.6, one important technical aspect we are currently addressing is training distribution aspects. Here, we are exploring how we could improve the deployment and distribution aspect of authored TrainAR trainings for users during the development phase and formative testing but also the actual deployment afterward, as currently, when authors create trainings, they would have to either manually distribute the training as compiled apps for Android or iOS or would have to publish them into an asset store, which, in our experience, is a daunting endeavor.

Furthermore, we are currently working on a higher-level onboarding utility for the TrainAR Authoring Tool. While the extensive documentation already includes getting-started guides, examples to utilize, and explanatory videos, we are currently working on video material in the "tutorial" format, which are more specifically targeted at potential users that are neither programmers nor media technologists, in the hopes to lower the barrier of entrance without limiting the expandability of TrainARs authoring tool.

Additionally, we plan to work on a didactic white paper that specifies more clearly on an abstract level, what we believe specific components of the framework can achieve and how we envision them to be used (e.g., feedback nodes, insights, instructions, or quiz elements). While this is discussed in detail from a scientific perspective in Section 6.4 and summarized in the technical documentation, we want to make this information more approachable and provide it in formats in line with domain experts' expectations.

We are also interested in expanding TrainAR towards incorporating physical object and marker tracking, to not only use virtual objects but also integrate physical objects into the flow during the training. While this was explored during development and is not particularly challenging from a technical perspective, as tracking libraries for this already exist, considerations must be made for the didactic concept and how to integrate the physical material seamlessly into the overall holistic framework, as this might impact not only the usability but also the current strengths of TrainARs' ideas of location-independent learning, material/cost-savings, and immediate scalability of authored AR trainings [55].

Finally, while interaction concepts that can be used for procedural trainings such as the MRTK exist for AR HMDs, we would like to explore if expanding TrainAR towards including HMDs as a target platform would be feasible and viable, as not only the visual representation of the flow of state but also some didactic considerations and modules could significantly add to the state of AR authoring tools in the HMD-based training context. Additionally, this comes with the synergy effect, that it would make TrainAR trainings immediately cross-platform. If we replace the interaction concept related components of the architecture of a TrainAR training (see Section 6.6.5), which is not technically challenging based on our modular design of the framework, all previously authored TrainAR trainings would immediately work on all newly supported hardware.

6.8 Summary

In this chapter, the TrainAR framework was introduced. First, the interaction concept was described from a theoretical perspective. Then it was described from one exemplary practical implementation from the context of procedural task training in academic midwifery during project Heb@AR, including the formative usability evaluations with subsequent improvements to the interaction concept, that led to its final design. Then the didactic framework was proposed, including guidelines on how procedures could be transformed into a TrainAR training. Finally, the framework components were discussed from a technical perspective, and a visual scripting-based authoring tool was proposed for the TrainAR framework: The TrainAR authoring tool. The authoring tool has been published as an open-source project (https://github.com/jblattgerste/TrainAR, accessed: 31.03.2023), containing the full source code for the TrainAR interaction concept and authoring functionality as a Unity extension under the MIT license. In addition to a full documentation (https://jblattgerste.github.io/TrainAR/, accessed: 31.03.2023), which is included with the authoring tool, getting-started guides, and tutorials are available and several trainings are already developed, published in App stores, and documented in reference videos.

7 TrainAR Evaluations

"Simple can be harder than complex: You have to work hard to get your thinking clean to make it simple." — Steve Jobs

Evaluating TrainAR's authoring tool inherently means evaluating TrainAR holistically as a framework for the creation and utilization of digital, procedural AR trainings. Consequently, this requires the evaluation of several of its components individually, making the evaluation challenging and extensive. Additionally, simple lab studies with preliminary prototypes would likely not suffice to evaluate TrainAR's most important aspects, or might even be misleading based on our perspective and usage vision. While extensive evaluations are ongoing, the following two sections, 7.1 and 7.2, provide preliminary insights into our current results. We believe that four questions have to be answered from the perspective of somebody trying to utilize TrainAR to entice them to apply it to their context:

1. Do TrainAR trainings elicit learning benefits (e.g., increased retention, conceptual understanding, or motivation, or providing self-paced learning opportunities)?
2. Are TrainAR trainings usable by and enjoyable for the trainee?
3. Is the TrainAR Authoring Tool usable by non-programmers to create such TrainAR trainings? More specifically,
   a) What is the required level of media competency, and who can realistically utilize the TrainAR authoring tool?
   b) How fast can the usage of the tool be learned, and what training or tutorial material is necessary?

These questions are in line with the accepted User Experience design principle *Utility + Usability = Usefulness*, which conveys that a product has to provide utility and be usable by the target group to be a useful product. We believe, in this specific case of the AR authoring tool, that this principle has two levels (see Figure 7.1). First, the Utility and Usability of the TrainAR training has to be shown to prove them to provide a useful training. If this is true, the utility of the AR authoring tool would consequently be a possibility for creating a useful AR training. Then, it has to be shown who can use the AR authoring tool to create these trainings, as usability is as dependent on the target user as it is on the implementation.

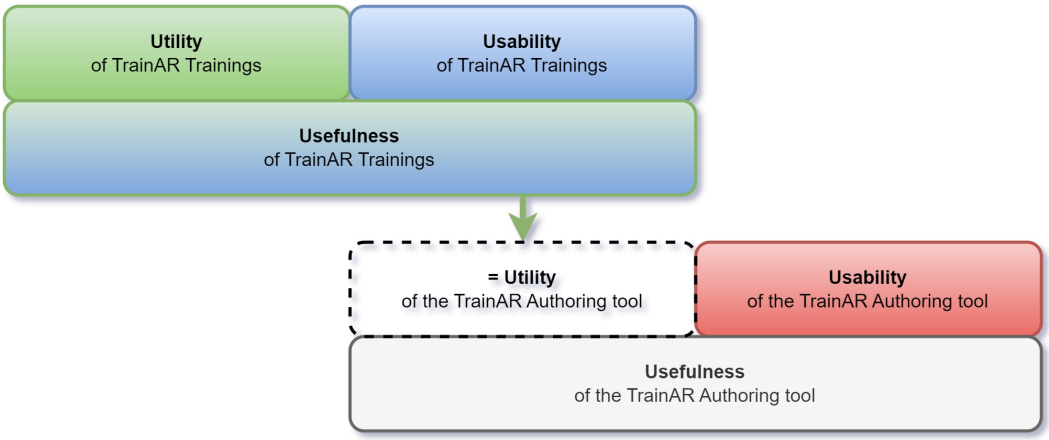


Figure 7.1: To show the usefulness of the TrainAR authoring tool, its utility and usability have to be evaluated. The utility of the authoring tool itself is the creation of trainings that themselves have to be useful, meaning they also have to prove their utility and usability.

7.1 Utility and Usability of TrainAR Trainings

To answer research questions 1 and 2, several TrainAR trainings are currently in development or were developed and evaluated using the TrainAR framework in different contexts. While evaluations have not concluded for all the trainings, five exemplary TrainAR Trainings are shown in Figure 7.2, and their utility and usability evaluations are described below. For usability assessment, the System Usability Scale (SUS) was used to make results comparable across the trainings (see Figure 7.3) and the System Usability Scale Analysis Toolkit [50] was provided as supporting utility (see Section 4.5). As the desired utility of each TrainAR training and its evaluation is highly dependent on the context, we only discuss the utility on an abstract level and refer to the authors' publications and the previous Heb@AR Chapter for more detailed insights and discussions.

Additionally, although TrainAR was originally envisioned as a holistic solution combining an interaction concept, didactic framework, and an authoring tool (see Figure 6.2), as already discussed in the previous chapter, each of the three components can also be used separately. It has to be noted that the scenarios shown in this section do not necessarily use each of the components. While all of them use the TrainAR interaction concept, the training of preparing a tocolytic injection was the starting point for the TrainAR framework abstraction and therefore was developed from scratch, not utilizing the TrainAR authoring tool. The denomination of the female pelvis, a game exploring the sourness of fruits, and the game for exploring ripeness all use the interaction concept and authoring tool but not the didactic framework, as they are envisioned more as rule-based learning games than strictly procedural trainings. Only the titration experiment utilizes all three components, though it also has to be noted that the authoring tool utilized in all trainings was in early preliminary stages, e.g., not including visual scripting and still requiring programming (see Table 7.1). The main focus in this stage was the evaluation of the created trainings, not the authoring tool, which was evaluated separately (see Section 7.2).

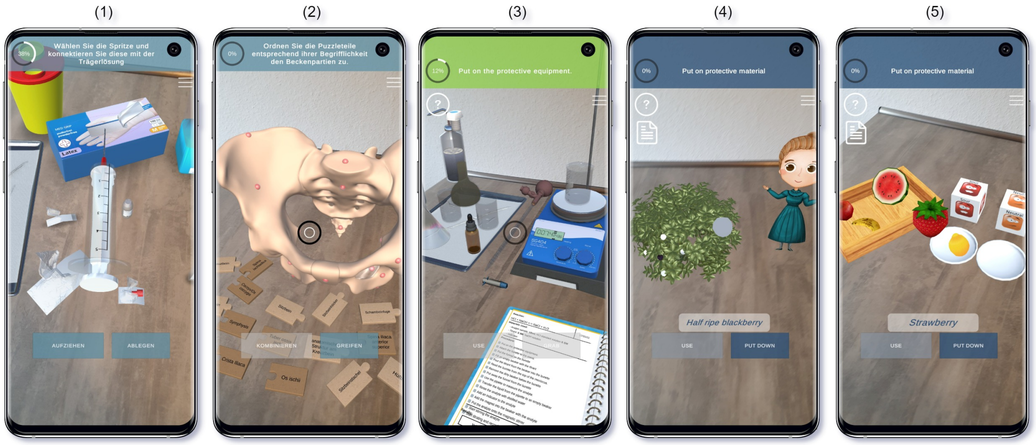


Figure 7.2: Five exemplary TrainAR trainings. (**1**) The preparation of a tocolytic injection in the context of academic midwifery, (**2**) the denomination and contextualization of German and Latin terminology of the female pelvis, (**3**) a titration experiment in the context of chemical engineering, (**4**) the exploration of chemical reactions in early school education, and (**5**) the exploration of ripeness as a chemistry learning game for children.

TrainAR Scenario	**Interaction Concept**	**Didactic Framework**	**Authoring Framework**
(1) Preparation of a Tocolytic Injection	✓	✓	
(2) Denominating the Female Pelvis	✓		✓
(3) Conduction of a Titration Experiment	✓	✓	✓
(4) Exploring Chemical Reactions	✓		✓
(5) Understanding Fruit Ripeness	✓		✓

Table 7.1: The five exemplary TrainAR trainings shown in Figure 7.2 and which parts of TrainAR (Interaction concept, didactic framework, or authoring framework) they utilize for their use case.

7.1.1 Preparation of a Tocolytic Injection

In the context of the Heb@AR project, a procedural TrainAR training was developed for the preparation of a tocolytic injection in the context of academic midwifery education (see Figure 7.2(1)). Here, the user elicits a sequence of actions to prepare a tocolytic syringe that is labeled and stored in a fridge, which is a common task in the daily midwifery routine [54]. For this purpose, the user has to interact with objects and grab, place, and combine objects while being instructed and tutored by a virtual professional midwife [55].

The desired utility of the training is an opportunity for self-directed, location-independent learning and an increase in self-efficacy for midwifery students [54]. While the more detailed results for the evaluation of the utility are forthcoming, preliminary analyses and qualitative feedback look promising. As reported in Section 4.6, the training did significantly increase student's self-reported perceived competency directly after conducting the training, qualitative feedback in learning journals indicated that some students did perceive an impact on their competency during their practical phase, and the OSCE exam results, though not statistically significant, show a positive tendency of improved exam results for the cohort which used the training. To measure the perceived usability of the training, the SUS questionnaire was used. A SUS study score of 83.11 (SD = 12.9) was reported (see Figure 7.3), which would indicate "Excellent" usability according to Bangor et al. [27] and surpasses the non-empirical, but commonly used, industry benchmark of SUS study scores of 80 [278]. With a sample size of $n = 33$ participants, the results are 100% conclusive, according to Tullis et al. [463].

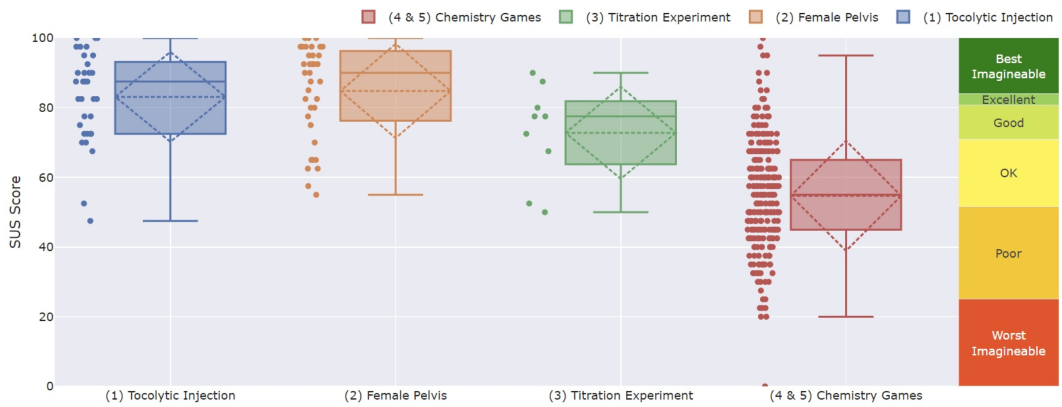


Figure 7.3: The perceived usability of the exemplary TrainAR scenarios in Figure 7.2 in the form of SUS study scores, taken from primary sources evaluating TrainAR trainings [18, 52, 55, 114] and plotted with the SUS Analysis Toolkit [50].

7.1.2 German–Latin Denomination of the Female Pelvis

Likewise, in the midwifery education context of project Heb@AR [54], a learning game for the denominating of the female pelvis [52] was developed (see Figure 7.2(2)). Here, the idea is to use TrainAR's gamification aspects to make the traditionally dry subject of learning all German and Latin names and their contextualization for the bones and regions of the female pelvis more enjoyable to students. The user has to grab and combine pieces of a puzzle with the Latin and German names with each other, and then contextualize them to corresponding bones and regions of the female pelvis.

In terms of utility, it was found to increase the students' intrinsic motivation to engage with the historically dry subject significantly, which was measured through a within-subject comparison using pre- and post-study questionnaires [52] (see Section 4.6.5 for more details). For usability, a SUS study score of 84.79 (SD = 13.51) was reported (see Figure 7.3). This would not only be interpreted as "Best Imaginable" Usability according to Bangor et al. [27] and surpass the non-empirical industry benchmark of 80 [278], but is also the highest recorded SUS study score of a TrainAR training recorded to date. The sample size of $n = 36$ participants should yield 100% conclusive results, according to Tullis et al. [463].

7.1.3 MARLabs Titration Experiment

In the context of academic chemical engineering education, Dominguez Alfaro et al. [114] from KU Leuven developed a TrainAR procedural training where students, preparing for their actual physical lab titration experiments as part of the curriculum, can train the necessary procedures of titration experiments beforehand. They can use their smartphone to combine chemicals, follow safety procedures, and document their experiment accordingly (see Figure 7.2(3)).

For the usability, a SUS study score of 72.8 (SD = 14.0) [114] was reported (see Figure 7.3), which would indicate above average or "Good" usability on the adjective contextualization scale proposed by Bangor et al. [27] and is an acceptable usability score [28]. According to Tullis et al. [463], this result is only between 75–80% conclusive, based on the small sample size of $n = 9$ participants. The desired utility of the training was an increased understanding of the users' knowledge of acid–base titration concepts. Likely because of the small sample size, the initial study failed to show statistically significant learning effects, but results from larger studies are forthcoming [114]. Nonetheless, Dominguez Alfaro et al. [114] could observe that the app was "well-received by the users", and they were able to independently download and utilize it in a remote experiment setting without an experimenter present.

7.1.4 Exploration of Fruit Ripeness and Sourness

Finally, for the K-12 chemistry education context, learning games for the exploration of fruit ripeness (see Figure 7.2(4)) and exploration of the sourness of fruits (see Figure 7.2(5)) were developed for iOS tablets by Arztmann et al. [18] at Utrecht University, using TrainAR. Dutch children aged between 11 and 15 used the application as part of their curriculum to have playful first points of contact with chemical principles such as ripeness and sourness by, for example, feeding beets with different levels of ripeness to a virtual avatar or analyzing fruits based on their sourness, using a pH strip, and then sorting them.

The intended utility of the training was the possibility for students to independently engage with these new concepts playfully and at their own pace. Therefore, the idea was "triggering students' interest in chemistry by providing a playful environment with relatable content" [18]. While this is challenging to quantify, it was observable that the students were able to independently utilize the game and were enjoying the experience. A non-validated Dutch translation of the simplified SUS questionnaire by Putnam et al. was used [383] to measure the perceived usability for this usage group. The resulting Dutch simplified SUS questionnaire had low internal consistency, with a Cronbach's alpha of 0.446. The calculated SUS study score of 54.7 (SD = 15.19) (see Figure 7.3) would be interpreted as "OK" [27], but below average, perceived usability and indicates only "marginally acceptable" usability according to Bangor et al. [28]. Besides the internal consistency issues, a SUS study score with n = 239 participants should be conclusive based on the sample size, according to Tullis et al. [463]. With low internal consistency, children as the target group instead of adults, and the usage of iOS tablets instead of smartphones as the delivery method, it is hard to determine where this low perceived usability, compared to the other TrainAR trainings, originates. It might be possible that, children, who are not the originally envisioned target group [55], require additional considerations [18]. Additionally, interaction effects are possible. These perceived usability results should therefore be interpreted with caution.

7.2 Usability of the TrainAR Authoring Tool

To answer research question 3, we carried out multiple steps. First, we shared the framework with researchers from Utrecht University and KU Leuven in 2020 for them to deploy it in their contexts; then, we iteratively used the authoring tool in two practical lectures to observe its usage in a non-representative setting. After this indicated sufficient maturity of the authoring tool, we conducted a systematic study to determine the tool's usability and to assess the required media competency.

7.2.1 Pre-Study and Non-Representative Observations

Initially, we shared early versions of the TrainAR framework with other Universities in 2020 to deploy them to their contexts. During this process, the trainings described in Section 7.1 were created. The TrainAR versions used by those collaborating researchers were early builds, e.g., not including the Visual Statemachine and providing a less convenient object-conversion utility. The researchers, while not computer scientists, had experience with programming. While this provided valuable first insights into the feasibility of TrainAR's set of utility and the effectiveness, usability, and enjoyability of the authored TrainAR trainings, these insights were not representative of the usability of the authoring tool itself.

Afterward, early versions of the authoring tool, then already including the full TrainAR Visual Statemachine functionality, were used during the practical part of an apprenticeship course to obtain preliminary insights into the usage of TrainAR by the main target group of the framework: domain experts with high levels of media competency but without programming knowledge. In this course, eight apprentices created four TrainAR scenarios of their choice through the course of four practical sessions, each lasting around 2 h. Throughout this course, the apprentices chose to create scenarios for the installation of a desktop computer set, the finishing work after 3D printing mechanical components, cutting and filing a workpiece, and the preparation of a steak with bacon and eggs. Besides some smaller hurdles and anticipated bugs, which could be either resolved by consulting the teaching assistants present or through smaller technical adjustments to the source code of the framework, the apprentices were able to create procedural action chains using TrainAR's authoring tool. They were even able to incorporate 3D scanning for model generation. The observations and feedback provided showed that the authoring tool was sufficiently usable for them. This indicates that it should also be usable for the envisioned target group and that it is possible to independently create TrainAR trainings for them. The most challenging aspect was the 3D model generation or gathering and the didactically conceptual, but not technical, chaining of instructions, actions, and feedback mechanisms. Their feedback furthermore highlighted that good documentation and especially in-depth onboarding and "Getting Started" utility would be helpful. After improving the documentation, the conversion utility, fixing bugs in the source code, and publishing TrainAR on GitHub, this procedure was repeated with another course of 12 students, and then six scenarios were created over the course of four practical sessions that were 2 h each. Here, students were again able to successfully create trainings with the authoring tool.

7.2.2 Systematic Usability Study Design

After the second pre-study iteration was successful and indicated that most major problems had been addressed, we conducted a systematic usability study, with a focus on the pragmatic qualities of the authoring tool and the required media competency to use the TrainAR Authoring Tool. While ideally the usability evaluation would be conducted with actual users of the authoring tool and correlations between the pre-existing media competency recorded through standardized tests would be investigated, this is challenging in practice for multiple reasons. Firstly, actual users are challenging to recruit for the study, because of resource and time constraints. Then, those users would have to be recruited systematically and in high numbers, so there are actual differences in media competency. Finally, systematic assessments of media competency often rely on self-reported measures and mostly focus on computer literacy, which would likely not be sensitive to differences in the media competencies we are interested in. Therefore, the study was designed as a between-subject comparison with students as participants from three groups: Computer Science (CS) students, Media Technology (MT) students, and non-technical students from our university. Before the experiment, we asked participants to self-assess their competency in 3D modeling, programming, and VR/AR/Game development. During the study, participants had to author three trainings based on provided 3D models and stateflow descriptions in line with task process analyses, totaling 47 sub-tasks to complete the study. The 47 sub-tasks consisted of 10 types/categories of tasks, e.g., placing or converting an object, placing an action node in the Visual Statemachine, or placing an instruction node. The three authoring tasks hereby increased in complexity, with the first one (mounting a lightbulb in a socket in order to subsequently switch it on) being a simple, linear flow of actions (the conceptual flow of the authoring task is visualized in the Appendix 1), the second task (a re-enactment of the East Frisian tea ceremony) introducing quizzes and UI elements (see Appendix 2), and the third task introducing non-linear flows of states (attachment of a needle to a syringe and subsequently filling it with medication), as visualized in Appendix 2. During the experiment, we recorded the Task-Completion Time (TCT) and Task-Completion Rate (TCR) for each of the 47 sub-tasks and cognitive load (NASA rTLX) [188] after each of the three tasks. After the experiment, we measured the perceived usability using the System Usability Scale (SUS) [65], which is one of the most widely used usability questionnaires and provides direct benchmarking and contextualization utility [50]. Finally, we asked the subjects for qualitative feedback to self-assess their ability to independently create AR trainings using TrainAR.

7.2.3 Setup and Procedure

A desktop computer (AMD Ryzen 7 5800X, 64 GB of ram, Nvidia GeForce 3080 Ti) with two 30-inch monitors and a stand microphone was used for the experiment. The experimenter was sitting beside the participants during the study and took notes about TCT and TCR, and audio was recorded during the experiment.

After greeting the participants and explaining the study, they were asked to fill out a pre-study questionnaire. Here, they filled out a declaration of consent, a demographic questionnaire, and a questionnaire on their relevant previous knowledge. Afterward, they were given a brief introduction to TrainAR in the form of a 4-min explanation video. This video explained the basic features and functionalities of TrainAR and its general use. Furthermore, the participants were given a short verbal introduction to the documentation of TrainAR and were encouraged to use it during the study. Then, participants used the TrainAR Authoring Tool to implement the three pre-defined TrainAR trainings. The tasks to create the trainings were divided into 47 sub-tasks, which were presented to the participants sequentially in a Google Forms document. For each of the sub-tasks, the desired end result was shown either as a short video clip or an image and their intended connection in the flow of states as a task-process-analysis-inspired visualization. The participants then had to author each sub-task on their own. In case they needed help, they were allowed to ask the experimenter for hints or help. These hints and the help were given systematically on four levels. Each of the sub-tasks had a documentation hint, meaning a pointer to the relevant passage in the TrainAR documentation. This hint was given as a first measure, should the participant run into problems. If this hint did not help, the participant was provided with a solution hint, meaning a pre-defined hint for the given sub-task that was read out to the participant. If participants were unable to solve the sub-task with this hint, they were explicitly helped by the experimenter. This ensured that each participant was exposed to all 47 sub-tasks during the experiment, which built upon each other. After a participant completed all sub-tasks of this authoring task, they were asked to test the scenario on a provided smartphone and then filled out a post-task questionnaire, containing a perceived cognitive workload questionnaire (NASA rTLX) [188] regarding the just-completed scenario. Before starting with the authoring of the next training, the participants were offered a short break and snacks. After all three authoring tasks had been completed in this fashion, the participants were finally asked to fill out the post-study questionnaires, containing the SUS, as well as a qualitative feedback questionnaire asking them what they liked, where they had problems, and to assess if they would be able to create an AR training independently.

7.2.4 Participants

Overall, 30 participants took part in the experiment. Twenty-one of the participants were male, and nine were female. The average age of the participants was 25.13 (SD = 3.24). Participants received monetary compensation for their participation in this study.

The participants were recruited from groups: 10 participants were Computer Science (CS) students, 10 were Media Technology (MT) students, and the remaining 10 students were from various non-technical study programs (i.e., business administration and social work studies) from our university. To validate if their self-reported competency matched our expectation of the groups, participants were asked to rate their experience in 3D modeling, programming, and VR/AR/game development on a seven-point Likert scale ranging from 1 ("no experience at all") to 7 ("very experienced"). CS students reported the highest experience in both programming (M = 5.40, SD = 1.11) and VR/AR/game development (M = 3.60, SD = 2.01), indicating they were experienced in programming and somewhat experienced in VR/AR/game development. MT students reported that they were somewhat experienced in programming (M = 3.00, SD = 1.18) and not at all experienced in VR/AR/game development (M = 1.6, SD = 0.92). Non-technical students reported no experience at all in programming (M = 1.10, SD = 0.30) or VR/AR/game development (M = 1.00, SD = 0.00). MT students reported the highest experience with 3D modeling (M = 4.00, SD = 1.00), followed by CS students (M = 3.10, SD = 1.45). Non-technical students again reported no experience at all with 3D modeling (M = 1.10, SD = 0.30).

7.2.5 Results

We recorded the objective measures of TCT, TCR, perceived cognitive load (NASA rTLX) [188], and perceived usability (SUS) [65]. As this study is exploratory in nature, the objective measures of TCT and TCR are descriptively reported on the task level to show trends in the data, but inferentially analyzed and reported at sub-task level across the three tasks to have sufficient power for the statistical tests. The perceived cognitive load is also descriptively reported on the authoring-task level, but the average cognitive load across the experiment is analyzed using inferential statistics.

Task Completion Times

In terms of the Task-Completion Times (see Figure 7.4) of the first authoring task, CS students achieved the fastest average TCT of 3.11 min (SD = 3.39 min), followed by MT students with a TCT of 3.67 min (SD = 4.24 min). With a TCT of 4.86 min (SD=3.84 min), non-technical students were on average the slowest during the first authoring task.

For the second authoring task, MT students achieved an average TCT of 1.33 min (SD = 1.01 min), closely followed by CS students with 1.47 min (SD = 1.03 min). Non-technical students on average needed 1.66 min (SD = 1.08 min) to complete the second authoring task. For the third authoring task, MT students were the fastest, where they on average needed 2.15 min (SD = 3.10 min) per sub-task. CS students achieved an average TCT of 2.21 min (SD = 2.74 min) and non-technical students an average of 2.6 min (SD = 3.45 min).

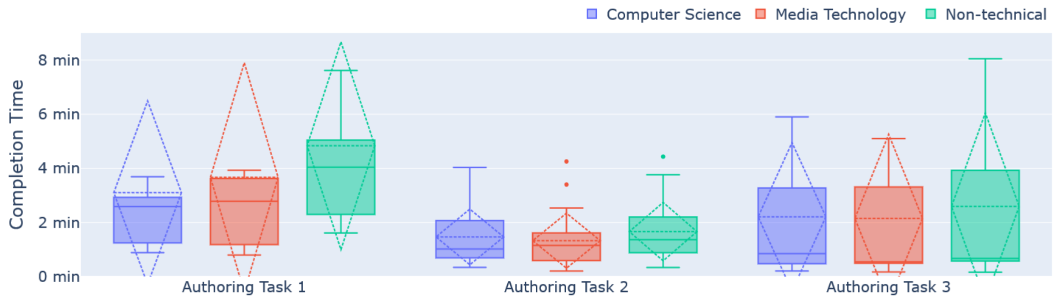


Figure 7.4: Average Task Completion Time (TCT) of each of the student groups (Computer Science, Media-Technology, and non-technical students) for each of the three authoring tasks the participants authored during the study.

Not shown in Figure 7.4 are six outliers. Three outliers in the first authoring task are the first occurrences of the sub-task of the “action” category (see Figure 7.5): CS students on average needed 13.45 min, MT students 16.6 min, and non-technical students 15.8 min to solve this sub-task. The other three outliers not visible are part of the third authoring task and are the first occurrences of the sub-task category “fork-action”. Here, CS students on average needed 9.80 min, MT students 11.96 min, and non-technical students 12.32 min to solve this sub-task.

Figure 7.5 shows the average task completion times for the occurrences of the sub-tasks in each sub-task-category, including the outliers. Notably, TCTs for all sub-tasks decreased not only consistently, but also to a similar degree in each of the student groups when occurring repeatedly.

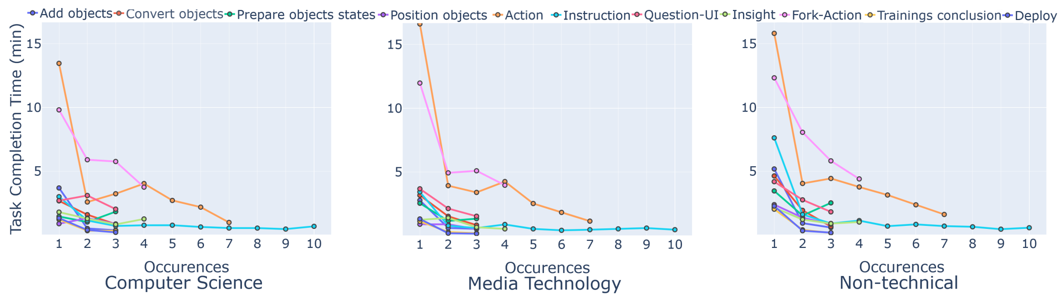


Figure 7.5: The average Task Completion Time (TCT) after each occurrence of one of the ten sub-task types for each of the student groups (Computer Science, Media-Technology, and non-technical students).

As the assumption of normality (Shapiro–Wilk test) was satisfied and Levene’s test considered the populations’ variance to be equal ($p = 0.646$), we conducted an ANOVA to check for differences of the average TCT across all 47 sub-tasks between the groups. The one-way ANOVA revealed no statistically significant differences in average TCT between CS students, MT students, or non-technical students ($F(2,27) = 2.79$, $p = 0.079$).

Task Completion Rates

Table 7.2 shows the average Task-Completion Rate (TCR) of the 47 sub-tasks across the three authoring tasks (11 for authoring task 1, 21 for 2, and 15 for 3) depending on the participant's group. Here, the reported TCR is split into four levels. On the first level, "no help", participants completed the task without any help or hints. On the second level, "documentation hint", participants were given a hint of where in the documentation the solution for their current task could be found. "Solution hint" was an explicit, predefined hint on how to solve the sub-task, which was shown to the participants when they were still not able to solve the task with the documentation hint. If this hint was also not sufficient, participants were helped by the experimenter to complete the sub-task ("explicit help").

While we do not have statistical power or sample size to deploy a two-way mixed-design analysis-of-variance model, there are some interesting descriptive trends that are apparent. For example, in the first authoring task, while in the CS and MT group, participants on average were able to complete over 80% of the sub-tasks without any help, this was only true for 64% of non-technical students. With an average percentage of completed sub-tasks without any help of 96% for the CS group, 98% for the MT group, and 92% for the non-technical students, this gap was narrowed with familiarity with the sub-tasks in the second authoring task. This is until non-linear action chains were introduced into the third authoring task, where the CS group and MT group both retained an average completion percentage of sub-tasks without help of above 90%, while the non-technical group reported the highest average percentage of sub-tasks only completed with explicit help by the experimenter (10%). This was mainly caused by the non-linear "Fork Actions", which the non-technical students struggled with. Here, the majority of them needed explicit help from the experimenter when it first occurred, while only one participant for the CS and MT students needed explicit help. Also notable is the fact that for the CS and MT students, the documentation hint was often sufficient (see the docu hint percentage >= solution hint percentage in Table 7.2), while for the non-technical student's solution hints were required more often (see the solution hint percentage > docu hint percentage in Table 7.2)

Combining the task completion rates of all 47 sub-tasks and interpreting them as ranks ranging from one (completed without help) to four (explicit help of the experimenter), we can check for differences between groups with a non-parametric test. Therefore, a Kruskal–Wallis H test was used. It indicated a significant difference in the TCR between the groups, $\chi^2(2) = 7.63$, $p = 0.022$, with a mean rank score of 10.85 for CS students, 14.2 for MT students, and 21.45 for non-technical students. Dunn's posthoc test using a Bonferroni corrected alpha of 0.017 indicated that the mean rank of CS students and non-technical students was significantly different ($p = 0.007$). The differences between CS students and MT students ($p = 0.39$) and MT and non-technical students ($p = 0.065$) were not significantly different.

Perceived Cognitive Load

To measure the perceived cognitive workload of the participants, the non-weighted version of the NASA TLX questionnaire [188] (NASA rTLX) was used. Participants had to answer the

Help/Hint	Computer Science	Media Technology	Non-Technical
	Authoring Task 1		
No Help	81.90% (SD = 23.00)	80.00% (SD = 18.00)	64.00% (SD = 26.00)
Docu Hint	9.90% (SD = 12.00)	11.00% (SD = 13.00)	12.00% (SD = 10.00)
Solution Hint	4.50% (SD = 8.70)	3.60% (SD = 4.60)	16.00% (SD = 15.00)
Explicit Help	3.60% (SD = 8.70)	5.40% (SD = 4.60)	8.10% (SD = 6.60)
	Authoring Task 2		
No Help	96.00% (SD = 7.80)	98.00% (SD =0 .00)	92.00% (SD = 9.00)
Docu Hint	1.50% (SD = 3.40)	1.00% (SD = 3.20)	1.00% (SD = 3.20)
Solution Hint	0.50% (SD = 1.60)	1.00% (SD = 3.20)	6.70% (SD = 5.90)
Explicit Help	1.90% (SD = 6.00)	0.00% (SD = 0.00)	0.50% (SD = 1.60)
	Authoring Task 3		
No Help	93.00% (SD = 9.70)	91.00% (SD = 9.50)	82.00% (SD = 10.00)
Docu Hint	2.00% (SD = 4.40)	3.40% (SD = 4.70)	1.40% (SD = 3.00)
Solution Hint	2.70% (SD = 4.60)	2.70% (SD = 4.60)	6.80% (SD = 6.30)
Explicit Help	2.70% (SD = 8.50)	2.70% (SD = 6.50)	10.00% (SD = 9.60)

Table 7.2: The average Task Completion Rate (TCR) for each of the three authoring tasks, grouped by the participants' study program and reported on 4 levels: without any help, with a hint of where in the documentation the solution is described, with a predefined solution hint, or with explicit help.

questionnaire items on a 7-point Likert scale. The results were then normalized towards a 1–100 score. This is not the standard application method to evaluate the NASA rTLX but is commonly used this way in the literature to achieve scale consistency while already utilizing Likert scales [94, 409]. Figure 7.6 shows the measured scores for each of the three tasks and participant groups. For the first authoring task, CS students reported the lowest perceived cognitive load with an average score of 31.67 (SD = 12.56), followed by MT students (M = 37.22, SD = 11.12), while non-technical students rated their perceived cognitive load the highest (M = 45.83, SD = 3.45). For authoring task two, CS students reported an average NASA rTLX score of 22.50 (SD = 6.51), MT students reported an average of 33.89 (SD = 10.67), and non-technical students an average of 34.44 (SD = 12.12). This trend continued for the final authoring task, where CS students again rated their perceived cognitive load the lowest (M = 29.77, SD = 12.30), followed by MT students (M = 35.56, SD = 7.11). Again, non-technical students rated their perceived cognitive load the highest (M = 42.24, SD = 16.89).

As the assumption of normality was satisfied (Shapiro–Wilk test) and Leven's test considered the population's variance to be equal (p = 0.39), a one-way ANOVA was used on the average NASA rTLX score of all three measurement points after each authoring task. It revealed statistically significant differences between at least two groups ($F(2, 27) = 3.3924$, p = 0.0485). Tukey's HSD test for multiple comparisons indicated that CS students perceived significantly lower average cognitive load throughout the authoring process compared to non-technical students (p = 0.0392, C.I. [0.5534, 25.18]). No statistically significant differences were found between the MT and CS (p = 0.29) and MT and non-technical students (p = 0.55).

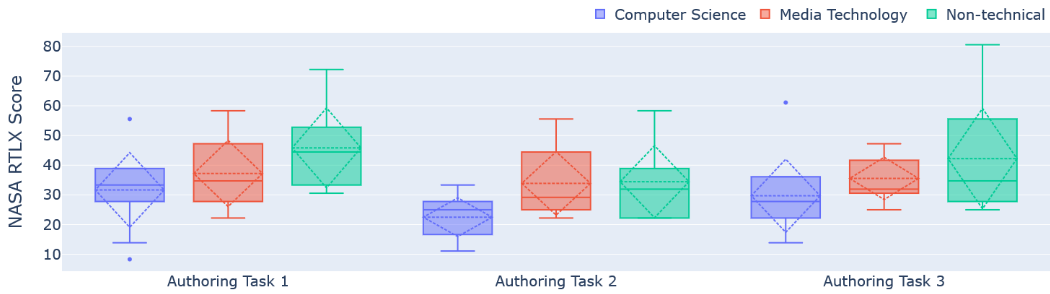


Figure 7.6: Average perceived cognitive load measures with the NASA rTLX [188] for each authoring task grouped by participants' study program, normalized to a 1–100 score.

Perceived Usability

Analyzing the perceived usability, reported through the SUS questionnaire after completing all three authoring tasks, using the SUS Analysis Toolkit [50], CS students reported a SUS study score of 85.75 (SD = 9.88). This would be considered "Best Imaginable" usability according to Bangor et al. [27] and is above the non-empirical, but commonly used, industry benchmark of SUS study scores of 80 [278]. MT students reported a SUS study score of 79.25 (SD = 9.36), which would be considered "Good" usability [27]. The non-technical students reported a SUS study score of 60.25 (SD = 15.14). This would be considered a below-average, "OK" usability according to Bangor et al. [27] but would still be marginally acceptable usability [28]. A sample size of $n = 10$ for each of the groups should be 80% conclusive, according to Tullis et al. [463]. The results of the SUS are visualized in Figure 7.7. The Appendix 32 includes the complete SUS score analysis from the SUS Analysis Toolkit [50] in the form of tables.

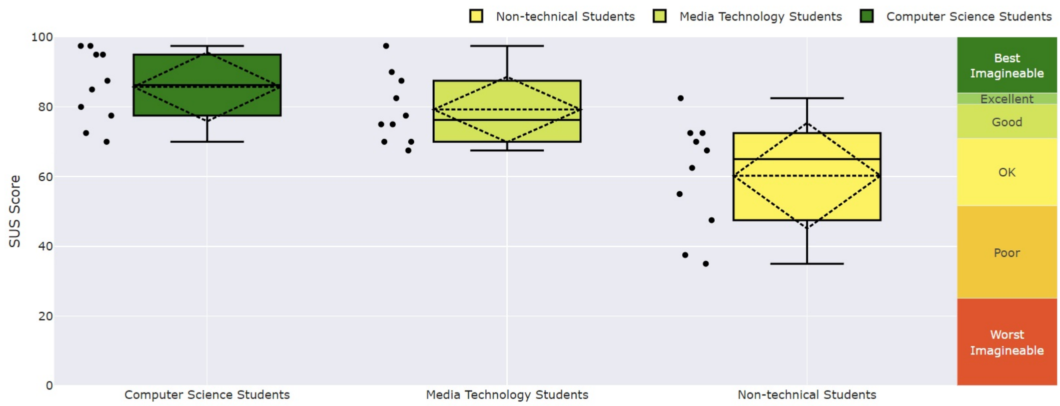


Figure 7.7: The perceived usability of the TrainAR authoring tool reported as SUS scores, grouped by the study program of the participants, plotted with the SUS Analysis Toolkit [50].

The assumption of normality was satisfied (Shapiro–Wilk test) and Levene's test considered the population's variance to be equal ($p = 0.293$). Therefore, a one-way ANOVA was conducted to compare the effect of the three groups on the perceived usability, reported through SUS scores. The one-way ANOVA revealed that there was a statistically significant difference in SUS scores between at least two groups ($F(2, 27) = 11.439$, $p = 0.00025$). Tukey's HSD test for multiple comparisons indicated that CS students reported significantly higher SUS scores compared to non-technical students ($p = 0.00025$, 95% C.I. = [11.7624, 39.24]), and MT students reported significantly higher SUS scores compared to non-technical students ($p = 0.00538$, 95% C.I. = [5.2624, 32.7376]). No statistically significant differences in SUS scores were found between CS students and MT students ($p = 0.479$).

Qualitative Feedback

To gather qualitative feedback for TrainAR, participants were asked what they liked and disliked while working with the authoring tool. When prompted to answer what the participants liked about TrainAR, eleven participants mentioned working with the visual scripting nodes and how these made it possible to implement complex training sequences without requiring any programming knowledge. Positively highlighted in particular was the visual representation of the processing logic. This was described as "intuitively understandable" and "clear". They stated that, once the basic concept was understood, it was "absolutely no problem applying them to the various tasks". Further, highlighted positively was the easy deployment process of a training to the handheld device. This made it possible to quickly identify and fix bugs.

When asked what the participants disliked about using TrainAR, three mentioned referencing the TrainAR Objects in the script nodes by their name. This was described as tedious and error-prone, because of typos, and participants wished that this could be performed by "drag and drop". Another three participants mentioned problems while moving objects in the 3D environment of the authoring tool (note: this was caused by a bug, where objects when dragged and dropped into the scene would sometimes not be placed at the correct height of the reference setup). Furthermore, the conversion process for TrainAR Objects was described as tedious by four of the participants, especially when multiple objects were involved, since the conversion had to be performed one-by-one and could not be performed in batches. Lastly, three participants from the MT and non-technical group mentioned that they may need help from a "technical person" to use TrainAR to its fullest potential.

When asked if they had any further remarks about TrainAR or the study, five participants from the MT and non-technical groups mentioned they had fun while using TrainAR. Three participants in the CS and MT group answered they thought TrainAR is versatile in use and could see its potential. Two CS students stated that it provides an interesting first insight into AR development.

Self-Assessment: Independently Creating a TrainAR Training

Finally, we asked participants how much they would agree with the statement "I think I would be able to create AR trainings on my own using TrainAR" on a seven-point Likert scale. Both the CS students (M = 6, SD = 1.15) and the MT students (M = 5.8, SD = 1.03) agreed with the statement, while the non-technical students somewhat agreed (M = 4.8, SD = 1.87) with it. As the sample size was small for Likert-scale data, and the normality assumption was violated, a Kruskal–Wallis H test was performed, which indicated that there was a non-significant difference in the dependent variable between the three groups, $\chi^2(2) = 2.64$, $p = 0.267$, with a mean rank score of 18.05 for CS students, 16.4 for MT students, 12.05 for the non-technical students. The effect size was small ($\eta^2 = 0.024$).

When asked to provide the reasoning behind their self-assessment, seven participants (all either MT or CS students) stated working with TrainAR, after a short familiarization phase, is easy to understand and use. Positively highlighted here were the visual state machine nodes, which were described as "intuitive" and "explained in an understandable way". Furthermore, three participants mentioned that the documentation was an enormous help in understanding what each element of TrainAR, especially the state machine nodes, does and made it possible to get them started quickly. Three participants, however, mentioned that they would not know how to acquire the 3D models necessary for creating trainings when they are not provided as they were during the study. One of the participants described the reasoning behind their self-assessment as "still too many questions" and another stated that they would need help from an expert to guide them if they had to create a training with TrainAR.

7.3 Domain Expert Requirements for the TrainAR Framework

As shown, in Figure 7.1, we believe that the utility of the TrainAR authoring tool is primarily the possibility to create useful TrainAR trainings. Because of this, from a scientific perspective, we do not state a research question regarding the utility of the AR authoring tool and the reported evaluations already address all stated research questions. Nonetheless, from a practical perspective, we also wanted to gather domain expert feedback on the utility of the TrainAR authoring tool and its created trainings directly, to ensure that our vision is shared with the domain experts, but also to see if they would have additional valuable input. Moreover, this helps to ensure that their requirements match with our expectations and current implementation of TrainAR.

7.3.1 Methodology

To accomplish this, we conducted semi-structured interviews in accordance with common interview guidelines [61, 495]. We conducted the interviews with 10 domain experts from three groups (Education and Social Affairs: 3 interview partners, Industry: 3 interview partners, Higher Education: 4 interview partners). All interview partners were domain experts, who could realistically be users of the TrainAR authoring tool based on their role. We gathered initial perceptions, requirements, wants and feedback based on several questions about how much demand there is for AR authoring tools in their context, how willing they would be to actually deploy an AR authoring tool, and what requirements for AR authoring tools they would have in general but also explicitly concerning TrainAR, after being shown explanatory videos about what TrainAR is and what can be accomplished with it. The interviews were designed in a way, so general feedback is gathered first, and afterward TrainAR is introduced to gather specific feedback regarding our vision and implementations.

The interviews were carried out, recorded, transcribed, and then inductively coded according to Mayring [314] using MAXQDA, by Jan Behrends [37]. They are reported in full, including the interview guide, explicit research questions, category definitions, and coding schemes in [37]. Here, the results are only reported in condensed textual form to highlight the most relevant findings. For readability, categories (the central schemes that emerged) are visualized in bold in the results.

7.3.2 Results

Generally, we found that interviewed domain experts all saw a need for AR as a technology, and that they had high hopes for AR to solve current and future educational challenges. More importantly, none of the interviewed domain experts rejected AR as a potential technology for their specific training purposes. But they did see hurdles. E.g., they indicated that there has to be actual added value, and it can't just be a gimmick, but also, that they fear that there could be potential hurdles with the usability of AR because of unfamiliarity of users.

General Domain Expert Requirements for Created AR Trainings

When coding the expert's answers concerning their requirements for the usage of AR trainings in their context in general, the following requirements emerged according to Behrends [37]:

Most often, the domain experts mentioned context-specific, or contextualized, additional information or the possibility to visualize processes which would not be visible normally as one of their requirements (**in-situ visualizations**). They mention that the **realism** of the displayed objects and their physical behavior is important. Ideally, they would want to utilize trainings that are based on **animations** and **interactive** AR applications. They also mentioned **accessibility** as one of their requirements. This was primarily mentioned by the "Education and Social Affairs" group, and here they mentioned that tools would need text-to-speech functionality or pictorial descriptions to be usable in their context. Besides the accessibility, the **usability** also emerged as a general requirement. Here, interviewed experts stated that they require the applications to "just instantly work", without it requiring extensive explanations, as this would often cause people to lose interest. Moreover, the experts stated that there has to be a proven **added value** for learning compared to conventional learning methods. Additionally, primarily mentioned by the industry group, the experts require AR trainings to be quickly **adaptable**, ideally without programming, to be able to accommodate constantly changing procedures.

General Domain Expert Requirements for AR Authoring Tools

Then Behrends [37] coded the domain expert's answers concerning the questions based on their general requirements for AR authoring tools to create such AR trainings, where the following schemes of requirements emerged:

Predominantly, the experts require an AR authoring tool generally **to not require programming** and, ideally, want the tool to be designed as easy to use as possible. They mention the requirement of having **supporting material** like documentation, tutorials, learning material or videos either directly in the application or as video series or workshops. Some experts explicitly suggest the usage of differentiating **expert/novice modes** to accommodate different levels of previous experience. Moreover, they state a good **usability, approachability, and learnability** as a requirement, where non-specialists can use the tool, and it is learnable in reasonable timeframes. Finally, they mention **collaboration** functionality as a requirement, where different parts can be designed by different domain experts, but also training and content can technically be stored in a database and shared with other authors to use.

Explicit Assessment of the TrainAR Authoring Tool

After the general requirements for AR trainings and a potential AR authoring tool, which could create such trainings, were gathered, we showed the interviewed domain experts and explanatory video of TrainAR and its functionality to gather explicit feedback on their perception regarding the TrainAR framework in the remainder of the interview.

Regarding their first impressions of TrainAR, most of the interviewed experts assessed the interaction concept as reasonable/useful. Some experts mentioned that they would wish for a more detailed interaction metaphor, e.g., incorporating more granular motor movements. Others explicitly state that they liked that it is "simple but realistic" with the requirement of gross motor movements to reach objects in the virtual space. They assessed the didactic elements of TrainAR as appropriate. Especially, the possibility to incorporate quiz elements for concept/declarative knowledge during the training was emphasized as a positive feature. Regarding the authoring tool, most experts found the overlay to be clear, but also mentioned that they would likely need some time to learn it. Other experts mentioned that they are already familiar with similar environments from other software, which uses comparable user interfaces. Especially for the visual statemachine, half of the interviewed experts reported already being familiar with similar concepts from the usage of other programs and almost all domain experts highlight the visual concept as positive, as it is "kept simple".

In terms of potential areas of application of TrainAR, they mention that they could see TrainAR trainings to be useful for flexible learning environments, for the familiarization of routine procedures in their context, and to relieve some teaching and training staff responsibilities for repeated learning of procedures. When asked about possible hurdles, some mention that they struggle to see the added valued compared to conventional methods, that it seems time-consuming to create such trainings, and that the requirements to acquire or create 3D models could be a potential hurdle in their context. Furthermore, some mention that the fully English tool and documentation would likely create a language hurdle during the authoring process.

7.4 An Initial Vertical Deployment & Evaluation of TrainAR

At this point, we evaluated the TrainAR interaction concept for its usability in several contexts, evaluated if TrainAR trainings can elicit desired utility (e.g., in form of learning or motivational benefits), evaluated the usability of the TrainAR authoring tool, and even gathered requirements from domain experts to assess if our vision and implementation of the TrainAR framework matches with their perspectives. As the results from the separate evaluations are promising, they do address our research questions as far as possible as separate entities. The next step is to combine the evaluation efforts and advance from the horizontal perspective, where components are evaluated separately, to vertically deploying and evaluating the deployment of the complete TrainAR framework holistically. Inspired by the “vertical slice” methodology [392], we hereby focus on covering all aspects vertically once, before expanding evaluation efforts horizontally again, increasing the sample sizes. Therefore, as a first step, we deploy TrainAR vertically, by letting one independent party develop one TrainAR training from scratch in their context, based on the documentation and our initial publication containing didactic considerations, but with no other guidance. We then instruct the independent party on how to evaluate the effects of the created AR trainings and ask them to reflect on the authoring process of that training.

For this, a media informatics student was instructed to create a TrainAR training in the context of the assembly and maintenance of radio relay stations as part of his master’s thesis in a company setting. The potential training task to be authored was hereby provided by the student, not by us. He evaluated the created training regarding its usability using the SUS Analysis Toolkit [50] and its utility by asking the participants for qualitative, verbal feedback. Furthermore, he was asked to reflect on the challenges during the creation process and to provide feedback based on his experience, using four guiding questions.

7.4.1 A TrainAR Training for Radio Relay Assembly

The TrainAR training that was developed by the student was a radio relay assembly for the offshore context. This assembly procedure, as visualized in Figure 7.8, is a complex procedure of assembling several of the relay’s components individually and then combining these components into the overall radio relay assembly, before attaching and aligning it. After developing the required models in Blender and utilizing them to create a procedural AR training with them in Unity, using the TrainAR Framework, the student evaluated the TrainAR training with nine workers of the company, who assemble and install the radio relays. Abstractly summarizing the findings [123], the workers reported “good”/“acceptable” [27, 28] usability (SUS study score: 77.5, SD = 18.3, Mdn = 87.5, n = 9), which was entirely in line with expectations from the previous evaluation efforts, and, as expected, provided non-representative qualitative feedback which indicated that they do see value in this kind of training.

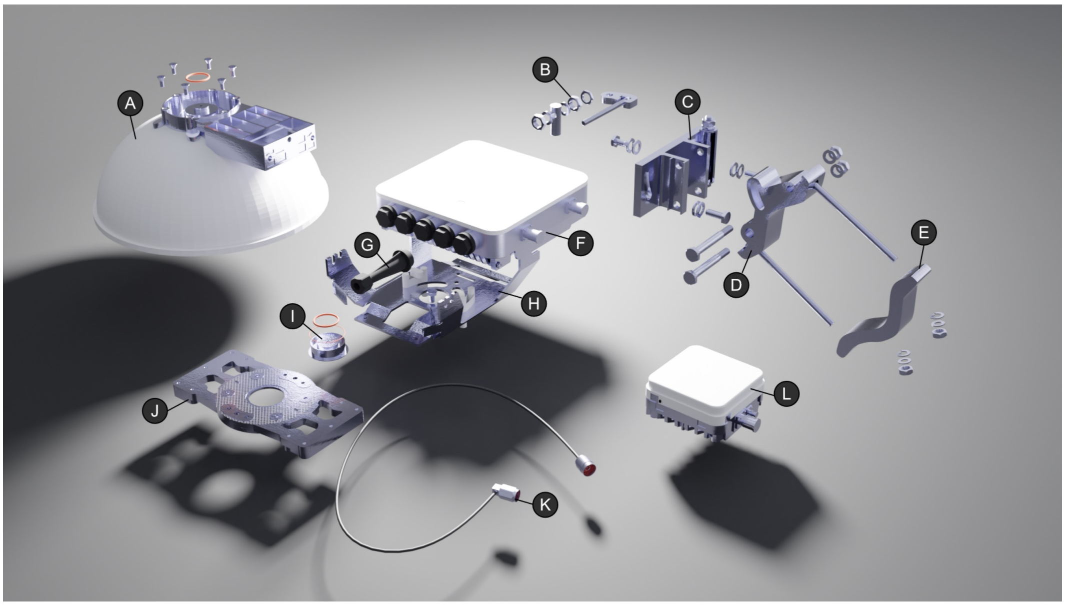


Figure 7.8: The training content of the radio relay assembly, including (a) an antenna, screws, and sealing ring, (b) an azimuth fine-adjuster, (c & d & h) brackets, (e) pipe clamps, (f) the main digital processing unit (MDU), (g) an ethernet cable insertion, (i) a transition element with sealing rings, (j) a mount for horizontal and vertical alignment of the antenna, (k) connection cable between MDU and ODU, and (l) the outdoor unit (ODU). The figure, depicted content, and descriptions are adopted from Jörg Eggeling [123], licensed under CC-BY 4.0

7.4.2 Author's Reflection on the Authoring Process

To reflect on the creation process, the author of this TrainAR training was first asked to reflect on: *"Where did the TrainAR authoring tool support you most? What did you particularly like about the authoring tool?"* Here, he provided feedback that indicated that he felt especially supported by the availability of a comprehensive documentation, that even goes as far as providing installation guides and helps with the "transition from the example project". He indicated that he liked the comparatively small scope of the framework, which he felt "focuses on the most important aspects" and stated that the framework was "quick and straightforward" to use.

Subsequently, when asked, *"What did you not like about the TrainAR authoring tool or where did you encounter problems?"*, his feedback indicated he felt the "script editor" (meaning the visual statemachine) was becoming increasingly convoluted with the growing complexity of the training. He also stated that he felt the linking of the visual stateflow and the TrainAR Objects through their names was error-prone, and that if errors were made in these instances, they were challenging to detect, as the deployed training would simply not proceed in these cases and there would be no indication that this was because of typing errors. He furthermore reported problems with building trainings for iOS devices through XCode on the silicon processor versions of Unity for macOS, but also reflected that this was likely a problem of Unity rather than the framework.

Then he was asked to reflect on the question: *"As how challenging did you perceive the following aspects of the authoring process of your TrainAR training? (1) Transferring the training procedure towards a conceptual flow of actions (2) Incorporating didactic considerations for the training (3) Ensuring the usability of the training (4) Creating or acquiring 3D Models for the training (5) Converting acquired or created 3D models into TrainAR Objects (6) Transferring the conceptual flow of actions into the TrainAR Stateflow."*. His feedback indicated that he found the creation (4) and the conversion (5) of the 3D models to be the most challenging and especially most time-consuming aspect. This was followed by the conceptual transfer (1) of the assembly procedure towards a flow that can be implemented in TrainAR, as the assembly of the radio relay was not well documented the conventional way before. He stated that it became apparent that it was not always clear which parts were meant to be strictly sequential. The transfer of the conceptual flow of states into the TrainAR statemachine on the other hand (6), he described as "easier to realize" "after looking into the project with the coffee pot" (the example training delivered with the TrainAR framework) and consulting the documentation. For the incorporation of didactic considerations (2) and ensuring the training's usability (3) he stated that he made no considerations, and relied on the frameworks delivered functionality, but he did incorporate quiz elements to "maintain the attention of the [trainees]".

Finally, he was asked *"Which aspects of the Authoring Tool of TrainAR did not meet your expectations or requirements, and how could those aspects be improved? Do you think these are context-specific or general requirements?"*. Here, he provided feedback that indicated that all aspects were meeting his expectations, but that he would wish for additional buttons to simplify the navigation in the authoring interface.

7.5 Discussion

7.5.1 Evaluations of TrainAR Trainings

As the current and forthcoming results of the utility and usability of TrainAR trainings look promising, this is at least a first indication that the TrainAR framework is providing the necessary utility to create TrainAR trainings that elicit the desired learning outcomes, are usable, and enjoyable. Though this is promising, and distinctive application contexts were chosen deliberately, the generalizability of these results has to be confirmed through the application of the framework to novel contexts. Notably, while the feedback of other researchers, utilizing TrainAR for more than three years at the time of publishing this thesis, indicates that the provided utility is largely sufficient, some implementations of TrainAR already required context-specific extensions, such as logbooks and interactive chemical graphs in the MARLab training application [114]. Moreover, while we were able to achieve consistently high perceived usability, this can only provide the upper limit and show that it is possible to create usable trainings but does not guarantee that usage of the TrainAR authoring tool will inherently lead to trainings with good usability.

The fact that the TrainAR authoring tool can be used effectively to create procedural and interactive handheld AR trainings is shown through our use and the use of partner universities that utilized TrainAR to create interactive procedural trainings and learning games for smartphones and tablets. These trainings span across a wide range of topics, targeted media competencies, and ages, with consistent observations, especially in terms of the enjoyability of the trainings, which previous work identified as the most important factor for usage intention [507].

7.5.2 Evaluation of the TrainAR Authoring Tool

The systematic usability evaluation of the TrainAR authoring tool revealed several interesting insights. While there were trends in the TCTs across all sub-tasks but also on the task level, no statistically significant differences were found. When categorizing the sub-task by the task category, a clear trend can be observed, where sub-tasks that occur multiple times are subsequently completed faster. This decline in subsequent TCT is especially steep after the first occurrence of a sub-task category, and is seemingly quite consistent across the groups (see Figure 7.5). This might be an indication that people can learn the handling of the authoring tool quickly, and there is only an initial hurdle to conceptual understanding. Interesting here are also the two outliers of the first occurrence of the action node and the first occurrence of the fork-action node, which require by far the most time to complete when they are introduced, but this decreases sharply. While we expected as much and already provided getting-started guides, we interpret this as a call for even more in-depth onboarding materials that provide the author with assistance on this initial hurdle, e.g., in more digestible formats such as video tutorials or practical course materials.

We tried to increase the difficulties of each of the three authoring tasks by first using a linear task of actions in the first, introducing quizzes and custom actions into the second, and then introducing non-linear flows in the third task. The TCT by task indicates that the complexity increment of quizzes was not nearly as challenging as the introduction of non-linear flows, as there is a decline in TCT between the first and second authoring task, but then, while the median stays consistent, a noticeable increase in the average and standard deviation of the TCT for the third authoring task, as can be seen in Figure 7.4. This was caused by fork actions. Looking at these and contextualizing them with the TCR, this was clearly caused by the introduction of non-linear flows across the groups.

Generally, significant differences were found for the TCR between the groups. Here, CS students performed significantly better than non-technical students, and the difference between MT and non-technical students was approaching significance ($p = 0.065$). An increase in TCR from the first to the second authoring task is consistent with the TCT and across the groups. Notably, non-technical students had very noticeable problems as soon as the non-linear fork-actions were introduced, which, in contrast to TCT, was not as apparent in the other groups. This is likely due to familiarity with programming and node-based systems in asset-creation pipelines that the MT and CS groups brought with them, and should be considered for the documentation and onboarding material.

This trend continued for the perceived cognitive load, where CS students reported significantly lower scores than the non-technical students, but no other differences were found. Again, in line with TCT and TCR, within the group, the perceived cognitive load first decreased after the second task but then for all groups increased with the introduction of the fork-actions again.

Most importantly, the perceived usability reported through the SUS was the self-reported measure of usability in our experiment. Here, significant differences were found between CS and non-technical and MT and non-technical students, but not between the CS and MT groups. As this is in line with the objective measures of TCT and TCR and somewhat in line with the perceived cognitive load, we believe this supports the hypothesis that the tool is usable by media technologists and domain experts with high media competency in its current state. The results indicate that domain experts with lower levels of media competency might struggle with the current state of the authoring tool, with the most difficult challenges being asset acquisition and non-linear state-flows. Nonetheless, it might be possible to lower the entrance barrier through documentation and getting-started guides, e.g., in more approachable formats such as videos or practical course material. Additionally, when asked for their self-assessment if they think they were able to create a TrainAR training, even the non-technical students somewhat agreed, which did not significantly differ from the other groups, which agreed.

While we think these results are promising, they are limited in several ways. For one, recruiting students from three groups with implied differences is not the ideal experimental setup, as stated in Section 7.2.2, but was chosen for practical reasons. Notably, at least the self-reported measures were in line with the group expectations. Additionally, the sample size was small, and we were aware that the prospective statistical power would be low, but larger sample sizes were not realistic, as the study took roughly two to three hours to conduct per participant and one-on-one study support was needed. Nonetheless, our results, and the trends that can be shown descriptively and through the qualitative feedback, are in line with our expectations. Importantly, they are also in line with the observations we made when sharing TrainAR with other researchers over the last two years, and using the authoring tool iteratively in several practical tutorial sessions at our university. We are therefore satisfied with these results and will address the observations in subsequent documentation and course material and move on to the next evaluation stage.

7.5.3 Impressions & Requirements from Domain Experts

While, from the scientific perspective, the utility of the authoring tool is the possibility to create useful AR trainings, which prove to elicit learning benefits, and therefore our research questions are answered with the evaluations discussed above, in actuality, the deployment by the domain experts might be dependent on further factors. The primary goal of the requirement analysis in the form of semi-structured interviews was to avoid overlooking these potential requirements. Occasionally these, from the developer's perspective peripheral, requirements can be the deciding factor for the adoption.

In our case, it appears that the requirements by the domain experts for the AR trainings match reasonably well with our vision and current implementation. In the "general requirements" parts, before introducing TrainAR specifically, domain experts wish for expressions like in-situ visualizations, specific accessibility features, and the want or need for 3D animations. But they also wish for a usable authoring tool, which can be used by non-programmers and be learned in a reasonable timeframe. These are contradictory requirements, as discussed in detail in Chapter 3. Likely, their answers here were influenced by the deliberate choice of the open structure of the interviews, where requirements for the AR trainings were asked abstractly, before the requirements for the authoring tool were asked, to not influence their answers regarding requirements. This provokes articulating a “wishlist” for both components without considering the inherent compromises that have to be made. The explicit feedback on the TrainAR authoring tool after introducing it through an explanatory video does furthermore indicate that they are aware of the potential need for a tradeoff between the fidelity of the created AR trainings and the complexity of the authoring tool. Most importantly for us, they repeatedly mentioned, in the general requirements but also specifically as feedback regarding TrainAR, that trainings cannot be a “gimmick” and some struggled to see the benefits of the TrainAR trainings as presented in the interview. The rest of the requirements, like good usability and learnability, realism, and adaptability are all in line with our vision and the current form of the TrainAR interaction concept and didactic considerations. Especially as they liked all the didactic elements and most understood and liked the “simple but realistic” interaction metaphors.

The general requirements for an AR authoring tool but also the TrainAR authoring tool-specific feedback from the interviews indicate that we identified all key aspects of the authoring process: In line with learnings from the literature, they mention that they would not want to program, they wish for supporting material to learn the authoring tool, they wish for collaboration functionality, and they wish for expert/novice modes of the authoring tool. All these concepts are central ideas and are satisfied by the current implementation of the TrainAR authoring tool.

Altogether, in line with the learnings from the usability study of the authoring tool, we believe that to address these remaining concerns, it would likely suffice, if we make the scientific results on the utility of TrainAR trainings more prominent. This could be achieved by, e.g., incorporating abstract findings from the evaluations into course materials, videos, or the online documentation. Alternatively, having more established, evaluated training and contexts might also help convey the benefits more easily in practice, as our previous non-representative observations indicate that letting lecturers actually deploy the training in their context promptly convinces them of the benefits.

7.5.4 Insights From the First Vertical Evaluation Efforts

Overall, as expected, the student was able to deploy the TrainAR framework independently, with some occasional questions and support, and the training he authored in the context of radio relay assembly achieved good perceived usability results, comparable to TrainAR trainings developed by us or partners. As he did not program any C# functionality beyond the scope of TrainAR, this adds towards research question 3 with a first, non-representative vertical evaluation after actually deploying the framework “in the wild” and letting independent parties use it. This sets promising expectations for future evaluations.

Additionally, in his reflection, some aspects which were already mentioned in other evaluations were repeated, e.g., he perceived the linking of objects in the stateflow by name as “error-prone”. Not visible in the systematic usability evaluations of the TrainAR authoring tool, he also provided feedback that, with the increasing complexity of the training, the statemachine would become convoluted. Though this is likely an inherent trade-off of using visual descriptions of complex logic, this could potentially be addressed with further navigational or overview components in the authoring interface. Furthermore, as is to be expected, he found the creation of the assets, in this case primarily the 3D models in Blender, to be one of the most challenging aspects of the creation of the training. This aspect was likely specific to his context, as he tried to implement a very specific training, where models for the assembly were not widely available (e.g., in asset stores). This aspect also created challenges in the conversion as it, e.g., required for the imported meshes to be set up correctly for Unity and pivot points to be set correctly.

7.5.5 In-The-Wild Testing Approach & Ongoing Evaluations

Realistically, there is no further lab-study or prototypical evaluation that can fully answer our questions. Gathering requirements before having an initial implementation of TrainAR trainings and an authoring tool to create them still faces the causality dilemma of the missing shared understanding of what procedural task trainings would even look like for Handheld AR, as discussed in Section 6.5. Additionally, it is challenging to distinguish general needs for the framework from context-specific needs. We are cautiously optimistic about the provided set of utility, based on our non-representative observations in our contexts, the feedback by other researchers, and the systematic usability evaluation of the authoring tool. Furthermore, our first vertical exploration and the insights gained from the domain expert interviews are largely in line with our separate evaluations. Still, generalizable insights will only be gathered after open-source publication of the authoring tool and its application by even more independent parties.

With the publication of the entire TrainAR framework, the evaluation results, and open sourcing of the authoring tool, we are therefore starting an evaluation period in line with the field usability testing methodology [118], and other researchers and educators are encouraged to deploy the tool and provide feedback (e.g., through email or GitHub Issues) from their real usage experience for additional refinements or future directions of TrainAR.

Additionally, scaling a combination of the requirement analysis and the conducted vertical slice evaluation up, we will use the TrainAR framework and apply it holistically in our next evaluation iteration. Here, we plan to let teachers author TrainAR trainings with the tool and not only evaluate the authoring process itself but also let them use the created trainings in a classroom setting as a multimedia learning intervention, where we will then gather feedback on the utility and usability of the trainings created by them, not us. In doing this, we explicitly expand upon our perspective of the usefulness evaluation over the two levels (see Figure 7.1), where we would streamline the usability evaluations of both the authoring and usage of the TrainAR training through usage of our open-source SUS Analysis Toolkit [50] (see Section 4.5) and qualitative questions, while discussing utility benchmarks of the TrainAR trainings with the teachers individually.

Furthermore, we will continue to deploy the authoring tool into the practical parts of an apprenticeship course and university lectures again. This allows us to iteratively evaluate TrainAR further and improve the usability of the authoring tool, while providing first insights into AR development for the students. While this was not our initial intention, in our experience, the TrainAR authoring tool can serve as a good starting point for AR development, which, and this was one of our intentions, allows for a seamless transition into actual Unity development. It is possible that we will investigate this preliminary observation in a systematic evaluation effort in future research endeavors.

7.6 Summary

TrainAR trainings showed good usability, were enjoyable for the trainees, and showed utility results such as increased motivation, increased perceived competency and autonomous usage in classroom settings. The systematic usability evaluation of the TrainAR authoring tool revealed that, while all groups were able to use the authoring tool and improved over time during the study, the likely target group of the authoring tool is media technologists or domain experts with high media competency. This is mainly due to concepts such as the formalization of non-linear flows of states, which are hard to grasp for people with lower levels of media competency, and the need to acquire appropriate 3D assets. Nonetheless, programming or expertise in software engineering does not seem to be required to utilize the authoring tool, and we believe that our current results suggest that the TrainAR authoring tool is a useful contribution to the current state of the AR authoring landscape. Finally, the first overarching TrainAR evaluations of the expert requirement assessments and vertical deployment supported the insights gained in the more granular evaluations across contexts and will be combined and scaled up in our future evaluation efforts.

8 Conclusion

"A complex system that works is invariably found to have evolved from a simple system that works." — John Gaule

One of the most enduring challenges remaining for the widespread adoption of AR is the creation of AR content at scale. Potential solutions to this issue, among others, come from the realm of AR authoring tools, which enable non-programmers to create AR content. But this field is comparatively diverse in its nature and not well understood.

As the overall holistic inquiry, this thesis contributed towards filling this gap by systematically establishing the design space of AR authoring tools and then exploring it from the practical perspective afterward. To accomplish this, it created a theoretical design space, then first developed procedural AR trainings in a real context the conventional way, and then finally explored the theoretical design space of AR authoring tools with the practical learnings from that context. Based on the exploration, an AR authoring tool was developed to enable non-programmers to create similar trainings at scale.

The thesis therefore contributes a very first proposal of the design space of AR authoring tools, with a guiding framework on how to explore it. While it is neither meant to be definitive, nor comprehensive, it should help others to more quickly understand and subsequently explore potential design decisions in the design of AR authoring tools. It is also a first step toward creating design guidelines for AR authoring tools, which are based on systematic analysis of previous efforts. Moreover, the thesis contributes an exemplary exploration of that design space in the context of academic procedural trainings, which can serve as an example of how the design space is envisioned to be explored.

Within this inquiry, the thesis also contributes a systematic scoping review of 20 years of research efforts in AR authoring tools, which reviews trends and gaps of the field between 2000 and 2020 based on 293 included publications. This scoping review was also used to contribute the mapping study to visualize trends of the field and structure efforts based on 26 dimensions. The results of this mapping study, beside being used for the construction of the design space itself, are furthermore contributed as a CC-BY licensed multi-variate dataset that other researchers can use to identify fitting previous efforts for their work on AR authoring tools or identify gaps.

8.1 Additional Contributions of this Thesis

Through contributing this first structured understanding of AR authoring tools in the form of a design space and then also actively exploring this established design space, several additional contributions beyond the scope of this guiding inquiry are reported in this thesis. These contributions, in all likelihood, hold intrinsic value, providing meaningful insights and are applicable independent of the overall inquiry. Reflecting on the contributions of this thesis, perhaps the contributions that were made in the process of addressing the main inquiry may hold even greater significance and provide more value to others than the guiding inquiry itself.

Foremost, TrainAR is contributed as a holistic open-source authoring framework based on the identified research gaps in the scoping review and the exploration of the established design space of AR authoring tools. In this, it combines the theoretical and practical learnings of the design space and the conventional development during project Heb@AR. TrainAR is fully open-source, was comprehensively evaluated for its utility and usability from both the authoring and usage perspective, includes a comprehensive technical documentation, is based on novel interaction concepts that can be realistically scaled today, and even provides first guidance for didactic considerations that authors using the framework would likely encounter.

Additionally, the SUS Analysis Toolkit is contributed as an open-source toolkit as a side contribution. While primarily developed as a benchmarking utility for the comprehensive perceived usability evaluation efforts encountered during Project Heb@AR and the evaluation efforts of TrainAR beyond the initial scope, the toolkit is technology-agnostic and hopefully provides value to HCI researchers and practitioners in general.

Finally, the Heb@AR App is contributed as an Open Educational Resource for the midwifery context that is available in the Android and iOS app stores, including supplementary material which is available as CC-BY 4.0 licensed open-source content. Beyond the app itself, the process of developing it is openly contributed for other researchers to learn from, a vision of ARBTs and the aspects of scalability are contributed as discussions, and comprehensive evaluation efforts are described. These evaluation efforts, partially through selective-variable analysis of evaluations during project Heb@AR, and partially through supplementary studies grounded in self-determination theory, contribute first exploratory evidence towards the learning benefits, which can be achieved through scalable AR concepts and hardware choices.

8.2 Limitations & Future Work

The limitations and potential for future work for each of the three major parts of the overall inquiry were already comprehensively discussed in their respective chapters and are not restated here. The limitations of the systematic scoping review were assessed in Section 2.4, the limitations of the subsequent mapping study in Section 2.8. Section 2.9 of that chapter furthermore discussed potential future work for scoping efforts in the field of AR authoring tools itself, but also beyond the authoring perspective. The limitations of the exploratory evaluations of the Heb@AR App during the Heb@AR project were discussed in Section 4.7.5 and potential subsequent technical and empirical inquiries for the Heb@AR App were discussed in Section 4.7.6. Potential technical future work for the TrainAR Framework was discussed in Section 6.7.1 and the limitations of the evaluation efforts of the framework were discussed in Section 7.5. Section 7.5.5 discusses “in-the-wild” evaluation efforts for the TrainAR Framework as the next steps towards its holistic evaluation.

What remains to be discussed are the limitations and the future work for the design space of AR authoring tools, and therefore the overall inquiry of this thesis. Overall, while the design space in its current form is built upon a systematically reviewed and mapped body of literature, it is exploratory in nature and therefore limited in several ways. Firstly, the literature map this design space is based on concludes at the end of 2020, and should be updated with newer publications. Secondly, it should be investigated whether commercially available tools should be incorporated into the design space. Common expressions of some design dimensions may have been missed because they only exist in commercially available solutions. Moreover, while the strictly systematic scoping of the literature was the enabling factor that allowed to construct the design space and perform association analysis on the mapped dimensions, it is possibly not the ideal choice to refine it to ensure the completeness of the expressions in the dimensions, but also the dimensions of relevance themselves. The design space should be seen as a first proposal and should be refined in future work through not only more systematic literature work but also narrative perspectives by other researchers, expert interviews, and the incorporation of practical perspectives. But all this is likely only possible after the first proposal of the design space is explored (as in utilizing it for design space exploration) by independent parties to identify its strength and remaining exploration challenges when trying to use it to inform the development of their AR authoring tools.

Besides the refinement of the design space as a framework for the exploration of potential design decisions during the development of future AR authoring tools, it can also be used to inform researchers what aspects need empirical evaluation. As discussed before, just because there is little asymmetrical association between the 13 design decisions of which the design space of AR authoring tool consists, does not mean that there should be no associated decisions. Moreover, just because it was not possibly to identify distinct "types" of tools, that does not mean that such types cannot emerge with a better understanding or progress of the field. Ultimately, this design space is the first effort of its kind and size, providing an overview for the first time and revealing all potentially relevant design decisions. Researchers should empirically evaluate the benefits but also the challenges of specific design combinations. In line with the efforts, researchers should also use the carefully reviewed, mapped and analyzed design space, which at this point is more of a systematic catalog of potential design decisions with a reference map to previous efforts, to work towards establishing design guidelines for AR authoring tools through adding narrative perspectives.

Finally, the design space incorporates the 5w-1H inspired guiding questions, which developers can use to reflect on their context and human factor considerations before exploring the design space itself. While this is already moving towards the direction of providing somewhat of a framework, and it is briefly reflected on which of the human factors and context reflections might influence which of the decisions in the design space (see Figure 3.3), those potential relationships should also be investigated in detail. Likely, through the usage of the design space, first tendencies will become apparent whether these reflections are correct and whether the guiding reflections are actually helpful to other researchers. At least in our context they were helpful and worked, but our context might have biased the method. On the other hand, even if this were to be the case, at least they will help researchers to report human and context factors in their publications more clearly, which was a gap in the literature.

Final Remarks

Reflecting on the overall inquiry of understanding the design space of AR authoring tools alone, this thesis tends to generate more questions for future work to answer than it resolves existing questions. It constructs the design space of AR authoring tools, revealing what design decisions to explore, provides some guiding framework to address the decisions, and it even provides a very first exploration. But it ultimately does not inform about what decisions to make within the dimensions. In many respects, this thesis concludes just prior to the most captivating subsequent segment of the broader inquiry: the point where the design space would be refined, expressions of dimensions would be subjected to empirical comparative evaluations, and design guidelines would be formulated. But this thesis is already more than long enough. In this, while this sentence signifies the conclusion of this thesis, it is my sincere hope that, conversely, it serves as a starting point for others to further explore and expand upon the field of AR authoring tools.

List of Figures

1.1 Exemplary Mixed Reality technologies like projection-based, handheld, or head-mounted AR, cave-based VR systems, 360-degree videos, and immersive VR, contextualized on the Virtuality Continuum by Milgram [321]. Moving toward the right on the continuum representation, incorporation of real-world elements decreases and incorporation of computer-generated content increases. Figure by Carolin Hainke & Jonas Blattgerste, licensed under CC-BY 4.0 2
1.2 The conventional development process (top) and the development process through AR authoring tools (bottom) to distribute AR content, here described as AR constructs, to the intended user of the AR content. 3
1.3 A schematic structure of this thesis, colored by component: Practical learnings & domain knowledge of Heb@AR (Chapter 4) are used as variables to explore (Chapter 5) the design space of AR authoring tools (Chapter 2 & 3), before the AR authoring tool TrainAR (Chapter 6 & 7) is developed and evaluated based on this exploration, combining theory and practice. 5

2.1 The "full-fledged hybrid search strategy" [506] with "parallel snowballing" [327] used for the systematic review. An initial database search with subsequent exclusion phase based on pre-determined criteria is followed up by *n* iterations of combined forward- and backward snowballing, until no new publications to include are found. This figure is recreated based on the combination of ideas in concept figures described by Wohlin et al. [504, 506]. 14
2.2 The 624 publications and their source database (left). The inclusion/exclusion funnel of publications from the retrieved database to the final 128 publications that were included in the initial database (right). 17
2.3 The snowballing process of the hybrid search strategy with the 128 publications of the initial database search (DB search) as a starting point, which identified 165 new publications after screening/reviewing 15068 publications (including duplicate references/citations) after 5 iterations of Backward- (BS) and Forward snowballing (FS). 20
2.4 The 293 publications included in this review by publication year, split into the search procedure, they were retrieved with. A table with all publications, grouped by the search strategy they were found with, can be found in Appendix 3. 21
2.5 The PRISMA 2020 flow diagram [361] with the combined results from the initial database search (with 128 publications, which are fully PRISMA conform) and the subsequent snowballing (165 publications), which cannot be properly reported through PRISMA [361]. 22

2.6 The chronological order of the 5 systematic reviews previously published on or closest to the topic of Augmented Reality authoring tools, that this review's results are compared against. 24
2.7 The 293 publications contributing an AR authoring tools, grouped by publication year. 29
2.8 The 293 publications by number of authors who contributed to the publication. 30
2.9 The 293 publications, split into their publication format as a donut chart (left) and used as a breakdown dimension when plotting the number of publications for each year (right) 31
2.10 The 293 publications mapped onto the 5 publishers with the most publications. Publications that were published through other publishers were grouped as "Others". 32
2.11 The 293 publications, split into whether the AR authoring tool is the main or a side contribution of the publication (left) and the trend of this distinction over the years (right). 32
2.12 The sum of citations at retrieval in March 2021 and 2 years later in March 2023, grouped by publication year of the literature. 33
2.13 The average number of references in a publication, grouped by publication format (left) and the publication year (right) 34
2.14 The 293 publications split into the deployment purpose the AR constructs are authored for (top right), and the publisher they are published with (top left). The bottom graph shows the trend of deployment purposes over the years. . . . 36
2.15 The 293 publications, grouped by deployment contexts the AR constructs created with the AR authoring tool are intended to be used in, broken down into the deployment purposes in that context. 37
2.16 The 293 publications split into whether they evaluate no aspect, the authoring perspective, the usage of authored AR construct perspective or both (top left). The graph on the top right visualizes this trend over the publication years. The bottom left graph shows the evaluation efforts grouped by the deployment purpose, and the bottom right graph shows the association between the average citation growth over two years and the evaluation efforts. 38
2.17 The 293 publications, grouped by the publication year and broken down into whether they are actually available as binaries or open source (top). The bottom left graph shows the publications grouped by the deployment purpose, broken down by the availability. The bottom middle plot shows the publications, grouped by availability, broken down by which aspect of the AR authoring tool was evaluated. The bottom right plot shows the publications, grouped by availability, with their average citation grows over the two year timespan. 40
2.18 The 293 publications, grouped by the envisioned author, who uses the AR authoring tool to create an AR construct, broken down by the deployment purpose of the construct. 43

2.19 The number of proposed AR authoring tools in the 293 publications which support this hardware platform for the authoring of constructs, including overlaps (top) and the one-to-one mapped choice of hardware, including "Multiple Device Types" (bottom). 44
2.20 The 293 publications grouped bey the deployment purpose (left) and the publication year (right), broken down into the markup notation used to serialize the AR construct. 45
2.21 The 293 publications grouped by whether they are a standalone application intended to be used in a host software, or both (left) and the average citation count in March 2023 based on the implementation decision (right). 45
2.22 The 293 publications split into whether they provide a 3D (not necessarily in-situ AR) scene preview or not (left) and grouped by the deployment purpose of the authored AR construct, broken down by this differentiation. 46
2.23 The 293 publications, grouped by the contextualization of the authoring process (left) and additionally as a stacked Barchart, broken down by the deployment purpose (right). 47
2.24 The 293 split into external and internal AR authoring tools (top left) and broken down into the deployment purpose of the created AR construct (top right). The bottom graph shows the publications grouped by year, broken down into the relationship between authoring and AR construct usage tool. 49
2.25 The 293 publication split into whether they distribute their AR constructs locally or through an external server (top left), this differentiation used as the breakdown dimension, when grouping the publications by the authoring tools' hardware choices (top right), deployment purpose (bottom left), and publication year (bottom right). 50
2.26 The 293 publications grouped by who is the envisioned user of the authored AR construct, broken down by the deployment purpose. 51
2.27 The number of supported hardware types in the 293 publications, which can be used to utilize the authored AR constructs, including overlaps (top) and the one-to-one mapped choice of the usage hardware, including "Multiple Device Types" as one expression (bottom). 52
2.28 The 293 publications, grouped by the interaction technique used for the tool that uses the authored AR construct, broken down by the deployment purpose of the AR construct. 53
2.29 The amount of AR authoring tools reported in the literature that support the 9 identified content types. This graph includes overlap between content types, as AR authoring tools usually support multiple content types. 53
2.30 The 293 publications, split into whether AR constructs encompass a user-controlled sequence of states (sequential) or a single constant state/visualization (left) and the publications grouped by the deployment purpose broken down by this differentiation (right). 54

2.31 The 293 reviewed publications grouped by which tracking techniques were utilized for the usage of the authored AR construct. 55
2.32 An alluvial diagram representing the relationships between three bibliographic dimensions of the publisher of each paper, the paper's format, and whether the AR authoring tools were the main or a side contribution of the publication. . . 56
2.33 An alluvial diagram representing the relationships between four dimensions of the literature map: Construct Author, Authoring Hardware, Usage Hardware, and Construct User. The diagram highlights how the dimensions are interconnected in the mapped 293 reported AR authoring tools, with the width of the flow lines indicating the extent of each relationship. 58
2.34 An alluvial diagram representing the relationships between four technical dimensions of the map: Authoring Hardware, Markup Notation, Content Distribution, and Usage Hardware. 59
2.35 The number of publications retrieved from the timeframe of 2000 to 2020 in blue, a linear trendline, and an estimate of publications expected to be found for the upcoming years in red. 62
3.1 Asymmetrical associations based on Theil's U [459] ($U(X, Y) \neq U(Y, X)$) between all 13 software and hardware design decisions mapped from the 293 AR authoring tools reported in the literature between 2000 and 2020. 71
3.2 The 5W-1H inspired five W-questions, which creators of AR authoring tools should reflect on before exploring the Design Space (H) of AR authoring tools. 75
3.3 The design space of AR authoring tools with its 5W-1H-inspired guidance to explore it. 76
4.1 A student disposes of a glass ampule after completing the virtual preparation of an emergency tocolysis in the Heb@AR App during curricular implementation at Bielefeld University of Applied Sciences and Arts [381]. Picture by Patrick Pollmeier, licensed under CC-BY 4.0 (cc)(i) 77
4.2 AR as the logical extension of Computer-based Trainings & Web-based Trainings. 79
4.3 The initial "bootstrapping" development methodology established during Project Heb@AR for the structured development of medical AR trainings in interdisciplinary teams. 84
4.4 Examples of the prototyping process of models and their interactions from paper prototyping, through low-poly models with simple interactions, to modeling the final versions. 87
4.5 The TrainAR interaction concept: Grabbing & moving, interacting with and combining virtual AR objects to complete a procedural chain of actions. 88

4.6 The Decide-Freeze-Imitate interaction concept: Deciding which action to perform next, getting detailed feedback on the decision, viewing the required motor component, freezing the AR view, putting the Smartphone aside, and performing the observed motor action. 90
4.7 An outline of the development and evaluation timeline of all AR trainigns developed during Project Heb@AR. Visualized from process description, over learning goal formulation, to technical development and formative evaluation as GANTT bars. Milestones visualize the evaluations. Black bars show the time between starting development to it's first curricular use. 92
4.8 Central design aspects of the app: A medical-inspired color pallete, flat UI design, consistent button and information placement, playful guidance and context, and realistic AR objects. 95
4.9 The medical inspired color palette used for the Heb@AR App: Two main colors, pastel blue (#6895AC) and white/light gray (#EEEEEE), combined with six supplementary colors (#776DC3, #6DC299, #64AB35, #5B1261, #1B4165, #ECB736). 95
4.10 The main menu (a) and the training selection menu (b) of the Heb@AR App. Moreover, menus for the supplementary materials for the trainings (c), information about project Heb@AR (d) and the settings menu (e). 96
4.11 Training 1, "the preparation of an emergency tocolysis" in the "guidance mode". Shown in the figure is the step of drawing up the carrier solution into the syringe. The top UI element guides the user on what to do next, while the expert midwife provides tips on what is important to take into consideration when pulling up the carrier solution in reality. 98
4.12 Training 2, "the preparation of a birthing person for a cesarean section". Shown in the figure is the step of positioning the legs of the birthing person in preparation for the placement of the permanent catheter. 99
4.13 The individual training version Training 3, the "reanimation of a newborn". Shown in the figure is the task of using the stethoscope to determine the newborn's heart rate. The Smartphone's AR view is frozen, while the action is performed, and the smartphone simulates the heart beat through audio and vibration. 101
4.14 To use the multi-user version, up to 6 smartphones can connect to each other using QR-Codes. (left) During the training, all participants make decisions on which action they would perform next. After everyone submitted their decision, they are shown on the UI. (right) 102
4.15 The multi-user version Training 3, the "reanimation of a newborn". Shown in the figure is the cooperative task, where one trainee is instructed to perform chest compression, while the other is instructed to perform bag-valve-mask ventilation. 103

4.16 Training 4, the virtual version of the "reanimation of a newborn using a reanimation unit". Shown in the figure is the step of using a stethoscope to determine the breathing rate of the newborn. While the trainee is not required to actually use the stethoscope, they have to listen to and count the breathing rate, which will be queried in the next step by the app. 104
4.17 Training 5, the "denomination of the female pelvis". Shown in the figure is a grabbed Latin puzzle piece to be connected with the correct bone structure of the virtual pelvis model. 106
4.18 Students train the preparation of a tocolytic injection, using the Heb@AR App, during the curricular implementation of the Heb@AR App at Bielefeld University of Applied Sciences and Arts [381]. Picture by Patrick Pollmeier, licensed under CC-BY 4.0 🅭🅯 108
4.19 The four basic components of the System Usability Scale Analysis Toolkit User Interface: The Analysis Tabs, Interactive Plots, Data Tables, and Customization Options. In this case, the SUS Score study comparison analysis tab is selected and shows the interactive notched box plot and corresponding data table according to the chosen customization options. 110
4.20 Conceptual visualization of the accepted user experience principle, that the usefulness of the Heb@AR App is dependent on its usability, but also its utility, which can be a combination of several learning benefits. 111
4.21 An overview of the evaluation efforts of the Heb@AR App reported in this section. self-efficacy + usability evaluations during curricular implementations for Training 1, 2, and 3 are combined with lecturer feedback, and summative evaluations. For Training 4 and 5, evaluations grounded in self-determination-theory + usability were supplemented. A timeline of the evaluations is included in the GANTT chart in Figure 4.7. **As multiple cohorts received the AR intervention, the results of the evaluation of Training 2 and 3 are combined results for both cohorts. The results of the evaluation of Training 1 are not the cohort included in the summative evaluation.* 112
4.22 The reported perceived competency to prepare a tocolytic injection on a scale from 0 to 100, before (pre) and after (post) completing the Training 1 of the Heb@AR app during curricular implementation. 114
4.23 The individual SUS scores of Training 1 as data points, its SUS study score as a box plot and the SUS contextualization scales: "Adjective Scale" [28], "Grade Scale" [416] and "Acceptability Scale" [27] for their interpretation (left). The conclusiveness percentage of the SUS study score based on the number of participants [463] (upper right) and the response distribution on the Likert scales of the ten questions of the SUS as stacked bar chart [50] (lower right). 115
4.24 The reported perceived competency to place a permanent catheter in preparation for a c-section, on a scale from 0 to 100, before (pre) and after (post) completing the Training 2 of the Heb@AR app during curricular implementation. 116

4.25 The individual SUS scores of Training 2 as data points, its SUS study score as a box plot and the SUS contextualization scales: “Adjective Scale” [28], “Grade Scale” [416] and “Acceptability Scale” [27] for their interpretation (left). The conclusiveness percentage of the SUS study score based on the number of participants [463] (upper right) and the response distribution on the Likert scales of the ten questions of the SUS as stacked bar chart [50] (lower right). 117
4.26 The individual SUS scores of Training 3 as data points, its SUS study score as a box plot and the SUS contextualization scales: “Adjective Scale” [28], “Grade Scale” [416] and “Acceptability Scale” [27] for their interpretation (left). The conclusiveness percentage of the SUS study score based on the number of participants [463] (upper right) and the response distribution on the Likert scales of the ten questions of the SUS as stacked bar chart [50] (lower right). 119
4.27 The individual SUS scores and SOS study scores as box plots for Training 3 in the single- and multi-user variants. Contextualized on the Adjective Scale by Bangor et al. [27]. 120
4.28 Measures of the motivational variables of the perceived competence, pressure/tension and intrinsic motivation of the IMI [406] before and after completing Training 5 (Female Pelvis). 123
4.29 The individual SUS scores of Training 5 as data points, its SUS study score as a box plot and the SUS contextualization scales: “Adjective Scale” [28], “Grade Scale” [416] and “Acceptability Scale” [27] for their interpretation (left). The conclusiveness percentage of the SUS study score based on the number of participants [463] (upper right) and the response distribution on the Likert scales of the ten questions of the SUS as stacked bar chart [50] (lower right). 124
4.30 Measures of the motivational variables of the perceived competence, pressure/tension and intrinsic motivation of the IMI [406] before and after completing Training 4, the Virtual Reanimation, using a reanimation unit. 127
4.31 The individual SUS scores of Training 4 as data points, its SUS study score as a box plot and the SUS contextualization scales: “Adjective Scale” [28], “Grade Scale” [416] and “Acceptability Scale” [27] for their interpretation (left). The conclusiveness percentage of the SUS study score based on the number of participants [463] (upper right) and the response distribution on the Likert scales of the ten questions of the SUS as stacked bar chart [50] (lower right). 129
4.32 SUS Results from the Lectures for Training 2 and 3, recorded after the practical usage during the workshops, compared to the SUS results from the students recorded after the curricular implementation of the scenario. 131
4.33 Lecturs agreement with the statement *“Overall, I find the didactic approach of the [...] in the AR app well done.”* on a 5 point likert scale, directly after completing each of the 3 trainings. 132

4.34 Lecturs agreement with the statement *"The AR [...] opens up new perspectives for my teaching methods."* on a 5 point likert scale, directly after completing each of the 3 trainings. . . . 133
4.35 Lectures choices, with which objective they would want to carry out the AR trainings, either in preparation for, or as consolidation of, practical SkillsLab trainings. . . . 133
4.36 The normalized OSCE scores of the control station for control and AR intervention cohort. . . . 135
4.37 The normalized OSCE scores of the Station 1, 2, and 3, grouped by the cohort. 136
4.38 The probabilities of normalized OSCE scores for Station 3, the distribution's means, and closest normal distribution, visualized for both cohorts (control cohort, AR intervention cohort). . . . 137
4.39 The retrospectively self-reported competency assessments of students which completed all three AR trainings before and after the training. . . . 139
4.40 Students' assessment of how well specific technical/didactic expressions across trainings helped them for their learning success. . . . 141
4.41 Students' assessment of how well components and didactic delivery methods of the Heb@AR app supported them in attaining their learning goals. . . . 142
4.42 The SUS results for Training 1, 2, and 3 of the Heb@AR App, contextualized with the adjective rating scale. Plotted with the SUS Analysis Toolkit [50] . . . 148
4.43 The SUS results for Training 4 and 5 of the Heb@AR App, contextualized with the adjective rating scale. Plotted with the SUS Analysis Toolkit [50] . . . 150

6.1 TrainAR Framework: The TrainAR authoring tool (left), that is used to create TrainAR trainings (right) using the TrainAR interaction concept and didactic framework. . . . 165
6.2 TrainAR: Didactic framework, interaction concept, and authoring tool. This chapter discusses the Unity-based authoring tool that is built based on AR Foundation, allowing the deployment to both Android and iOS devices through the ARCore and ARKit SDKs, respectively. . . . 167
6.3 The TrainAR interaction concept: After the **Onboarding & Setup** of the **Training Assembly**, the internal state process model provides **Adaptive Instructions** to the user of the application. Additionally, it can provide **Insights** and hints if needed. The user can trigger **User Actions** with virtual objects **Interacting** with them, **Grabbing** them, **Combining** objects or trigger **Custom Actions** (e.g., quizes or mini-games). The internal state process model checks those actions against an internal state catalog, advances the procedural task, and provides **Layered Feedback** to the user. . . . 169

6.4 TrainARs first two trainings, which are fully evaluated: The conduction of a titration experiment in the context of chemical education in project CHARMING (**left**) and the preparing of a tocolytic injection in the midwifery context of project Heb@AR (**right**). 170

6.5 UI components implemented for midwifery AR training scenario. **Top, a-e:** The scenario and interaction onboarding. **Bottom, f-j:** The onboarding of placing the training assembly into the room, menus for rewatching onboarding tutorials, replacing the assembly or exiting the scenario, exemplary warning for AR tracking problems (Others warn of problems with too little illumination or insufficient feature) and the End-screen of the scenario providing contextualized performance feedback and an additional training assessment with professional feedback. 174

6.6 Interactions with virtual objects in the midwifery AR training scenario (Displayed actions do not represent a sequence of events. They visualize specific functionalities described in Section 6.2). **Top, a-e:** Selection of an object with context triggered insights, no selection, a grabbed object, an object before "combining" and a scenario-specific custom action of drawing the syringe. **Bottom, f-j:** Positive feedback for an interaction, positive feedback with additional feedback, negative feedback for an error, an overlay for severe or repeated errors and an example of a custom action in the form of a quiz. 175

6.7 The SUS results for both the explicit and implicit interaction concept from usability study 2 contextualized on the percentile curve of SUS scores according to Kortum et al. [240] (left) and an example of combining a grabbed object with another object through both interaction concepts (right). 179

6.8 System Usability Scale results contextualized on the adjective interpretation scale according to Kortum et al. [240], sorted from lowest to highest average score. Datapoints are represented as dots, mean values as solid lines and average values with their standard deviation as dotted lines. 179

6.9 User Experience Questionnaire scores of all 3 usability studies in relation to the UEQ benchmark data set. 180

6.10 Possibilities for the curricular embedding of TrainAR as pre-trainings to practical on-site trainings, as retention opportunities after practical trainings, or a combination of both. Contextualized with the instructional design theory of Gagne's 5 learning outcomes and 9 events of learning [145, 146]. 184

6.11 The transformation of a procedural action sequence into a TrainAR training scenario. 187

6.12 TrainAR automatically includes onboarding screens and technical utility for (**a**) grabbing objects, (**b**) interacting with objects, (**c**) combining objects, (**d**) scanning the training area, and (**e**) placing the training assembly. 192

6.13 The basic AR actions of the TrainAR Interaction Concept that allow trainees to (**f**) select and (**g**) grab TrainAR objects. Selected or grabbed objects can be (**h**) interacted with. Grabbed objects can be (**i**) combined with another TrainAR object by overlapping them. 193
6.14 Custom Actions include UI-based "quiz" actions such as (**j**) text input fields, (**k**) questionnaire elements, and (**l**) list selection elements. Authors can also create their own UI overlays that trigger custom actions, for example, (**m**) a slider to pull up a syringe (see Section 7.1.1). 194
6.15 The output modalities of the TrainAR Interaction Concept consist of (**n**) feedback to aid technical problems, (**o**) instructions and progress indicators, (**p**) expert tips and insights, (**q**) error feedback overlays, and (**r**) a training summary at the end of each training. 194
6.16 The TrainAR Authoring Tool layout combines the Unity Inspector and projects folder with the TrainAR Training Assembly, TrainAR Visual Statemachine, and a Scene preview, allowing authors to create procedural TrainAR trainings. . . . 196
6.17 The concepts introduced in the TrainAR interaction concept and didactic framework (**left**) and their corresponding Statemachine nodes in the TrainAR Authoring Tool, which can be used to author the TrainAR training. 198
6.18 The modal window for the conversion of 3D objects from a variety of sources and in different formats into a consistent TrainAR Object, which is simplified, compatible with TrainAR, automatically inherits all intended behaviors, and can be referenced in the TrainAR Visual Statemachine. 199
6.19 An abstract conceptual visualization of the architecture of an authored TrainAR training. 201

7.1 To show the usefulness of the TrainAR authoring tool, its utility and usability have to be evaluated. The utility of the authoring tool itself is the creation of trainings that themselves have to be useful, meaning they also have to prove their utility and usability. 210
7.2 Five exemplary TrainAR trainings. (**1**) The preparation of a tocolytic injection in the context of academic midwifery, (**2**) the denomination and contextualization of German and Latin terminology of the female pelvis, (**3**) a titration experiment in the context of chemical engineering, (**4**) the exploration of chemical reactions in early school education, and (**5**) the exploration of ripeness as a chemistry learning game for children. 211
7.3 The perceived usability of the exemplary TrainAR scenarios in Figure 7.2 in the form of SUS study scores, taken from primary sources evaluating TrainAR trainings [18, 52, 55, 114] and plotted with the SUS Analysis Toolkit [50]. 212
7.4 Average Task Completion Time (TCT) of each of the student groups (Computer Science, Media-Technology, and non-technical students) for each of the three authoring tasks the participants authored during the study. 218

7.5 The average Task Completion Time (TCT) after each occurrence of one of the ten sub-task types for each of the student groups (Computer Science, Media-Technology, and non-technical students). 218
7.6 Average perceived cognitive load measures with the NASA rTLX [188] for each authoring task grouped by participants' study program, normalized to a 1–100 score. 221
7.7 The perceived usability of the TrainAR authoring tool reported as SUS scores, grouped by the study program of the participants, plotted with the SUS Analysis Toolkit [50]. 221
7.8 The training content of the radio relay assembly, including (a) an antenna, screws, and sealing ring, (b) an azimuth fine-adjuster, (c & d & h) brackets, (e) pipe clamps, (f) the main digital processing unit (MDU), (g) an ethernet cable insertion, (i) a transition element with sealing rings, (j) a mount for horizontal and vertical alignment of the antenna, (k) connection cable between MDU and ODU, and (l) the outdoor unit (ODU). The figure, depicted content, and descriptions are adopted from Jörg Eggeling [123], licensed under CC-BY 4.0 227

1 The first authoring task, the construction of a lamp setup, that was used in the systematic TrainAR Authoring Tool usability evaluation study. Translated and adapted from [37]. 342
2 The second authoring task, the traditional East Frisian tea ceremony, that was used in the systematic TrainAR Authoring Tool usability evaluation study. Translated and adapted from [37]. 343
3 The third authoring task, the preparation of an injection, that was used in the systematic TrainAR Authoring Tool usability evaluation study. Translated and adapted from [37]. 345

List of Tables

2.1 The six databases used to retrieve articles, the types of searches used (title, abstract, keyword, text or combination searches) and the search strings that were used to retrieve publications. A combined 624 publications (including duplicates) were retrieved from this initial database search. 16

2.2 The overlap of the retrieved literature from this systematic review compared to the 5 most similar reviews previously published. Numbers in brackets, followed by an asterisk (*), represent the overlap, when different publications on the same AR authoring tool are counted as overlap. 25

2.3 The 26 primary dimensions, the literature, is mapped on. Each is shown with the coding scheme used (descriptive, deductive, or inductive), the levels for the deductively coded themes, and descriptions of the questions this dimension tries to answer. The table is sorted alphabetically and colored by the coding scheme used. 28

2.4 The 13 AR authoring tools which are available as open source, with their publication year, date of last commit (accessed on the 22.06.2023) and link to their repositories. 41

4.1 The 5 AR trainings included in the Heb@AR app, with their training name, training type, abstract learning goal, whether they need physical material during the training, a rough estimate of the training duration, and individual other features of the trainings. 94

4.2 Results of the three paired t-test examining the change of the motivational variables of the perceived competence, pressure/tension and intrinsic motivation of the IMI [406] before and after completing the Pelvis AR training. 123

4.3 Results of the paired t-tests examining the change of the motivational variables (Competence, pressure, and intrinsic motivation) of the IMI [406] before and after completing Training 4. 128

7.1 The five exemplary TrainAR trainings shown in Figure 7.2 and which parts of TrainAR (Interaction concept, didactic framework, or authoring framework) they utilize for their use case. 211

7.2 The average Task Completion Rate (TCR) for each of the three authoring tasks, grouped by the participants' study program and reported on 4 levels: without any help, with a hint of where in the documentation the solution is described, with a predefined solution hint, or with explicit help. 220

1 The official PRISMA checklist, transferred from [361]. 309

2 The official PRISMA-ScR checklist, transferred from [461]. 311

3 The 293 publications of the literature review, grouped by the search strategy (database search, forward snowballing, backward snowballing) and iteration, they were found with. . . . 312
4 The 293 publications of the literature review, grouped by the year they were published in. . . . 313
5 The 293 publications of the literature review, grouped by the number of authors who contributed to the publication. . . . 314
6 The 293 publications of the literature review, grouped by publication format, they were published in. . . . 315
7 The 293 publications of the literature review, grouped by Publisher they were published with. Publishers with less than 5 publications published in them are grouped in "Others". . . . 316
8 The 293 publications of the literature review, grouped by whether the AR authoring tool is main contribution, the main or a side contribution of the paper. 317
9 The 293 publications of the literature review, grouped by the deployment purpose. 318
10 The 293 publications of the literature review, grouped by context they were deployed in. . . . 319
11 The 293 publications of the literature review, grouped by which part of the process was evaluated by the authors of the publication in terms of usability. . . . 320
12 The 293 publications of the literature review, grouped by whether they are available as binaries or even open source. . . . 321
13 The 293 publications of the literature review, grouped by the (described or inferrable) intended author of the AR construct. . . . 322
14 The 293 publications of the literature review, grouped by the hardware supported for the authoring. As some AR authoring tools support multiple hardware platforms, this includes overlap. . . . 323
15 The 293 publications of the literature review, grouped by the markup notation used. . . . 324
16 The 293 publications of the literature review, grouped by whether they are standalone applications or plugins for host software. . . . 325
17 The 293 publications of the literature review, grouped by whether the AR authoring tool provides preview functionality of the AR content. . . . 326
18 The 293 publications of the literature review, grouped by whether content is authored in-situ or decontextualized. . . . 327
19 The publications of the literature review, grouped by primary authoring interaction concept. Only publications which are using concepts beyond the "traditional interaction concepts" (163 publications) of the authoring hardware utilized are mapped in this table. . . . 328

20 The 293 publications of the literature review, grouped by whether the application that is used to author the AR construct is also the application for its usage (internal), or one application authors content to be used with a second application (external). . . . 329
21 The 293 publications of the literature review, grouped whether the AR constructs are distributed locally or through a server. . . . 330
22 The 293 publications of the literature review, grouped by the (described or inferrable) intended user of the AR construct. . . . 331
23 The 293 publications of the literature review, grouped by the hardware supported for the usage of the authored AR constructs. As some authored AR constructs can be used on multiple hardware platforms, this includes overlap. . . . 332
24 The 293 publications of the literature review, grouped by main interaction concept used to interact with the authored AR constructs on the usage hardware. . 333
25 The 293 publications of the literature reviewed, grouped by the AR content type they support. As some authored AR constructs utilize multiple content types, this includes overlap. . . . 334
26 The 293 publications of the literature review, grouped by the complexity of provided content; therefore, whether it is sequential or constant. . . . 335
27 The 293 publications of the literature review, grouped by combination of tracking techniques utilized to contextualize the AR content. . . . 336
28 All System Usability Scale metrics for the formative usability evaluations for the design of the TrainAR interaction concept reported in Section 6.3. Calculated with the SUS Analysis Toolkit [50]. . . . 337
29 All System Usability Scale metrics for Training 1 (Tocolysis), Training 2 (Sectio), Training 3 (Reanimation) and the multi-user version of Training 3, which are reported in Section 4.6.2, 4.6.3, and 4.6.4. Calculated with the SUS Analysis Toolkit [50]. . . . 338
30 All System Usability Scale metrics for Training 4 (Virtual reanimation), Training 5 (Pelvis Termini), reported in Sections 4.6.6 and 4.6.5. Calculated with the SUS Analysis Toolkit [50]. . . . 339
31 All System Usability Scale metrics for Training 2 and Training 3 from the lecturers' perspective after completing the trainings during workshops (see Section 4.6.7). Calculated with the SUS Analysis Toolkit [50]. . . . 340
32 All System Usability Scale metrics for the TrainAR Usability Study evaluation reported in Section 7.2.2. Calculated with the SUS Analysis Toolkit [50]. . . . 341

Bibliography

1. D. Abawi, R. Dorner, M. Raller, and J. Zauner. “Efficient mixed reality application development”. In: *1st European Conference on Visual Media Production (CVMP) 2004*. 2004, pp. 289–294.
2. D. Abawi, R. Dörner, and P. Grimm. “A Component-based Authoring Environment for Creating Multimedia-RichMixed Reality”. In: *Eurographics Multimedia Workshop*. Ed. by N. Correia, J. Jorge, T. Chambel, and Z. Pan. The Eurographics Association, 2004. ISBN: 3-905673-17-7. DOI: 10.2312/EGMM/MM04/031-040.
3. D. F. Abawi, S. Reinhold, and R. Dörner. “A Toolkit for Authoring Non-linear Storytelling Environments Using Mixed Reality”. In: *Technologies for Interactive Digital Storytelling and Entertainment*. Ed. by S. Göbel, U. Spierling, A. Hoffmann, I. Iurgel, O. Schneider, J. Dechau, and A. Feix. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004, pp. 113–118. ISBN: 978-3-540-27797-2.
4. J. L. B. Acosta, S. M. B. Navarro, R. F. Gesa, and K. Kinshuk. “Framework for designing motivational augmented reality applications in vocational education and training”. *Australasian Journal of Educational Technology* 35:3, 2019. ISSN: 1449-5554, 1449-3098. DOI: 10.14742/ajet.4182.
5. S. Ahn, H. Ko, and B. Yoo. “Webizing mobile augmented reality content”. *New Review of Hypermedia and Multimedia* 20:1, 2014, pp. 79–100. DOI: 10.1080/13614568.2013.857727.
6. M. Akçayır and G. Akçayır. “Advantages and challenges associated with augmented reality for education: A systematic review of the literature”. *Educational Research Review* 20, 2017, pp. 1–11. ISSN: 1747-938X. DOI: 10.1016/j.edurev.2016.11.002.
7. M. V. Almeida and J. M. S. Dias. “Kit Assembly Helper in Augmented Reality”, 2006. Ed. by P. Brunet, N. Correia, and G. Baranoski. DOI: 10.2312/LocalChapterEvents/siacg/siacg06/089-096.
8. N. M. Alzahrani and S. Lajmi. “AugmentedBook: A Collaborative E-Learning Augmented Reality Platform”. In: *Human Interaction and Emerging Technologies*. Ed. by T. Ahram, R. Taiar, S. Colson, and A. Choplin. Springer International Publishing, Cham, 2020, pp. 282–288. ISBN: 978-3-030-25629-6. DOI: 10.1007/978-3-030-25629-6_44.
9. N. M. Alzahrani. “Augmented Reality: A Systematic Review of Its Benefits and Challenges in E-learning Contexts”. *Applied Sciences* 10:16, 2020. ISSN: 2076-3417. DOI: 10.3390/app10165660.
10. K. Anagnostou and P. Vlamos. “Square AR: Using Augmented Reality for Urban Planning”. In: *2011 Third International Conference on Games and Virtual Worlds for Serious Applications*. 2011, pp. 128–131. DOI: 10.1109/VS-GAMES.2011.24.

11. L. W. Anderson, B. S. Bloom, et al. *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman, 2001.
12. Y. Apaza, R. Tumailla, W. Hancco, A. Paz-Valderrama, C. Corrales-Delgado, and M. Loaiza. "Systematic Mapping Study on High-Level Content Design Frameworks for Augmented Reality". In: *2018 20th Symposium on Virtual and Augmented Reality (SVR)*. 2018, pp. 192–201. DOI: 10.1109/SVR.2018.00037.
13. Y. Apaza-Yllachura, A. Paz-Valderrama, and C. Corrales-Delgado. "SimpleAR: Augmented Reality high-level content design framework using visual programming". In: *2019 38th International Conference of the Chilean Computer Science Society (SCCC)*. 2019, pp. 1–7. DOI: 10.1109/SCCC49216.2019.8966427.
14. Apple. *ARKit*. https://developer.apple.com/arkit/. Accessed: 08.04.2023. 2017.
15. J. Arango-López, C. C. Cerón Valdivieso, C. A. Collazos, F. L. Gutiérrez Vela, and F. Moreira. "CREANDO: Tool for creating pervasive games to increase the learning motivation in higher education students". *Telematics and Informatics* 38, 2019, pp. 62–73. ISSN: 0736-5853. DOI: 10.1016/j.tele.2018.08.005.
16. L. Arenas, T. Zarraonandia, P. Díaz, and I. Aedo. "A Platform for Supporting the Development of Mixed Reality Environments for Educational Games". In: *Learning and Collaboration Technologies*. Ed. by P. Zaphiris and A. Ioannou. Springer International Publishing, Cham, 2015, pp. 537–548. ISBN: 978-3-319-20609-7. DOI: 10.1007/978-3-319-20609-7_51.
17. R. Arora, R. Habib Kazi, T. Grossman, G. Fitzmaurice, and K. Singh. "SymbiosisSketch: Combining 2D & 3D Sketching for Designing Detailed 3D Objects in Situ". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18. Association for Computing Machinery, Montreal QC, Canada, 2018, pp. 1–15. ISBN: 9781450356206. DOI: 10.1145/3173574.3173759.
18. M. Arztmann, J. L. D. Alfaro, J. Blattgerste, J. Jeuring, and P. van Puyvelde. "Marie's ChemLab: a Mobile Augmented Reality Game to Teach Basic Chemistry to Children". In: *European Conference on Games Based Learning*. ECGBL '22. Lisbon, Portugal, 2022.
19. A.-r. Asadi and R. Hemadi. "Augmented Reality Game Creator for on-site Job Training". In: *2018 2nd National and 1st International Digital Games Research Conference: Trends, Technologies, and Applications (DGRC)*. 2018, pp. 128–133. DOI: 10.1109/DGRC.2018.8712020.
20. N. Ashtari, A. Bunt, J. McGrenere, M. Nebeling, and P. K. Chilana. "Creating augmented and virtual reality applications: Current practices, challenges, and opportunities". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, 2020, pp. 1–13. DOI: 10.1145/3313831.3376722.

21. D. Avola, L. Cinque, G. L. Foresti, C. Mercuri, and D. Pannone. "A Practical Framework for the Development of Augmented Reality Applications by Using ArUco Markers". In: *Proceedings of the 5th International Conference on Pattern Recognition Applications and Methods*. ICPRAM 2016. SCITEPRESS - Science and Technology Publications, Lda, Rome, Italy, 2016, pp. 645–654. ISBN: 9789897581731. DOI: 10.5220/0005755806450654.

22. R. T. Azuma. "A survey of augmented reality". *Presence: Teleoperators & Virtual Environments* 6:4, 1997, pp. 355–385. DOI: 10.1162/pres.1997.6.4.355.

23. B. Bach, R. Sicat, H. Pfister, and A. Quigley. "Drawing into the AR-CANVAS: Designing embedded visualizations for augmented reality". In: *Workshop on Immersive Analytics, IEEE Vis*. 2017.

24. D. Badampudi, C. Wohlin, and K. Petersen. "Experiences from using snowballing and database searches in systematic literature studies". In: *Proceedings of the 19th international conference on evaluation and assessment in software engineering*. ACM, 2015, pp. 1–10. DOI: 10.1145/2745802.2745818.

25. H. Bae, M. Golparvar-Fard, and J. White. "Image-Based Localization and Content Authoring in Structure-from-Motion Point Cloud Models for Real-Time Field Reporting Applications". *Journal of Computing in Civil Engineering* 29:4, 2015, B4014008. DOI: 10.1061/(ASCE)CP.1943-5487.0000392.

26. S. Balcisoy, M. Kallmann, P. Fua, and D. Thalmann. "A Framework for Rapid Evaluation of Prototypes with Augmented Reality". In: *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*. VRST '00. Association for Computing Machinery, Seoul, Korea, 2000, pp. 61–66. ISBN: 1581133162. DOI: 10.1145/502390.502403.

27. A. Bangor, P. Kortum, and J. Miller. "Determining what individual SUS scores mean: Adding an adjective rating scale". *Journal of usability studies* 4:3, 2009, pp. 114–123.

28. A. Bangor, P. T. Kortum, and J. T. Miller. "An empirical evaluation of the system usability scale". *Intl. Journal of Human–Computer Interaction* 24:6, 2008, pp. 574–594. DOI: 10.1080/10447310802205776.

29. I. Barakonyi and D. Schmalstieg. "Immersive content authoring for computer entertainment in augmented reality". In: *DVD-ROM Proc. Int. Workshop on Mixed Reality Technology for Filmmaking*. 2006, pp. 26–30.

30. J. Barbadillo, N. Barrena, V. Goñi, and J. R. Sánchez. "Collaborative E-Learning Framework for Creating Augmented Reality Mobile Educational Activities". In: *Ubiquitous Computing and Ambient Intelligence. Personalisation and User Adapted Services*. Ed. by R. Hervás, S. Lee, C. Nugent, and J. Bravo. Springer International Publishing, Cham, 2014, pp. 52–59. ISBN: 978-3-319-13102-3. DOI: 10.1007/978-3-319-13102-3_11.

31. J. Barbadillo and J. R. Sánchez. "A Web3D Authoring Tool for Augmented Reality Mobile Applications". In: *Proceedings of the 18th International Conference on 3D Web Technology*. Web3D '13. Association for Computing Machinery, San Sebastian, Spain, 2013, p. 206. ISBN: 9781450321334. DOI: 10.1145/2466533.2466564.

32. L. Barbieri, E. Marino, and F. Bruno. "A Knowledge-Based Augmented Reality Tool for Managing Design Variations". In: *Advances on Mechanics, Design Engineering and Manufacturing II*. Ed. by F. Cavas-Martínez, B. Eynard, F. J. Fernández Cañavate, D. G. Fernández-Pacheco, P. Morer, and V. Nigrelli. Springer International Publishing, Cham, 2019, pp. 430–439. ISBN: 978-3-030-12346-8. DOI: 10.1007/978-3-030-12346-8_42.

33. A. Barone Rodrigues, D. R. C. Dias, V. F. Martins, P. A. Bressan, and M. de Paiva Guimarães. "WebAR: A Web-Augmented Reality-Based Authoring Tool with Experience API Support for Educational Applications". In: *Universal Access in Human–Computer Interaction. Designing Novel Interactions*. Ed. by M. Antona and C. Stephanidis. Springer International Publishing, Cham, 2017, pp. 118–128. ISBN: 978-3-319-58703-5. DOI: 10.1007/978-3-319-58703-5_9.

34. N. Barrena, A. Navarro, S. García, and D. Oyarzun. "CoolTour: VR and AR Authoring Tool to Create Cultural Experiences". In: *Intelligent Interactive Multimedia Systems and Services 2016*. Ed. by G. D. Pietro, L. Gallo, R. J. Howlett, and L. C. Jain. Springer International Publishing, Cham, 2016, pp. 483–489. ISBN: 978-3-319-39345-2. DOI: 10.1007/978-3-319-39345-2_42.

35. N. H. Bauer, A. Bernloehr, K. Vogel, and T. Willmeroth. *Forschungsprojekt Heb@AR, Teilprojekt Hebammenwissenschaft*. Project partner during the BMBF (Federal Ministry of Education and Research, Germany) research project Heb@AR (16DHB3019).

36. P. Bégout, T. Duval, S. Kubicki, B. Charbonnier, and E. Bricard. "WAAT: A Workstation AR Authoring Tool for Industry 4.0". In: *Augmented Reality, Virtual Reality, and Computer Graphics*. Ed. by L. T. De Paolis and P. Bourdot. Springer International Publishing, Cham, 2020, pp. 304–320. ISBN: 978-3-030-58468-9.

37. J. Behrends. "Anforderungsanalyse, Entwicklung und Evaluation von Augmented Reality Autorenwerkzeugen am Beispiel TrainAR". Masters's Thesis. University of Applied Sciences Emden/Leer, 2022.

38. J. Behrends. "Konzeption und Entwicklung eines Webbasierten System Usability Scale Analyse Toolkits". Bachelors's Thesis. University of Applied Sciences Emden/Leer, 2021.

39. A. H. Behzadan and V. R. Kamat. "Automated Generation of Operations Level Construction Animations in Outdoor Augmented Reality". *Journal of Computing in Civil Engineering* 23:6, 2009, pp. 405–417. DOI: 10.1061/(ASCE)0887-3801(2009)23:6(405).

40. P. Belimpasakis, Y. You, and P. Selonen. "Enabling Rapid Creation of Content for Consumption in Mobile Augmented Reality". In: *2010 Fourth International Conference on Next Generation Mobile Applications, Services and Technologies*. 2010, pp. 1–6. DOI: 10.1109/NGMAST.2010.13.

41. O. Ben-Kiki, C. Evans, and B. Ingerson. "Yaml ain't markup language (yaml™) version 1.1". *Working Draft 2008* 5, 2009, p. 11.

42. O. Bergig, N. Hagbi, J. El-Sana, and M. Billinghurst. "In-place 3D sketching for authoring and augmenting mechanical systems". In: *2009 8th IEEE International Symposium on Mixed and Augmented Reality*. 2009, pp. 87–94. DOI: 10.1109/ISMAR.2009.5336490.

43. R. Berry, N. Hikawa, M. Makino, M. Suzuki, and T. Furuya. "Authoring Augmented Reality: A Code-Free Approach". In: *ACM SIGGRAPH 2004 Posters*. SIGGRAPH '04. Association for Computing Machinery, Los Angeles, California, 2004, p. 43. ISBN: 1581138962. DOI: 10.1145/1186415.1186465.

44. R. Berry, M. Makino, N. Hikawa, M. Naemura, Y. Pisan, and E. Edmonds. "Programming in the world". *Digital Creativity* 17:1, 2006, pp. 36–48. DOI: 10.1080/14626260600665728.

45. D. R. Berryman. "Augmented Reality: A Review". *Medical Reference Services Quarterly* 31:2, 2012. PMID: 22559183, pp. 212–218. DOI: 10.1080/02763869.2012.670604. eprint: https://doi.org/10.1080/02763869.2012.670604.

46. F. Bhanji, R. Gottesman, W. de Grave, Y. Steinert, and L. R. Winer. "The Retrospective Pre–Post: A Practical Method to Evaluate Learning from an Educational Program". *Academic Emergency Medicine* 19:2, 2012, pp. 189–194. DOI: 10.1111/j.1553-2712.2011.01270.x.

47. B. Bhattacharya and E. H. Winer. "Augmented reality via expert demonstration authoring (AREDA)". *Computers in Industry* 105, 2019, pp. 61–79. ISSN: 0166-3615. DOI: 10.1016/j.compind.2018.04.021.

48. M. L. Bianco, S. Pedell, and G. Renda. "Augmented Reality and Home Modifications: A Tool to Empower Older Adults in Fall Prevention". In: *Proceedings of the 28th Australian Conference on Computer-Human Interaction*. OzCHI '16. Association for Computing Machinery, Launceston, Tasmania, Australia, 2016, pp. 499–507. ISBN: 9781450346184. DOI: 10.1145/3010915.3010929.

49. J. Blattgerste. *Data for the Dissertation "The Design Space of Augmented Reality Authoring Tools and its Exploration for the Procedural Training Context"*. 2024. DOI: 10.4119/unibi/2986422.

50. J. Blattgerste, J. Behrends, and T. Pfeiffer. "A Web-Based Analysis Toolkit for the System Usability Scale". In: *Proceedings of the 15th International Conference on PErvasive Technologies Related to Assistive Environments*. PETRA '22. Association for Computing Machinery, Corfu, Greece, 2022, pp. 237–246. ISBN: 9781450396318. DOI: 10.1145/3529190.3529216.

51. J. Blattgerste, J. Behrends, and T. Pfeiffer. “TrainAR: An Open-Source Visual Scripting-Based Authoring Tool for Procedural Mobile Augmented Reality Trainings”. *Information* 14:4, 2023. ISSN: 2078-2489. DOI: 10.3390/info14040219.

52. J. Blattgerste, J. Franssen, M. Arztmann, and T. Pfeiffer. “Motivational benefits and usability of a handheld Augmented Reality game for anatomy learning”. In: *2022 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. 2022, pp. 266–274. DOI: 10.1109/AIVR56993.2022.00056.

53. J. Blattgerste, C. Lewa, K. Vogel, T. Willmeroth, S. Janßen, J. Franssen, J. Behrends, M. Joswig, T. Schäfer, N. H. Bauer, A. Bernloehr, and T. Pfeiffer. “Die Heb@AR App - Eine Android & iOS App mit Augmented Reality Trainings für selbstbestimmtes und curriculares Lernen in der hochschulischen Hebammenausbildung”. In: *Wettbewerbsband AVRiL 2022*. Ed. by H. Söbke and R. Zender. Gesellschaft für Informatik e.V., Bonn, 2022, pp. 4–9. DOI: 10.18420/avril2022_01.

54. J. Blattgerste, K. Luksch, C. Lewa, M. Kunzendorf, N. H. Bauer, A. Bernloehr, M. Joswig, T. Schäfer, and T. Pfeiffer. “Project Heb@AR: Exploring handheld Augmented Reality training to supplement academic midwifery education”. In: *DELFI 2020 – Die 18. Fachtagung Bildungstechnologien der Gesellschaft für Informatik e.V.* Ed. by R. Zender, D. Ifenthaler, T. Leonhardt, and C. Schumacher. Gesellschaft für Informatik e.V., Bonn, 2020, pp. 103–108. ISBN: 978-3-88579-702-9.

55. J. Blattgerste, K. Luksch, C. Lewa, and T. Pfeiffer. “TrainAR: A Scalable Interaction Concept and Didactic Framework for Procedural Trainings Using Handheld Augmented Reality”. *Multimodal Technologies and Interaction* 5:7, 2021. ISSN: 2414-4088. DOI: 10.3390/mti5070030.

56. J. Blattgerste and T. Pfeiffer. “TrainAR: Ein Augmented Reality Training Autorensystem”. In: *Wettbewerbsband AVRiL 2022*. Ed. by H. Söbke and R. Zender. Gesellschaft für Informatik e.V., Bonn, 2022, pp. 40–45. DOI: 10.18420/avril2022_06.

57. J. Blattgerste, P. Renner, and T. Pfeiffer. “Augmented Reality Action Assistance and Learning for Cognitively Impaired People: A Systematic Literature Review”. In: *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments*. PETRA ’19. Association for Computing Machinery, Rhodes, Greece, 2019, pp. 270–279. ISBN: 9781450362320. DOI: 10.1145/3316782.3316789.

58. J. Blattgerste, P. Renner, and T. Pfeiffer. “Authorable Augmented Reality Instructions for Assistance and Training in Work Environments”. In: *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia*. MUM ’19. Association for Computing Machinery, Pisa, Italy, 2019. ISBN: 9781450376242. DOI: 10.1145/3365610.3365646.

59. J. Blattgerste, P. Renner, B. Strenge, and T. Pfeiffer. "In-Situ Instructions Exceed Side-by-Side Instructions in Augmented Reality Assisted Assembly". In: *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference*. PETRA '18. Association for Computing Machinery, Corfu, Greece, 2018, pp. 133–140. ISBN: 9781450363907. DOI: 10.1145/3197768.3197778.

60. J. Blattgerste, K. Vogel, C. Lewa, T. Willmeroth, M. Joswig, T. Schäfer, N. H. Bauer, A. Bernloehr, and T. Pfeiffer. "The Heb@AR App – Five Augmented Reality Trainings for Self-Directed Learning in Academic Midwifery Education". In: *20. Fachtagung Bildungstechnologien (DELFI)*. Ed. by P. A. Henning, M. Striewe, and M. Wölfel. Gesellschaft für Informatik e.V., Bonn, 2022, pp. 245–246. ISBN: 978-3-88579-716-6. DOI: 10.18420/delfi2022-052.

61. J. Bortz and N. Döring. *Forschungsmethoden und Evaluation für Human-und Sozialwissenschaftler: Limitierte Sonderausgabe*. Springer-Verlag, 2007.

62. J. Brings, M. Daun, M. Kempe, and T. Weyer. "On Different Search Methods for Systematic Literature Reviews and Maps: Experiences from a Literature Search on Validation and Verification of Emergent Behavior". In: *Proceedings of the 22nd International Conference on Evaluation and Assessment in Software Engineering 2018*. EASE'18. Association for Computing Machinery, Christchurch, New Zealand, 2018, pp. 35–45. ISBN: 9781450364034. DOI: 10.1145/3210459.3210463.

63. W. Broll, J. Herling, and L. Blum. "Interactive Bits: Prototyping of Mixed Reality Applications and Interaction Techniques through Visual Programming". In: *2008 IEEE Symposium on 3D User Interfaces*. 2008, pp. 109–115. DOI: 10.1109/3DUI.2008.4476600.

64. W. Broll, I. Lindt, and M. Wittkämper. "Mixed Reality User Interface Description Language". In: *ACM SIGGRAPH 2006 Research Posters*. SIGGRAPH '06. Association for Computing Machinery, Boston, Massachusetts, 2006, 156–es. ISBN: 1595933646. DOI: 10.1145/1179622.1179802.

65. J. Brooke. "SUS: A 'Quick and Dirty' Usability Scale". *Usability Evaluation in Industry* 189, 1996. DOI: 10.1201/9781498710411-35.

66. J. Brooke. "SUS: a retrospective". *Journal of usability studies* 8:2, 2013, pp. 29–40.

67. F. Bruno, L. Barbieri, E. Marino, M. Muzzupappa, L. D'Oriano, and B. Colacino. "An augmented reality tool to detect and annotate design variations in an Industry 4.0 approach". *The International Journal of Advanced Manufacturing Technology* 105:1, 2019, pp. 875–887.

68. C. Buabeng-Andoh. "Factors influencing teachers adoption and integration of information and communication technology into teaching: A review of the literature". *International Journal of Education and Development using ICT* 8:1, 2012. ISSN: 1814-0556.

69. D. Budgen, M. Turner, P. Brereton, and B. A. Kitchenham. "Using Mapping Studies in Software Engineering." In: *Ppig*. Vol. 8. 2008, pp. 195–204.

70. Bundesministerium des Inneren Bundesverwaltungsamt. *Handbuch für Organisationuntersuchungen und Personalbedarfsermittlung*. Berlin, 2018.

71. S. Buntin. "Generating Furniture for Augmented Reality Applications Using Natural Language". In: *Proceedings of the 8th Augmented Human International Conference*. AH '17. Association for Computing Machinery, Silicon Valley, California, USA, 2017. ISBN: 9781450348355. DOI: 10.1145/3041164.3041189.

72. S. Büttner, H. Mucha, M. Funk, T. Kosch, M. Aehnelt, S. Robert, and C. Röcker. "The Design Space of Augmented and Virtual Reality Applications for Assistive Environments in Manufacturing: A Visual Approach". In: *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments*. PETRA '17. Association for Computing Machinery, Island of Rhodes, Greece, 2017, pp. 433–440. ISBN: 9781450352277. DOI: 10.1145/3056540.3076193.

73. J. D. Camba and M. Contero. "From reality to augmented reality: Rapid strategies for developing marker-based AR content using image capturing and authoring tools". In: *2015 IEEE Frontiers in Education Conference (FIE)*. 2015, pp. 1–6. DOI: 10.1109/FIE.2015.7344162.

74. J. Capper. "E-learning growth and promise for the developing world". *TechKnowLogia* 2:2, 2001, pp. 7–10.

75. G. Caruso, G. M. Re, M. Carulli, and M. Bordegoni. "Novel Augmented Reality system for Contract Design Sector". *Computer-Aided Design and Applications* 11:4, 2014, pp. 389–398. DOI: 10.1080/16864360.2014.881181.

76. R. I. B. Castillo, V. G. C. Sánchez, O. O. V. Villegas, and A. L. Lozoya. "Node based visual editor for mobile augmented reality". *International Journal of Combinatorial Optimization Problems and Informatics* 7:2, 2016, pp. 35–48. ISSN: 2007-1558.

77. M. Cavallo and A. G. Forbes. "CAVE-AR: A VR Authoring System to Interactively Design, Simulate, and Debug Multi-user AR Experiences". In: *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 2019, pp. 872–873. DOI: 10.1109/VR.2019.8798148.

78. R. Chalumattu, S. Schaub-Meyer, R. Wiethuchter, S. Klingler, and M. Gross. "Simplifying the Process of Creating Augmented Outdoor Scenes". In: *2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. 2020, pp. 1–6. DOI: 10.1109/ICMEW46912.2020.9106030.

79. Y. S. Chang, B. Nuernberger, B. Luan, and T. Höllerer. "Evaluating gesture-based augmented reality annotation". In: *2017 IEEE Symposium on 3D User Interfaces (3DUI)*. 2017, pp. 182–185. DOI: 10.1109/3DUI.2017.7893337.

80. H.-M. Chen and P.-H. Huang. "3D AR-based modeling for discrete-event simulation of transport operations in construction". *Automation in Construction* 33, 2013, pp. 123–136. ISSN: 0926-5805. DOI: 10.1016/j.autcon.2012.09.015.

81. Z. Chen, Y. Su, Y. Wang, Q. Wang, H. Qu, and Y. Wu. "MARVisT: Authoring Glyph-Based Visualization in Mobile Augmented Reality". *IEEE Transactions on Visualization and Computer Graphics* 26:8, 2020, pp. 2645–2658. DOI: 10.1109/TVCG.2019.2892415.

82. Z. Chen, W. Tong, Q. Wang, B. Bach, and H. Qu. "Augmenting Static Visualizations with PapARVis Designer". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI '20. Association for Computing Machinery, Honolulu, HI, USA, 2020, pp. 1–12. ISBN: 9781450367080. DOI: 10.1145/3313831.3376436.

83. K.-H. Cheng and C.-C. Tsai. "Affordances of augmented reality in science learning: Suggestions for future research". *Journal of science education and technology* 22:4, 2013, pp. 449–462. DOI: 10.1007/s10956-012-9405-9.

84. W. Chinthammit and A. Thomas. "iFiction: Mobile technology, new media, Mixed Reality and literary creativity in English teaching". In: *2012 IEEE International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities (ISMAR-AMH)*. 2012, pp. 39–46. DOI: 10.1109/ISMAR-AMH.2012.6483987.

85. H. Choi, S. Oh, S. Choi, and J. Yoon. "Innovation Topic Analysis of Technology: The Case of Augmented Reality Patents". *IEEE Access* 6, 2018, pp. 16119–16137. DOI: 10.1109/ACCESS.2018.2807622.

86. J. Choi, Y. Kim, M. Lee, G. J. Kim, Y. Nam, and Y. Kwon. "k-MART: Authoring tool for mixed reality contents". In: *2010 IEEE International Symposium on Mixed and Augmented Reality*. 2010, pp. 219–220. DOI: 10.1109/ISMAR.2010.5643576.

87. J. Christian, H. Krieger, A. Holzinger, and R. Behringer. "Virtual and Mixed Reality Interfaces for e-Training: Examples of Applications in Light Aircraft Maintenance". In: *Universal Access in Human-Computer Interaction. Applications and Services*. Ed. by C. Stephanidis. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007, pp. 520–529. ISBN: 978-3-540-73283-9. DOI: 10.1007/978-3-540-73283-9_58.

88. K. Chu, W. Lu, K. Oka, K. Takashima, and Y. Kitamura. "A Study on AR Authoring using Mobile Devices for Educators". In: *ICAT-EGVE 2018 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*. Ed. by G. Bruder, S. Yoshimoto, and S. Cobb. The Eurographics Association, 2018. ISBN: 978-3-03868-058-1. DOI: 10.2312/egve.20181331.

89. S. Claudino Daffara, A. Brewer, B. Thoravi Kumaravel, and B. Hartmann. "Living Paper: Authoring AR Narratives Across Digital and Tangible Media". In: *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI EA '20. Association for Computing Machinery, Honolulu, HI, USA, 2020, pp. 1–10. ISBN: 9781450368193. DOI: 10.1145/3334480.3383091.

90. W. S. Cleveland and R. McGill. "The Many Faces of a Scatterplot". *Journal of the American Statistical Association* 79:388, 1984, pp. 807–822. DOI: 10.1080/01621459.1984.10477098.

91. E. Coelho, S. Julier, and B. MacIntyre. "OSGAR: a scene graph with uncertain transformations". In: *Third IEEE and ACM International Symposium on Mixed and Augmented Reality*. 2004, pp. 6–15. DOI: 10.1109/ISMAR.2004.44.
92. I. Coma-Tatay, S. Casas-Yrurzum, P. Casanova-Salas, and M. Fernández-Marín. "FI-AR learning: a web-based platform for augmented reality educational content". *Multimedia Tools and Applications* 78:5, 2019, pp. 6093–6118. DOI: 10.1007/s11042-018-6395-5.
93. D. Constancio, D. Casanova, and A. R. Fernandes. "MARS: Mobile augmented reality script framework". In: *2017 24º Encontro Português de Computação Gráfica e Interação (EPCGI)*. 2017, pp. 1–8. DOI: 10.1109/EPCGI.2017.8124302.
94. F. Corazza, D. Snijders, M. Arpone, V. Stritoni, F. Martinolli, M. Daverio, M. G. Losi, L. Soldi, F. Tesauri, L. Da Dalt, and S. Bressan. "Development and Usability of a Novel Interactive Tablet App (PediAppRREST) to Support the Management of Pediatric Cardiac Arrest: Pilot High-Fidelity Simulation-Based Study". *JMIR Mhealth Uhealth* 8:10, 2020, e19070. ISSN: 2291-5222. DOI: 10.2196/19070.
95. I. Cruz, B. Marques, J. Alves, P. Dias, and B. S. Santos. "Configuration and visualization of AR environments: a prototype". *ICGI 2019 - International Conference on Graphics and Interaction* 4:5, p. 6.
96. J. Cubillo, S. Martin, M. Castro, and I. Boticki. "Preparing augmented reality learning content should be easy: UNED ARLE—an authoring tool for augmented reality learning environments". *Computer Applications in Engineering Education* 23:5, 2015, pp. 778–789. DOI: 10.1002/cae.21650.
97. J. Cubillo, S. Martin, M. Castro, and I. Boticki. "Preparing augmented reality learning content should be easy: UNED ARLE—an authoring tool for augmented reality learning environments". *Computer Applications in Engineering Education* 23:5, 2015, pp. 778–789. DOI: 10.1002/cae.21650.
98. J. Cubillo, S. Martín, M. Castro, G. Diaz, A. Colmenar, and I. Botički. "A learning environment for augmented reality mobile learning". In: *2014 IEEE Frontiers in Education Conference (FIE) Proceedings*. 2014, pp. 1–8. DOI: 10.1109/FIE.2014.7044039.
99. P. Dähne. "Entwurf eines Rahmensystems für mobile Augmented-reality-Anwendungen". PhD thesis. Technische Universität Darmstadt, 2008.
100. C. S. C. Dalim, H. Kolivand, H. Kadhim, M. S. Sunar, and M. Billinghurst. "Factors influencing the acceptance of augmented reality in education: A review of the literature". *Journal of computer science* 13:11, 2017, pp. 581–589.
101. F. D. Davis. "Perceived usefulness, perceived ease of use, and user acceptance of information technology". *MIS quarterly*, 1989, pp. 319–340. DOI: 10.2307/249008.
102. L. F. de Souza Cardoso, F. C. M. Q. Mariano, and E. R. Zorzal. "A survey of industrial augmented reality". *Computers & Industrial Engineering* 139, 2020, p. 106159. ISSN: 0360-8352. DOI: 10.1016/j.cie.2019.106159.

103. A. M. Demiris, V. Vlahakis, A. Makri, M. Papaioannou, and N. Ioannidis. "intGuide: A platform for context-aware services featuring augmented-reality, based on the outcome of European Research Projects". *Signal Processing: Image Communication* 20:9, 2005, pp. 927–946. ISSN: 0923-5965. DOI: 10.1016/j.image.2005.05.007.

104. A. Dengel, M. Iqbal, S. Grafe, and E. Mangina. "A Review on Augmented Reality Authoring Toolkits for Education. Front". *Virtual Real. 3: 798032. doi: 10.3389/frvir*, 2022.

105. A. Dey, M. Billinghurst, R. W. Lindeman, and J. E. Swan. "A Systematic Review of 10 Years of Augmented Reality Usability Studies: 2005 to 2014". *Frontiers in Robotics and AI* 5, 2018. ISSN: 2296-9144. DOI: 10.3389/frobt.2018.00037.

106. A. Diamantopoulos, M. Sarstedt, C. Fuchs, P. Wilczynski, and S. Kaiser. "Guidelines for choosing between multi-item and single-item scales for construct measurement: a predictive validity perspective". *Journal of the Academy of Marketing Science* 40, 2012, pp. 434–449. DOI: 10.1007/s11747-011-0300-3.

107. J. Dias, L. Monteiro, P. Santos, R. Silvestre, and R. Bastos. "Developing and authoring mixed reality with MX toolkit". In: *2003 IEEE International Augmented Reality Toolkit Workshop*. 2003, pp. 18–26. DOI: 10.1109/ART.2003.1320420.

108. M. Dias, J. Jorge, J. Carvalho, P. Santos, and J. Luzio. "Usability evaluation of tangible user interfaces for augmented reality". In: *2003 IEEE International Augmented Reality Toolkit Workshop*. 2003, pp. 54–61. DOI: 10.1109/ART.2003.1320428.

109. W. Dick, L. Carey, and J. O. Carey. "The systematic design of instruction", 2005.

110. *Die Crossmedia-Reihe "Praxisorientiert ausbilden"*. http://www.kompetenzwerkstatt.net/lernbausteine. Accessed: 16.04.2023.

111. T. V. Do and J. W. Lee. "Creating 3D E-Books with ARBookCreator". In: *Proceedings of the International Conference on Advances in Computer Enterntainment Technology*. ACE '09. Association for Computing Machinery, Athens, Greece, 2009, pp. 429–430. ISBN: 9781605588643. DOI: 10.1145/1690388.1690484.

112. A. M. Dominguez, T. Robles, R. Alcarria, and E. Cedeño. "Mixed Reality-Aware Service Architecture for Mobile Environments". In: *2012 Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*. 2012, pp. 817–822. DOI: 10.1109/IMIS.2012.124.

113. J. L. Domínguez Alfaro. "XR Laboratory Prototype for Chemical Engineering Education". Forthcoming. PhD thesis. Faculty of Engineering Science, Katholieke Universiteit Leuven, Belgium, 2023.

114. J. L. Domínguez Alfaro, S. Gantois, J. Blattgerste, R. De Croon, K. Verbert, T. Pfeiffer, and P. Van Puyvelde. "Mobile Augmented Reality Laboratory for Learning Acid–Base Titration". *Journal of Chemical Education* 99:2, 2022, pp. 531–537. DOI: 10.1021/acs.jchemed.1c00894.

115. V. Driessen. *A successful Git branching model*. https://nvie.com/posts/a-successful-git-branching-model/ (Accessed: 2023-05-25). 2010.

116. N. Drljević, L. H. Wong, and I. Botički. "Where Does My Augmented Reality Learning Experience (ARLE) Belong? A Student and Teacher Perspective to Positioning ARLEs". *IEEE Transactions on Learning Technologies* 10:4, 2017, pp. 419–435. DOI: 10.1109/TLT.2017.2690426.

117. J. J. Dudley, H. Schuff, and P. O. Kristensson. "Bare-Handed 3D Drawing in Augmented Reality". In: *Proceedings of the 2018 Designing Interactive Systems Conference*. DIS '18. Association for Computing Machinery, Hong Kong, China, 2018, pp. 241–252. ISBN: 9781450351980. DOI: 10.1145/3196709.3196737.

118. J. S. Dumas and J. Redish. *A practical guide to usability testing*. Intellect books, 1999.

119. A. Dünser, L. Walker, H. Horner, and D. Bentall. "Creating Interactive Physics Education Books with Augmented Reality". In: *Proceedings of the 24th Australian Computer-Human Interaction Conference*. OzCHI '12. Association for Computing Machinery, Melbourne, Australia, 2012, pp. 107–114. ISBN: 9781450314381. DOI: 10.1145/2414536.2414554.

120. D. H. e.V. *Zahlenspiegel zur Situation der Hebammen*. Technical report. https://www.unsere-hebammen.de/w/files/kampagnenmaterial/202111-zahlenspiegel-zur-situation-der-hebammen.pdf. Accessed: 31.05.2023. Deutscher Hebammenverband e.V., 2021.

121. D. Eckhoff, C. Sandor, C. Lins, U. Eck, D. Kalkofen, and A. Hein. "TutAR: Augmented Reality Tutorials for Hands-Only Procedures". In: *Proceedings of the 16th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*. VRCAI '18. Association for Computing Machinery, Tokyo, Japan, 2018. ISBN: 9781450360876. DOI: 10.1145/3284398.3284399.

122. M. Edlund. *UnityMeshSimplifier*. https://github.com/Whinarn/UnityMeshSimplifier. Accessed: 02.05.2023, GitHub repository. 2017.

123. J. Eggeling. "Evaluation der Entwicklung von TrainAR für Richtfunk im Offshore Bereich". Masters's Thesis. University of Applied Sciences Emden/Leer, 2023.

124. H. Endo, S. Furuya, and K.-I. Okada. "MR Manual and Authoring Tool with Afterimages". In: *2015 IEEE 29th International Conference on Advanced Information Networking and Applications*. 2015, pp. 890–895. DOI: 10.1109/AINA.2015.284.

125. T. Engelke, J. Keil, P. Rojtberg, F. Wientapper, M. Schmitt, and U. Bockholt. "Content First: A Concept for Industrial Augmented Reality Maintenance Applications Using Mobile Devices". In: *Proceedings of the 6th ACM Multimedia Systems Conference*. MMSys '15. Association for Computing Machinery, Portland, Oregon, 2015, pp. 105–111. ISBN: 9781450333511. DOI: 10.1145/2713168.2713169.

126. T. Engelke, S. Webel, U. Bockholt, H. Wuest, N. Gavish, F. Tecchia, and C. Preusche. "Towards automatic generation of multimodal AR-training applications and workflow descriptions". In: *19th International Symposium in Robot and Human Interactive Communication*. 2010, pp. 434–439. DOI: 10.1109/ROMAN.2010.5598613.

127. J. A. Erkoyuncu, I. F. del Amo, M. Dalle Mura, R. Roy, and G. Dini. "Improving efficiency of industrial maintenance with context aware adaptive authoring in augmented reality". *CIRP Annals* 66:1, 2017, pp. 465–468. ISSN: 0007-8506. DOI: 10.1016/j.cirp.2017.04.006.

128. L. Escobedo, M. Tentori, E. Quintana, J. Favela, and D. Garcia-Rosas. "Using Augmented Reality to Help Children with Autism Stay Focused". *IEEE Pervasive Computing* 13:1, 2014, pp. 38–46. DOI: 10.1109/MPRV.2014.19.

129. A. Evangelista, L. Ardito, A. Boccaccio, M. Fiorentino, A. Messeni Petruzzelli, and A. E. Uva. "Unveiling the technological trends of augmented reality: A patent analysis". *Computers in Industry* 118, 2020, p. 103221. ISSN: 0166-3615. DOI: 10.1016/j.compind.2020.103221.

130. F. E. Fadzli, M. A. M. Yusof, A. W. Ismail, M. S. H. Salam, and N. A. Ismail. "ARGarden: 3D outdoor landscape design using handheld augmented reality with multi-user interaction". *IOP Conference Series: Materials Science and Engineering* 979:1, 2020, p. 012001. DOI: 10.1088/1757-899X/979/1/012001.

131. M. E. Falagas, E. I. Pitsouni, G. A. Malietzis, and G. Pappas. "Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses". *The FASEB Journal* 22:2, 2008, pp. 338–342. DOI: 10.1096/fj.07-9492LSF.

132. L. Farias, R. Dantas, and A. Burlamaqui. "Educ-AR: A tool for assist the creation of augmented reality content for education". In: *2011 IEEE International Conference on Virtual Environments, Human-Computer Interfaces and Measurement Systems Proceedings*. 2011, pp. 1–5. DOI: 10.1109/VECIMS.2011.6053850.

133. D. Fehling, L. Goertz, and T. Hagenhofer. *Didaktisches Konzept des Projektes Social Augmented Learning*. https://www.social-augmented-learning.de/wp-content/uploads/2015/04/SAL_Didaktisches_Konzept_20150409.pdf. Accessed: 28.05.2023.

134. G. Fei, X. Li, and R. Fei. "AVATAR: Autonomous Visual Authoring of Tangible Augmented Reality". In: *Proceedings of the 11th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*. VRCAI '12. Association for Computing Machinery, Singapore, Singapore, 2012, pp. 21–24. ISBN: 9781450318259. DOI: 10.1145/2407516.2407521.

135. K. R. Felizardo, E. Mendes, M. Kalinowski, É. F. Souza, and N. L. Vijaykumar. "Using Forward Snowballing to Update Systematic Reviews in Software Engineering". In: *Proceedings of the 10th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*. ESEM '16. Association for Computing Machinery, Ciudad Real, Spain, 2016. ISBN: 9781450344272. DOI: 10.1145/2961111.2962630.

136. C. Fenu and F. Pittarello. "Svevo tour: The design and the experimentation of an augmented reality application for engaging visitors of a literary museum". *International Journal of Human-Computer Studies* 114, 2018, pp. 20–35. ISSN: 1071-5819. DOI: 10.1016/j.ijhcs.2018.01.009.

137. I. Fernández del Amo, J. Erkoyuncu, R. Vrabič, R. Frayssinet, C. Vazquez Reynel, and R. Roy. "Structured authoring for AR-based communication to enhance efficiency in remote diagnosis for complex equipment". *Advanced Engineering Informatics* 45, 2020, p. 101096. ISSN: 1474-0346. DOI: 10.1016/j.aei.2020.101096.

138. I. Fernández del Amo, J. A. Erkoyuncu, R. Roy, R. Palmarini, and D. Onoufriou. "A systematic review of Augmented Reality content-related techniques for knowledge transfer in maintenance applications". *Computers in Industry* 103, 2018, pp. 47–71. ISSN: 0166-3615. DOI: 10.1016/j.compind.2018.08.007.

139. S. Feuerstack, Á. C. M. de Oliveira, M. dos Santos Anjo, R. B. Araujo, and E. B. Pizzolato. "Model-Based Design of Multimodal Interaction for Augmented Reality Web Applications". In: *Proceedings of the 20th International Conference on 3D Web Technology*. Web3D '15. Association for Computing Machinery, Heraklion, Crete, Greece, 2015, pp. 259–267. ISBN: 9781450336475. DOI: 10.1145/2775292.2775293.

140. P. Figueroa, R. Dachselt, and I. Lindt. "A conceptual model and specification language for mixed reality interface components". In: *VR 2006, In Proc. of the Workshop "Specification of Mixed Reality User Interfaces: Approaches, Languages, Standardization*. 2006, pp. 4–11.

141. S. S. Fisher. "An Authoring Toolkit for Mixed Reality Experiences". In: *Entertainment Computing: Technologies and Application*. Ed. by R. Nakatsu and J. Hoshino. Springer US, Boston, MA, 2003, pp. 487–494. ISBN: 978-0-387-35660-0. DOI: 10.1007/978-0-387-35660-0_59.

142. J. Franssen. "Gestaltung und Implementierung eines Technischen Ansatzes zur Erfassung von Benutzungsdaten in Augmented Reality Trainings". Bachelors's Thesis. University of Applied Sciences Emden/Leer, 2022.

143. G. Freitas, M. S. Pinho, M. S. Silveira, and F. Maurer. "A Systematic Review of Rapid Prototyping Tools for Augmented Reality". In: *2020 22nd Symposium on Virtual and Augmented Reality (SVR)*. 2020, pp. 199–209. DOI: 10.1109/SVR51698.2020.00041.

144. M. Funk, L. Lischke, S. Mayer, A. S. Shirazi, and A. Schmidt. "Teach Me How! Interactive Assembly Instructions Using Demonstration and In-Situ Projection". In: *Assistive Augmentation*. Ed. by J. Huber, R. Shilkrot, P. Maes, and S. Nanayakkara. Springer Singapore, Singapore, 2018, pp. 49–73. ISBN: 978-981-10-6404-3. DOI: 10.1007/978-981-10-6404-3_4.

145. R. M. Gagne. "Learning outcomes and their effects: Useful categories of human performance." *American psychologist* 39:4, 1984, p. 377. DOI: 10.1037/0003-066x.39.4.377.

146. R. M. Gagne and L. J. Briggs. *Principles of instructional design.* Holt, Rinehart & Winston, 1974.

147. Z. Ganlin, F. Pingfa, Z. Jianfu, Y. Dingwen, and W. Zhijun. "Information integration and instruction authoring of augmented assembly systems". *International Journal of Intelligent Systems* 36:9, 2021, pp. 5028–5050. DOI: 10.1002/int.22501.

148. K. Gao, H. L. Xu, D. M. Zhang, Z. Ouyang, and C. Y. Wei. "A Framework for Virtual Assembly Based on AR Environments". In: *Applied Science, Materials Science and Information Technologies in Industry*. Vol. 513. Applied Mechanics and Materials. Trans Tech Publications Ltd, 2014, pp. 854–857. DOI: 10.4028/www.scientific.net/AMM.513-517.854.

149. M. Gao, P. Kortum, and F. L. Oswald. "Multi-Language Toolkit for the System Usability Scale". *International Journal of Human–Computer Interaction* 36:20, 2020, pp. 1883–1901. DOI: 10.1080/10447318.2020.1801173.

150. I. García-Pereira, J. Gimeno, P. Morillo, and P. Casanova-Salas. "A Taxonomy of Augmented Reality Annotations." In: *VISIGRAPP (1: GRAPP)*. 2020, pp. 412–419.

151. I. García-Pereira, C. Portalés, J. Gimeno, and S. Casas. "A collaborative augmented reality annotation tool for the inspection of prefabricated buildings". *Multimedia Tools and Applications* 79:9, 2020, pp. 6483–6501. DOI: 10.1007/s11042-019-08419-x.

152. G. Gartlehner, L. Affengruber, V. Titscher, A. Noel-Storr, G. Dooley, N. Ballarini, and F. König. "Single-reviewer abstract screening missed 13 percent of relevant studies: a crowd-based, randomized controlled trial". *Journal of Clinical Epidemiology* 121, 2020, pp. 20–28. ISSN: 0895-4356. DOI: 10.1016/j.jclinepi.2020.01.005.

153. J. Garzón, J. Pavón, and S. Baldiris. "Systematic review and meta-analysis of augmented reality in educational settings". *Virtual Reality* 23:4, 2019, pp. 447–459. DOI: 10.1007/s10055-019-00379-9.

154. D. Gasques, J. G. Johnson, T. Sharkey, and N. Weibel. "What You Sketch Is What You Get: Quick and Easy Augmented Reality Prototyping with PintAR". In: *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI EA '19. Association for Computing Machinery, Glasgow, Scotland Uk, 2019, pp. 1–6. ISBN: 9781450359719. DOI: 10.1145/3290607.3312847.

155. A. Gates, S. Guitard, J. Pillay, S. A. Elliott, M. P. Dyson, A. S. Newton, and L. Hartling. "Performance and usability of machine learning for screening in systematic reviews: a comparative evaluation of three tools". *Systematic reviews* 8:1, 2019, pp. 1–11. DOI: 10.1186/s13643-019-1222-2.

156. M. Gattullo, A. Evangelista, A. E. Uva, M. Fiorentino, and J. L. Gabbard. "What, How, and Why are Visual Assets Used in Industrial Augmented Reality? A Systematic Review and Classification in Maintenance, Assembly, and Training (From 1997 to 2019)". *IEEE Transactions on Visualization and Computer Graphics* 28:2, 2022, pp. 1443–1456. DOI: 10.1109/TVCG.2020.3014614.

157. J. Gausemeier, J. Fruend, and C. Matysczok. "AR-Planning Tool: Designing Flexible Manufacturing Systems with Augmented Reality". In: *Proceedings of the Workshop on Virtual Environments 2002*. EGVE '02. Eurographics Association, Barcelona, Spain, 2002, pp. 19–25. ISBN: 1581135351.

158. G. Gauthier, C. St-Onge, and W. Tavares. "Rater cognition: review and integration of research findings". *Medical Education* 50:5, 2016, pp. 511–522. DOI: 10.1111/medu.12973.

159. N. Gazcón and S. Castro. "ARBS: An Interactive and Collaborative System for Augmented Reality Books". In: *Augmented and Virtual Reality*. Ed. by L. T. De Paolis and A. Mongelli. Springer International Publishing, Cham, 2015, pp. 89–108. ISBN: 978-3-319-22888-4. DOI: 10.1007/978-3-319-22888-4_8.

160. A. P. Gee, A. Calway, and W. Mayol-Cuevas. "Visual Mapping and Multi-modal Localisation for Anywhere AR Authoring". In: *Computer Vision – ACCV 2010 Workshops*. Ed. by R. Koch and F. Huang. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 31–41. ISBN: 978-3-642-22819-3. DOI: 10.1007/978-3-642-22819-3_4.

161. C. Geiger, V. Paelke, C. Reimann, and W. Rosenbach. "A Framework for the Structured Design of VR/AR Content". In: *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*. VRST '00. Association for Computing Machinery, Seoul, Korea, 2000, pp. 75–82. ISBN: 1581133162. DOI: 10.1145/502390.502405.

162. C. Geiger, C. Reimann, W. Rosenbach, and J. Stöcklein. "Rapid Prototyping of Mixed Reality Applications That Entertain and Inform". In: *Entertainment Computing: Technologies and Application*. Ed. by R. Nakatsu and J. Hoshino. Springer US, Boston, MA, 2003, pp. 479–486. ISBN: 978-0-387-35660-0. DOI: 10.1007/978-0-387-35660-0_58.

163. J. Geng, X. Song, Y. Pan, J. Tang, Y. Liu, D. Zhao, and Y. Ma. "A systematic design method of adaptive augmented reality work instruction for complex industrial operations". *Computers in Industry* 119, 2020, p. 103229. ISSN: 0166-3615. DOI: 10.1016/j.compind.2020.103229.

164. E. Geronikolakis, P. Zikas, S. Kateros, N. Lydatakis, S. Georgiou, M. Kentros, and G. Papagiannakis. "A True AR Authoring Tool for Interactive Virtual Museums". In: *Visual Computing for Cultural Heritage*. Ed. by F. Liarokapis, A. Voulodimos, N. Doulamis, and A. Doulamis. Springer International Publishing, Cham, 2020, pp. 225–242. ISBN: 978-3-030-37191-3. DOI: 10.1007/978-3-030-37191-3_12.

165. *Gesetz über das Studium und den Beruf von Hebammen (Hebammengesetz - HebG) (2020) in der Fassung des Gesetzes zur Reform der Hebammenausbildung und zur Änderung des Fünften Buches Sozialgesetzbuch (Hebammenreformgesetzes- HebRefG) vom 22.11.2019 (BGBl. I S. 1759)*. 2020.

166. J. Gimeno, P. Morillo, J. M. Orduña, and M. Fernández. "An Easy-to-Use AR Authoring Tool for Industrial Applications". In: *Computer Vision, Imaging and Computer Graphics. Theory and Application*. Ed. by G. Csurka, M. Kraus, R. S. Laramee, P. Richard, and J. Braz. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 17–32. ISBN: 978-3-642-38241-3. DOI: 10.1007/978-3-642-38241-3_2.

167. D. Giustini and M. N. Kamel Boulos. "Google Scholar is not enough to be used alone for systematic reviews". *Online journal of public health informatics* 5:2, 2013. ISSN: 1947-2579. DOI: 10.5210/ojphi.v5i2.4623.

168. K. Gökhan and İ. Gökhan. "ARgent: A web based augmented reality framework for dynamic content generation". *Avrupa Bilim ve Teknoloji Dergisi*, 2020, pp. 244–257. DOI: 10.31590/ejosat.779946.

169. M. S. Gonzalez, C. L. Gómez, and D. Computing. "ARUIML: A Three-dimensional User Interface Description Language for Augmented Reality". Note: Full bibliographic details could not be retrieved (29.05.2023).

170. Google. *ARCore*. https://developers.google.com/ar. Accessed: 08.04.2023. 2017.

171. S. Gottschalk, E. Yigitbas, E. Schmidt, and G. Engels. "Model-Based Product Configuration in Augmented Reality Applications". In: *Human-Centered Software Engineering*. Ed. by R. Bernhaupt, C. Ardito, and S. Sauer. Springer International Publishing, Cham, 2020, pp. 84–104. ISBN: 978-3-030-64266-2. DOI: 10.1007/978-3-030-64266-2_5.

172. D. Greyson, E. Rafferty, L. Slater, N. MacDonald, J. A. Bettinger, È. Dubé, and S. E. MacDonald. "Systematic review searches must be systematic, comprehensive, and transparent: a critique of Perman et al". *BMC Public Health* 19:1, 2019, pp. 1–6. DOI: 10.1186/s12889-018-6275-y.

173. P. Grimm, M. Haller, V. Paelke, S. Reinhold, C. Reimann, and R. Zauner. "AMIRE-authoring mixed reality". In: *The First IEEE International Workshop Agumented Reality Toolkit,* IEEE. 2002, 2–pp.

174. M. Groenendyk. "Cataloging the 3D web: The availability of educational 3D models on the internet". *Library Hi Tech*, 2016. DOI: 10.1108/LHT-09-2015-0088.

175. A. Guo, I. Canberk, H. Murphy, A. Monroy-Hernández, and R. Vaish. "Blocks: Collaborative and Persistent Augmented Reality Experiences". *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3:3, 2019. DOI: 10.1145/3351241.
176. M. Gusenbauer. "Google Scholar to overshadow them all? Comparing the sizes of 12 academic search engines and bibliographic databases". *Scientometrics* 118:1, 2019, pp. 177–214. DOI: 10.1007/s11192-018-2958-5.
177. M. Gusenbauer and N. R. Haddaway. "Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources". *Research Synthesis Methods* 11:2, 2020, pp. 181–217. DOI: 10.1002/jrsm.1378.
178. S. Guven and S. Feiner. "A hypermedia authoring tool for augmented and virtual reality". *New Review of Hypermedia and Multimedia* 9:1, 2003, pp. 89–116. DOI: 10.1080/13614560410001725329.
179. S. Guven, S. Feiner, and O. Oda. "Mobile augmented reality interaction techniques for authoring situated media on-site". In: *2006 IEEE/ACM International Symposium on Mixed and Augmented Reality*. 2006, pp. 235–236. DOI: 10.1109/ISMAR.2006.297821.
180. T. Ha and W. Woo. "Graphical tangible user interface for a ar authoring tool in product design environment". In: *International Symposium on Ubiquitous VR*. Vol. 260. 2007.
181. T. Ha, W. Woo, Y. Lee, J. Lee, J. Ryu, H. Choi, and K. Lee. "ARtalet: Tangible User Interface Based Immersive Augmented Reality Authoring Tool for Digilog Book". In: *2010 International Symposium on Ubiquitous Virtual Reality*. 2010, pp. 40–43. DOI: 10.1109/ISUVR.2010.20.
182. N. Hagbi, R. Grasset, O. Bergig, M. Billinghurst, and J. El-Sana. "In-Place Sketching for Augmented Reality Games". *Comput. Entertain.* 12:3, 2015. DOI: 10.1145/2702109.2633419.
183. M. Haller, E. Stauder, and J. Zauner. "AMIRE-ES: Authoring Mixed Reality once, run it anywhere". In: *Proceedings of the 11th International Conference on Human-Computer Interaction (HCII)*. Vol. 2005. 2005.
184. A. Hampshire, H. Seichter, R. Grasset, and M. Billinghurst. "Augmented Reality Authoring: Generic Context from Programmer to Designer". In: *Proceedings of the 18th Australia Conference on Computer-Human Interaction: Design: Activities, Artefacts and Environments*. OZCHI '06. Association for Computing Machinery, Sydney, Australia, 2006, pp. 409–412. ISBN: 1595935452. DOI: 10.1145/1228175.1228259.
185. R. M. Harden. "What is an OSCE?" *Medical Teacher* 10:1, 1988, pp. 19–22. DOI: 10.3109/01421598809019321.

186. J. Hardy and J. Alexander. "Toolkit Support for Interactive Projected Displays". In: *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia*. MUM '12. Association for Computing Machinery, Ulm, Germany, 2012. ISBN: 9781450318150. DOI: 10.1145/2406367.2406419.

187. M. Haringer and H. Regenbrecht. "A pragmatic approach to augmented reality authoring". In: *Proceedings. International Symposium on Mixed and Augmented Reality*. 2002, pp. 237–245. DOI: 10.1109/ISMAR.2002.1115093.

188. S. G. Hart and L. E. Staveland. "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research". In: *Advances in psychology*. Vol. 52. Elsevier, 1988, pp. 139–183. DOI: 10.1016/S0166-4115(08)62386-9.

189. P. Haynes, S. Hehl-Lange, and E. Lange. "Mobile Augmented Reality for Flood Visualisation". *Environmental Modelling & Software* 109, 2018, pp. 380–389. ISSN: 1364-8152. DOI: 10.1016/j.envsoft.2018.05.012.

190. F. He, S. Ong, and A. Nee. "A Mobile Solution for Augmenting a Manufacturing Environment with User-Generated Annotations". *Information* 10:2, 2019. ISSN: 2078-2489. DOI: 10.3390/info10020060.

191. R. A. Heinlein. *Starship Troopers*. G. P. Putnam's Sons, United States, 1959. ISBN: 978-0450044496.

192. A. Henrysson, M. Ollila, and M. Billinghurst. "Mobile Phone Based AR Scene Assembly". In: *Proceedings of the 4th International Conference on Mobile and Ubiquitous Multimedia*. MUM '05. Association for Computing Machinery, Christchurch, New Zealand, 2005, pp. 95–102. ISBN: 0473106582. DOI: 10.1145/1149488.1149504.

193. D. Herr, J. Reinhardt, G. Reina, R. Krüger, R. V. Ferrari, and T. Ertl. "Immersive Modular Factory Layout Planning using Augmented Reality". *Procedia CIRP* 72, 2018, pp. 1112–1117. ISSN: 2212-8271. DOI: 10.1016/j.procir.2018.03.200.

194. J. Hertel, S. Karaosmanoglu, S. Schmidt, J. Bräker, M. Semmann, and F. Steinicke. "A Taxonomy of Interaction Techniques for Immersive Augmented Reality based on an Iterative Literature Review". In: *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 2021, pp. 431–440. DOI: 10.1109/ISMAR52148.2021.00060.

195. A. Hill, B. MacIntyre, M. Gandy, B. Davidson, and H. Rouzati. "KHARMA: An open KML/HTML architecture for mobile augmented reality applications". In: *2010 IEEE International Symposium on Mixed and Augmented Reality*. 2010, pp. 233–234. DOI: 10.1109/ISMAR.2010.5643583.

196. Z. Hodaie, J. Haladjian, and B. Brügge. "ISAR: An Authoring System for Interactive Tabletops". In: *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces*. ISS '19. Association for Computing Machinery, Daejeon, Republic of Korea, 2019, pp. 355–360. ISBN: 9781450368919. DOI: 10.1145/3343055.3360751.

197. R. J. Holden. "A Simplified System Usability Scale (SUS) for Cognitively Impaired and Older Adults". *Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care* 9:1, 2020, pp. 180–182. DOI: 10.1177/2327857920091021.

198. J. S. Hong and J. W. Lee. "Picture Book-Based Augmented Reality Content Authoring System". In: *HCI International 2018 – Posters' Extended Abstracts*. Ed. by C. Stephanidis. Springer International Publishing, Cham, 2018, pp. 253–257. ISBN: 978-3-319-92279-9. DOI: 10.1007/978-3-319-92279-9_34.

199. H. Huang, Y. Zhang, T. Weiss, R. W. Perry, and L.-F. Yu. "Interactive Design of Gallery Walls via Mixed Reality". In: *2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. 2020, pp. 17–26. DOI: 10.1109/AIVR50618.2020.00013.

200. K. Huo, Vinayak, and K. Ramani. "Window-Shaping: 3D Design Ideation by Creating on, Borrowing from, and Looking at the Physical World". In: *Proceedings of the Eleventh International Conference on Tangible, Embedded, and Embodied Interaction*. TEI '17. Association for Computing Machinery, Yokohama, Japan, 2017, pp. 37–45. ISBN: 9781450346764. DOI: 10.1145/3024969.3024995.

201. R. Ichikari, K. Kawano, A. Kimura, F. Shibata, and H. Tamura. "Mixed reality previsualization and camera-work authoring in filmmaking". In: *2006 IEEE/ACM International Symposium on Mixed and Augmented Reality*. 2006, pp. 239–240. DOI: 10.1109/ISMAR.2006.297823.

202. R. Ichikari, R. Tenmoku, F. Shibata, T. Ohshima, and H. Tamura. "MR-based PreViz systems for filmmaking: On-set camera-work authoring and action rehearsal". In: *Proc. 6th ISMAR Workshop Mixed Reality Entertain. Art*. 2007, pp. 21–26.

203. A. Ioannidi, D. Gavalas, and V. Kasapakis. "Flaneur: Augmented exploration of the architectural urbanscape". In: *2017 IEEE Symposium on Computers and Communications (ISCC)*. 2017, pp. 529–533. DOI: 10.1109/ISCC.2017.8024582.

204. S. Irawati, S. Green, M. Billinghurst, A. Duenser, and H. Ko. ""Move the couch where?" : developing an augmented reality multimodal interface". In: *2006 IEEE/ACM International Symposium on Mixed and Augmented Reality*. 2006, pp. 183–186. DOI: 10.1109/ISMAR.2006.297812.

205. A. Irlitti, R. T. Smith, S. Von Itzstein, M. Billinghurst, and B. H. Thomas. "Challenges for Asynchronous Collaboration in Augmented Reality". In: *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*. 2016, pp. 31–35. DOI: 10.1109/ISMAR-Adjunct.2016.0032.

206. A. Jain, B. Mildenhall, J. T. Barron, P. Abbeel, and B. Poole. "Zero-Shot Text-Guided Object Generation with Dream Fields". In: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2022, pp. 857–866. DOI: 10.1109/CVPR52688.2022.00094.

207. S. Jalali and C. Wohlin. "Systematic Literature Studies: Database Searches vs. Backward Snowballing". In: *Proceedings of the ACM-IEEE International Symposium on Empirical Software Engineering and Measurement*. ESEM '12. Association for Computing Machinery, Lund, Sweden, 2012, pp. 29–38. ISBN: 9781450310567. DOI: 10.1145/2372251.2372257.

208. J. Jang, Y. Ko, W. S. Shin, and I. Han. "Augmented Reality and Virtual Reality for Learning: An Examination Using an Extended Technology Acceptance Model". *IEEE Access* 9, 2021, pp. 6798–6809. DOI: 10.1109/ACCESS.2020.3048708.

209. S. Janßen. "Implementierung und Evaluation gestenbasierter Interaktionskonzepte für prozedurale Abläufe auf headmounted AR-Displays". Bachelors's Thesis. University of Applied Sciences Emden/Leer, 2020.

210. S. Janßen. "Skalierbare Distribution von Augmented Reality Training anhand des Authoring Tools TrainAR". Masters's Thesis. University of Applied Sciences Emden/Leer, 2023.

211. H.-K. Jee, S. Lim, J. Youn, and J. Lee. "An Immersive Authoring Tool for Augmented Reality-Based E-Learning Applications". In: *2011 International Conference on Information Science and Applications*. 2011, pp. 1–5. DOI: 10.1109/ICISA.2011.5772399.

212. J. Jeon, M. Hong, M. Yi, J. Chun, J. S. Kim, and Y.-J. Choi. "Interactive authoring tool for mobile augmented reality content". *Journal of Information Processing Systems* 12:4, 2016, pp. 612–630. DOI: 10.3745/jips.02.0048.

213. S. Jiang, S. Ong, and A. Nee. "An AR-based hybrid approach for facility layout planning and evaluation for existing shop floors". *The International Journal of Advanced Manufacturing Technology* 72:1, 2014, pp. 457–473. DOI: 10.1007/s00170-014-5653-6.

214. Q. Jin, D. Wang, and F. Sun. "TanCreator: A Tangible Tool for Children to Create Augmented Reality Games". In: *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*. UbiComp '18. Association for Computing Machinery, Singapore, Singapore, 2018, pp. 82–85. ISBN: 9781450359665. DOI: 10.1145/3267305.3267603.

215. D. Jo and G. J. Kim. "ARIoT: scalable augmented reality framework for interacting with Internet of Things appliances everywhere". *IEEE Transactions on Consumer Electronics* 62:3, 2016, pp. 334–340. DOI: 10.1109/TCE.2016.7613201.

216. D. H. Jonassen, W. H. Hannum, and M. Tessmer. *Handbook of task analysis procedures.* Praeger Publishers, 1989.

217. M. Joswig, C. Lewa, and T. Schäfer. *Forschungsprojekt Heb@AR, Teilprojekt Medizindidaktik*. Project partner during the BMBF (Federal Ministry of Education and Research, Germany) research project Heb@AR (16DHB3020).

218. B. Jung, M. Lenk, and A. Vitzthum. "Structured development of 3D applications: round-trip engineering in interdisciplinary teams". *Computer Science-Research and Development* 30:3, 2015, pp. 285–301. DOI: 10.1007/s00450-014-0258-8.

219. J. Jung, J. Hong, S. Park, and H. S. Yang. "Smartphone as an Augmented Reality Authoring Tool via Multi-Touch Based 3D Interaction Method". In: *Proceedings of the 11th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*. VRCAI '12. Association for Computing Machinery, Singapore, Singapore, 2012, pp. 17–20. ISBN: 9781450318259. DOI: 10.1145/2407516.2407520.

220. K. Jung, S. Lee, S. Jeong, and B.-U. Choi. "Virtual Tactical Map with Tangible Augmented Reality Interface". In: *2008 International Conference on Computer Science and Software Engineering*. Vol. 2. 2008, pp. 1170–1173. DOI: 10.1109/CSSE.2008.1305.

221. Y. Jung, T. Franke, P. Dähne, and J. Behr. "Enhancing X3D for Advanced MR Appliances". In: *Proceedings of the Twelfth International Conference on 3D Web Technology*. Web3D '07. Association for Computing Machinery, Perugia, Italy, 2007, pp. 27–36. ISBN: 9781595936523. DOI: 10.1145/1229390.1229394.

222. A. Kampa and U. Spierling. "Smart Authoring for location-based augmented reality storytelling applications". *INFORMATIK 2017*, 2017, pp. 915–922. DOI: 10.18420/in2017_93.

223. S. Kasahara, V. Heun, A. S. Lee, and H. Ishii. "Second Surface: Multi-User Spatial Collaboration System Based on Augmented Reality". In: *SIGGRAPH Asia 2012 Emerging Technologies*. SA '12. Association for Computing Machinery, Singapore, Singapore, 2012, pp. 1–4. ISBN: 9781450319126. DOI: 10.1145/2407707.2407727.

224. H. Kato and M. Billinghurst. "Marker tracking and HMD calibration for a video-based augmented reality conferencing system". In: *Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99)*. 1999, pp. 85–94. DOI: 10.1109/IWAR.1999.803809.

225. A. Kelly, R. B. Shapiro, J. de Halleux, and T. Ball. "ARcadia: A Rapid Prototyping Platform for Real-Time Tangible Interfaces". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18. Association for Computing Machinery, Montreal QC, Canada, 2018, pp. 1–8. ISBN: 9781450356206. DOI: 10.1145/3173574.3173983.

226. M. Kerres. "Mediendidaktik". In: *Mediendidaktik*. Oldenbourg Wissenschaftsverlag, 2013.

227. M. Kerres, M. Bormann, and M. Vervenne. "Didaktische konzeption von serious games: Zur verknüpfung von spiel-und lernangeboten". *MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung*, 2009, pp. 1–16. ISSN: 1424-3636.

228. K. Khowaja, B. Banire, D. Al-Thani, M. T. Sqalli, A. Aqle, A. Shah, and S. S. Salim. "Augmented Reality for Learning of Children and Adolescents With Autism Spectrum Disorder (ASD): A Systematic Review". *IEEE Access* 8, 2020, pp. 78779–78807. DOI: 10.1109/ACCESS.2020.2986608.

229. N. Kiesler and D. Schiffner. "On the lack of recognition of software artifacts and it infrastructure in educational technology research". *20. Fachtagung Bildungstechnologien (DELFI)*, 2022. DOI: 10.18420/delfi2022-034.

230. E. Kim, J. Kim, and W. Woo. "Metadata schema for context-aware augmented reality applications in cultural heritage domain". In: *2015 Digital Heritage*. Vol. 2. 2015, pp. 283–290. DOI: 10.1109/DigitalHeritage.2015.7419509.

231. H. Kim, G. Reitmayr, and W. Woo. "IMAF: in situ indoor modeling and annotation framework on mobile phones". *Personal and ubiquitous computing* 17:3, 2013, pp. 571–582. DOI: 10.1007/s00779-012-0516-3.

232. H. Kim, H. Yoon, A. Choi, W. Baek, I. Lee, D. Kim, and W. Woo. "Data markup representation for mixed reality contents". In: *International AR Standards Meeting-February*. 2011, pp. 17–19.

233. K. Kim, M. Billinghurst, G. Bruder, H. B.-L. Duh, and G. F. Welch. "Revisiting trends in augmented reality research: A review of the 2nd decade of ISMAR (2008–2017)". *IEEE transactions on visualization and computer graphics* 24:11, 2018, pp. 2947–2962.

234. K. Kim, J. Park, and W. Woo. "Marker-Less Tracking for Multi-Layer Authoring in AR Books". In: *Proceedings of the 8th International Conference on Entertainment Computing*. ICEC '09. Springer-Verlag, Paris, France, 2009, pp. 48–59. ISBN: 9783642040511. DOI: 10.1007/978-3-642-04052-8_5.

235. C. Kirner and R. Santin. "Interaction, collaboration and authoring in augmented reality environments". *XI Symposium on Virtual and Augmented Reality 16*, 2009.

236. P. Knierim, F. Kiss, M. Rauh, and A. Schmidt. "Tangibility is Overrated: Comparing Learning Experiences of Physical Setups and Their Virtual Equivalent in Augmented Reality". In: *19th International Conference on Mobile and Ubiquitous Multimedia*. MUM '20. Association for Computing Machinery, Essen, Germany, 2020, pp. 299–305. ISBN: 9781450388702. DOI: 10.1145/3428361.3428379.

237. C. Knöpfle, J. Weidenhausen, L. Chauvigné, and I. Stock. "Template Based Authoring for Augmented Reality based Service Scenarios". DOI: 10.1109/vr.2005.75.

238. S. M. Ko, W. S. Chang, and Y. G. Ji. "Usability Principles for Augmented Reality Applications in a Smartphone Environment". *International Journal of Human–Computer Interaction* 29:8, 2013, pp. 501–515. DOI: 10.1080/10447318.2012.722466.

239. B. Koleva, S. R. Egglestone, H. Schnädelbach, K. Glover, C. Greenhalgh, T. Rodden, and M. Dade-Robertson. "Supporting the Creation of Hybrid Museum Experiences". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '09. Association for Computing Machinery, Boston, MA, USA, 2009, pp. 1973–1982. ISBN: 9781605582467. DOI: 10.1145/1518701.1519001.

240. P. T. Kortum and A. Bangor. "Usability Ratings for Everyday Products Measured With the System Usability Scale". *International Journal of Human–Computer Interaction* 29:2, 2013, pp. 67–76. DOI: 10.1080/10447318.2012.681221.

241. K. I. Kotsopoulos, P. Chourdaki, D. Tsolis, R. Antoniadis, G. Pavlidis, and N. Assimakopoulos. "An authoring platform for developing smart apps which elevate cultural heritage experiences: A system dynamics approach in gamification". *Journal of Ambient Intelligence and Humanized Computing*, 2019, pp. 1–17. DOI: 10.1007/s12652-019-01505-w.

242. D. R. Krathwohl. "A Revision of Bloom's Taxonomy: An Overview". *Theory Into Practice* 41:4, 2002, pp. 212–218. DOI: 10.1207/s15430421tip4104_2.

243. V. Krauß, A. Boden, L. Oppermann, and R. Reiners. "Current Practices, Challenges, and Design Implications for Collaborative AR/VR Application Development". In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI '21. Association for Computing Machinery, Yokohama, Japan, 2021. ISBN: 9781450380966. DOI: 10.1145/3411764.3445335.

244. S. Kurkovsky, R. Koshy, V. Novak, and P. Szul. "Current issues in handheld augmented reality". In: *2012 International Conference on Communications and Information Technology (ICCIT)*. 2012, pp. 68–72. DOI: 10.1109/ICCITechnol.2012.6285844.

245. K. C. Kwan and H. Fu. "Mobi3DSketch: 3D Sketching in Mobile AR". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. Association for Computing Machinery, Glasgow, Scotland Uk, 2019, pp. 1–11. ISBN: 9781450359702. DOI: 10.1145/3290605.3300406.

246. M. Lahti, H. Hätönen, and M. Välimäki. "Impact of e-learning on nurses' and student nurses knowledge, skills, and satisfaction: A systematic review and meta-analysis". *International Journal of Nursing Studies* 51:1, 2014, pp. 136–149. ISSN: 0020-7489. DOI: https://doi.org/10.1016/j.ijnurstu.2012.12.017.

247. T. H. Laine, E. Nygren, A. Dirin, and H.-J. Suk. "Science Spots AR: a platform for science learning games with augmented reality". *Educational Technology Research and Development* 64:3, 2016, pp. 507–531. DOI: 10.1007/s11423-015-9419-0.

248. T. G. Lane. "Studying software architecture through design spaces and rules", 1990.

249. T. Lang, B. MacIntyre, and I. J. Zugaza. "Massively Multiplayer Online Worlds as a Platform for Augmented Reality Experiences". In: *2008 IEEE Virtual Reality Conference*. 2008, pp. 67–70. DOI: 10.1109/VR.2008.4480752.

250. T. Langlotz, S. Mooslechner, and D. Schmalstieg. "In-Situ Content Creation for Mobile Augmented Reality". Note: Full bibliographic details could not be retrieved (29.05.2023).

251. T. Langlotz, S. Mooslechner, S. Zollmann, C. Degendorfer, G. Reitmayr, and D. Schmalstieg. "Sketching up the world: in situ authoring for mobile augmented reality". *Personal and ubiquitous computing* 16:6, 2012, pp. 623–630. DOI: 10.1007/s00779-011-0430-0.

252. T. Langlotz, H. Regenbrecht, and D. Schmalstieg. "Large scale content creation for mobile augmented reality". *Postgraduate Day*, 2011, p. 9.

253. T. Langlotz, H. Regenbrecht, S. Zollmann, and D. Schmalstieg. "Audio Stickies: Visually-Guided Spatial Audio Annotations on a Mobile Augmented Reality Platform". In: *Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration*. OzCHI '13. Association for Computing Machinery, Adelaide, Australia, 2013, pp. 545–554. ISBN: 9781450325257. DOI: 10.1145/2541016.2541022.

254. T. Langlotz, D. Wagner, A. Mulloni, and D. Schmalstieg. "Online Creation of Panoramic Augmented Reality Annotations on Mobile Phones". *IEEE Pervasive Computing* 11:2, 2012, pp. 56–63. DOI: 10.1109/MPRV.2010.69.

255. T. Langlotz, M. Zingerle, R. Grasset, H. Kaufmann, and G. Reitmayr. "AR Record&Replay: Situated Compositing of Video Content in Mobile Augmented Reality". In: *Proceedings of the 24th Australian Computer-Human Interaction Conference*. OzCHI '12. Association for Computing Machinery, Melbourne, Australia, 2012, pp. 318–326. ISBN: 9781450314381. DOI: 10.1145/2414536.2414588.

256. M. Lau, M. Hirose, A. Ohgawara, J. Mitani, and T. Igarashi. "Situated Modeling: A Shape-Stamping Interface with Tangible Primitives". In: *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction*. TEI '12. Association for Computing Machinery, Kingston, Ontario, Canada, 2012, pp. 275–282. ISBN: 9781450311748. DOI: 10.1145/2148131.2148190.

257. B. Laugwitz, T. Held, and M. Schrepp. "Construction and Evaluation of a User Experience Questionnaire". In: *HCI and Usability for Education and Work*. Ed. by A. Holzinger. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008, pp. 63–76. ISBN: 978-3-540-89350-9. DOI: 10.1007/978-3-540-89350-9_6.

258. T. LAVRIC, E. BRICARD, M. PREDA, and T. ZAHARIA. "An AR Work Instructions Authoring Tool for Human-Operated Industrial Assembly Lines". In: *2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. 2020, pp. 174–183. DOI: 10.1109/AIVR50618.2020.00037.

259. M. Lechner. "ARML 2.0 in the context of existing AR data formats". In: *2013 6th Workshop on Software Engineering and Architectures for Realtime Interactive Systems (SEARIS)*. 2013, pp. 41–47. DOI: 10.1109/SEARIS.2013.6798107.

260. M. Lechner and M. Tripp. "ARML—an augmented reality standard". *coordinates* 13:47. 797222, 2010, pp. 432–440.

261. F. Lécuyer, V. Gouranton, A. Reuzeau, R. Gaugne, and B. Arnaldi. "Authoring AR Interaction by AR". In: *ICAT-EGVE 2019 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*. Ed. by Y. Kakehi and A. Hiyama. The Eurographics Association, 2019, pp. 1–8. ISBN: 978-3-03868-083-3. DOI: 10.2312/egve.20191285.
262. F. Ledermann and D. Schmalstieg. "APRIL: a high-level framework for creating augmented reality presentations". In: *IEEE Proceedings. VR 2005. Virtual Reality, 2005.* 2005, pp. 187–194. DOI: 10.1109/VR.2005.1492773.
263. F. Ledermann and D. Schmalstieg. "APRIL: a high-level framework for creating augmented reality presentations". In: *IEEE Proceedings. VR 2005. Virtual Reality, 2005.* 2005, pp. 187–194. DOI: 10.1109/VR.2005.1492773.
264. F. Ledermann. "An authoring framework for augmented reality presentations". Diploma Thesis. Technischen Universität Wien, 2004.
265. F. Ledermann, I. Barakonyi, and D. Schmalstieg. "Abstraction and implementation strategies for augmented reality authoring". In: *Emerging Technologies of Augmented Reality: Interfaces and Design*. IGI Global, 2007, pp. 138–159. DOI: 10.4018/978-1-59904-066-0.ch007.
266. G. Lee, C. Nelles, M. Billinghurst, and G. Kim. "Immersive authoring of tangible augmented reality applications". In: *Third IEEE and ACM International Symposium on Mixed and Augmented Reality*. 2004, pp. 172–181. DOI: 10.1109/ISMAR.2004.34.
267. G. A. Lee and M. Billinghurst. "A Component Based Framework for Mobile Outdoor AR Applications". In: *Proceedings of the 12th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*. VRCAI '13. Association for Computing Machinery, Hong Kong, Hong Kong, 2013, pp. 207–210. ISBN: 9781450325905. DOI: 10.1145/2534329.2534344.
268. G. A. Lee and G. J. Kim. "Immersive authoring of Tangible Augmented Reality content: A user study". *Journal of Visual Languages & Computing* 20:2, 2009, pp. 61–79. ISSN: 1045-926X. DOI: 10.1016/j.jvlc.2008.07.001.
269. H. Lee, J. Lee, S. Ahn, H. Ko, and G. J. Kim. "Collaborative AR Authoring Framework using Remote Panorama and In-situ Sensors", 2014.
270. J. Y. Lee and G. Rhee. "Context-aware 3D visualization and collaboration services for ubiquitous cars using augmented reality". *The International Journal of Advanced Manufacturing Technology* 37:5, 2008, pp. 431–442. DOI: 10.1007/s00170-007-0996-x.
271. J. Y. Lee and D. W. Seo. "A Context-Aware and Augmented Reality-Supported Service Framework in Ubiquitous Environments". In: *Proceedings of the 2005 International Conference on Embedded and Ubiquitous Computing*. EUC'05. Springer-Verlag, Nagasaki, Japan, 2005, pp. 258–267. ISBN: 3540308032. DOI: 10.1007/11596042_27.

272. J. Y. Lee, D. W. Seo, and G. W. Rhee. "Tangible authoring of 3D virtual scenes in dynamic augmented reality environment". *Computers in Industry* 62:1, 2011, pp. 107–119. ISSN: 0166-3615. DOI: 10.1016/j.compind.2010.07.003.

273. S.-W. Lee, J. Jung, J. Hong, S. Lee, H. Cho, and H. S. Yang. "Efficient 3D content authoring framework based on mobile AR". In: *2012 18th International Conference on Virtual Systems and Multimedia*. 2012, pp. 95–102. DOI: 10.1109/VSMM.2012.6365912.

274. C. Lewa, M. Joswig, K. Luksch, J. Blattgerste, T. Pfeiffer, N. Bauer, A. Bernloehr, and T. Schäfer. "Praxisorientierte arbeitsprozessgeleitete Lernzieldefinition und deren Transformation in eine Augmented Reality Lehr-/Lernbegleitung für die hochschulische Hebammenausbildung". *Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA)*, 2021. DOI: 10.3205/21gma016.

275. C. Lewa, M. Joswig, T. Willmeroth, K. Luksch, J. Blattgerste, T. Pfeiffer, A. Bernloehr, N. Bauer, and T. Schäfer. "Notfallszenarien transformiert in eine Augmented Reality Lehr-/Lernbegleitung für ein realitätsnahes Training in der hochschulischen Hebammenausbildung". *Gemeinsame Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA) und des Arbeitskreises zur Weiterentwicklung der Lehre in der Zahnmedizin (AKWLZ)*, 2022. DOI: 10.3205/22gma175.

276. J. R. Lewis. *Perceived Usability - Usefulness and measurement*. https://www.torchi.org/resources/Documents/2020-Apr-7%20Jim%20Lewis%20-%20Perceived%20Usability%20-%20slides.pdf. Accessed: 2022-01-12. 2020.

277. J. R. Lewis and J. Sauro. "Can I Leave This One out? The Effect of Dropping an Item from the SUS". *J. Usability Studies* 13:1, 2017, pp. 38–46.

278. J. R. Lewis and J. Sauro. "Item Benchmarks for the System Usability Scale". *J. Usability Studies* 13:3, 2018, pp. 158–167.

279. J. R. Lewis and J. Sauro. "The Factor Structure of the System Usability Scale". In: *Human Centered Design*. Ed. by M. Kurosu. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, pp. 94–103. ISBN: 978-3-642-02806-9.

280. A. Lharchi, M. Tamke, and M. Ramsgaard Thomsen. "Connected Augmented Assembly: Cloud based Augmented Reality applications in architecture". English. 1, 2020. Ed. by L. Werner and D. Koering, pp. 179–186.

281. B. Li, Q. Dong, J. Dong, J. Wang, W. Li, and S. Li. "Instruction Manual for Product Assembly Process Based on Augmented Visualization". In: *2018 Chinese Automation Congress (CAC)*. 2018, pp. 3248–3253. DOI: 10.1109/CAC.2018.8623583.

282. Z. Li, M. Annett, K. Hinckley, K. Singh, and D. Wigdor. "HoloDoc: Enabling Mixed Reality Workspaces That Harness Physical and Digital Content". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. Association for Computing Machinery, Glasgow, Scotland Uk, 2019, pp. 1–14. ISBN: 9781450359702. DOI: 10.1145/3290605.3300917.

283. F. Liarokapis, N. Mourkoussis, M. White, J. Darcy, M. Sifniotis, P. Petridis, A. Basu, P. F. Lister, et al. "Web3D and augmented reality to support engineering education". *World transactions on engineering and technology education* 3:1, 2004, pp. 11–14.

284. S. Lim and J. Lee. "An Immersive Augmented-Reality-Based e-Learning System Based on Dynamic Threshold Marker Method". *ETRI Journal* 35:6, 2013, pp. 1048–1057. DOI: 10.4218/etrij.13.2013.0081.

285. B. H. Limbu, H. Jarodzka, R. Klemke, and M. Specht. "Using sensors and augmented reality to train apprentices using recorded expert performance: A systematic literature review". *Educational Research Review* 25, 2018, pp. 1–22. ISSN: 1747-938X. DOI: 10.1016/j.edurev.2018.07.001.

286. R. Lindgren, M. Tscholl, S. Wang, and E. Johnson. "Enhancing learning and engagement through embodied interaction within a mixed reality simulation". *Computers & Education* 95, 2016, pp. 174–187. ISSN: 0360-1315. DOI: https://doi.org/10.1016/j.compedu.2016.01.001.

287. M. S. Linneberg and S. Korsgaard. "Coding qualitative data: a synthesis guiding the novice". *Qualitative Research Journal* 19:3, 2019, pp. 259–270. ISSN: 1443-9883. DOI: 10.1108/qrj-12-2018-0012.

288. C. Liu, J. Diehl, S. Huot, and J. Borchers. "Mobile Augmented Note-taking to Support Operating Physical Devices", 2011.

289. J. Looser, R. Grasset, H. Seichter, and M. Billinghurst. "OSGART-A pragmatic approach to MR", 2006.

290. M. Lorenz, S. Knopp, J. Kim, and P. Klimant. "Industrial Augmented Reality: 3D-Content Editor for Augmented Reality Maintenance Worker Support System". In: *2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. 2020, pp. 203–205. DOI: 10.1109/ISMAR-Adjunct51615.2020.00060.

291. Y. Lu and T. Ishida. "Implementation and Evaluation of a High-presence Interior Layout Simulation System using Mixed Reality." *J. Internet Serv. Inf. Secur.* 10:1, 2020, pp. 50–63. DOI: 10.22667/JISIS.2020.02.29.050.

292. M. Lucrecia, S. Cecilia, P. Patricia, and B. Sandra. "AuthorAR: Authoring tool for building educational activities based on Augmented Reality". In: *2013 International Conference on Collaboration Technologies and Systems (CTS)*. 2013, pp. 503–507. DOI: 10.1109/CTS.2013.6567277.

293. K. Luksch, N. Bauer, J. Blattgerste, C. Lewa, T. Schäfer, and A. Bernloehr. "Notfalltraining per Smartphone–Augmented Reality (AR) gestütztes Training zur Reanimation eines Neugeborenen". *Zeitschrift für Geburtshilfe und Neonatologie* 225:S 01, 2021, P–118. DOI: 10.1055/s-0041-1739881.

294. C. Lytridis, A. Tsinakos, and I. Kazanidis. "ARTutor—An Augmented Reality Platform for Interactive Distance Learning". *Education Sciences* 8:1, 2018. ISSN: 2227-7102. DOI: 10.3390/educsci8010006.

295. T. L. MacDonald Colla J. ; Thompson. "Structure, Content, Delivery, Service, and Outcomes: Quality e-Learning in higher education". *International Review of Research in Open and Distributed Learning* 6:2, 2005, pp. 1–25. DOI: 10.19173/irrodl.v6i2.237.

296. B. MacIntyre, M. Gandy, S. Dow, and J. D. Bolter. "DART: A Toolkit for Rapid Design Exploration of Augmented Reality Experiences". In: *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology*. UIST '04. Association for Computing Machinery, Santa Fe, NM, USA, 2004, pp. 197–206. ISBN: 1581139578. DOI: 10.1145/1029632.1029669.

297. B. MacIntyre, A. Hill, H. Rouzati, M. Gandy, and B. Davidson. "The Argon AR Web Browser and standards-based AR application environment". In: *2011 10th IEEE International Symposium on Mixed and Augmented Reality*. 2011, pp. 65–74. DOI: 10.1109/ISMAR.2011.6092371.

298. B. MacIntyre and M. Lohse. "Ghosts in the Machine: Integrating 2D Video Actors into a 3D AR System". In: *In Proc. International Symposium on Mixed Reality*. Citeseer. 2001.

299. A. MacLean, V. Bellotti, and S. Shum. "Developing the design space with design space analysis". *Computers, Communication and Usability: Design issues, research and methods for integrated services*, 1993, pp. 197–219.

300. J. Madar, C. C. Roehr, S. Ainsworth, H. Ersdal, C. Morley, M. Rüdiger, C. Skåre, T. Szczapa, A. te Pas, D. Trevisanuto, B. Urlesberger, D. Wilkinson, and J. P. Wyllie. "European Resuscitation Council Guidelines 2021: Newborn resuscitation and support of transition of infants at birth". *Resuscitation* 161, 2021, pp. 291–326. ISSN: 0300-9572. DOI: 10.1016/j.resuscitation.2021.02.014.

301. T. Madeira, B. Marques, P. Neves, P. Dias, and B. S. Santos. "Comparing Desktop vs. Mobile Interaction for the Creation of Pervasive Augmented Reality Experiences". *Journal of Imaging* 8:3, 2022. ISSN: 2313-433X. DOI: 10.3390/jimaging8030079.

302. S. Mader and B. Urban. "Creating instructional content for augmented reality based on controlled natural language concepts". In: *Proceedings of 20th International Conference on Artificial Reality and Telexistence (ICAT 2010)*. 2010, pp. 313–321. ISBN: 978-4-904490-03-7.

303. L. Maia, W. Rodrigues, W. Viana, and F. Trinta. "Auragame: A Case Study of a Zero Programming Augmented Reality Game". In: 2016.

304. L. F. Maia, C. Nolêto, M. Lima, C. Ferreira, C. Marinho, W. Viana, and F. Trinta. "LAGARTO: A LocAtion based Games AuthoRing TOol enhanced with augmented reality features". *Entertainment Computing* 22, 2017, pp. 3–13. ISSN: 1875-9521. DOI: 10.1016/j.entcom.2017.05.001.

305. S. Makris, G. Pintzos, L. Rentzos, and G. Chryssolouris. “Assembly support using AR technology based on automatic sequence generation”. *CIRP Annals* 62:1, 2013, pp. 9–12. ISSN: 0007-8506. DOI: 10.1016/j.cirp.2013.03.095.

306. D. Markov-Vetter, J. Millberg, and O. Staadt. “Mobile Augmented Reality for Space Operation Procedures: A Generic Approach of Authoring and Guiding On-Board Payload Activities”. *Proceedings of the International Astronautical Congress, IAC* 6, 2013, pp. 1–14.

307. I. M. Marstio, J. S. Sääski, T. S. Salonen, C. W. Woodward, K. R. Rainio, and O. K. Korkalo. “Data pipeline from CAD to AR based assembly instructions”. In: *ASME world conference on innovative virtual reality*. Vol. 43376. 2009, pp. 165–168. DOI: 10.1115/WINVR2009-705.

308. A. Martin-Martin, E. Orduna-Malea, A.-W. Harzing, and E. Delgado López-Cózar. “Can we use Google Scholar to identify highly-cited documents?” *Journal of Informetrics* 11:1, 2017, pp. 152–163. ISSN: 1751-1577. DOI: 10.1016/j.joi.2016.11.008.

309. A. Martín-Martín, M. Thelwall, E. Orduna-Malea, and E. Delgado López-Cózar. “Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations’ COCI: a multidisciplinary comparison of coverage via citations”. *Scientometrics* 126:1, 2021, pp. 871–906. DOI: 10.1007/s11192-020-03690-4.

310. H. Martínez and S. Laukkanen. “STEDUS, a new educational platform for Augmented Reality applications”. In: *At the Edge of the Rift*. Brill, Leiden, The Netherlands, 2014, pp. 37–49. ISBN: 9781848883215. DOI: 10.1163/9781848883215_005.

311. H. Martínez, S. Laukkanen, and J. Mattila. “A new flexible augmented reality platform for development of maintenance and educational applications”. *International Journal of Virtual Worlds and Human Computer Interaction* 2:1, 2014, pp. 18–27. DOI: 10.11159/vwhci.2014.003.

312. T. Masood and J. Egger. “Augmented reality in support of Industry 4.0—Implementation challenges and success factors”. *Robotics and Computer-Integrated Manufacturing* 58, 2019, pp. 181–195. ISSN: 0736-5845. DOI: 10.1016/j.rcim.2019.02.003.

313. C. Matysczok, P. Ebbesmeyer, H. Krumm, and J. Maciej. “Efficient Creation of Augmented Reality Content by Using an Intuitive Authoring System”. In: vol. Volume 4: 24th Computers and Information in Engineering Conference. International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. 2004, pp. 53–60. DOI: 10.1115/DETC2004-57653.

314. P. Mayring et al. “Qualitative content analysis”. *A companion to qualitative research* 1:2, 2004, pp. 159–176.

315. F. Meawad. “InterAKT: A mobile augmented reality browser for geo-social mashups”. In: *2016 4th International Conference on User Science and Engineering (i-USEr)*. 2016, pp. 167–171. DOI: 10.1109/IUSER.2016.7857954.

316. M. Meccawy. "Creating an Immersive XR Learning Experience: A Roadmap for Educators". *Electronics* 11:21, 2022. ISSN: 2079-9292. DOI: 10.3390/electronics11213547.

317. L. I. Meho and Y. Rogers. "Citation counting, citation ranking, and h-index of human-computer interaction researchers: A comparison of Scopus and Web of Science". *Journal of the American Society for Information Science and Technology* 59:11, 2008, pp. 1711–1726. DOI: https://doi.org/10.1002/asi.20874.

318. M. Mengoni, M. Peruzzini, M. Bordegoni, and M. Mecella. "Toward an integrated platform to support contract furniture industry". *Computer-Aided Design and Applications* 13:5, 2016, pp. 662–674. DOI: 10.1080/16864360.2016.1150712.

319. L. Merino, M. Schwarzl, M. Kraus, M. Sedlmair, D. Schmalstieg, and D. Weiskopf. "Evaluating Mixed and Augmented Reality: A Systematic Literature Review (2009-2019)". In: *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 2020, pp. 438–451. DOI: 10.1109/ISMAR50242.2020.00069.

320. Microsoft. *C# Coding Conventions*. https://learn.microsoft.com/en-us/dotnet/csharp/fundamentals/coding-style/coding-conventions. Accessed: 02.05.2023. 2021.

321. P. Milgram and F. Kishino. "A taxonomy of mixed reality visual displays". *IEICE TRANSACTIONS on Information and Systems* 77:12, 1994, pp. 1321–1329.

322. P. Milgram, H. Takemura, A. Utsumi, and F. Kishino. "Augmented reality: a class of displays on the reality-virtuality continuum". In: *Telemanipulator and Telepresence Technologies*. Ed. by H. Das. Vol. 2351. International Society for Optics and Photonics. SPIE, 1995, pp. 282–292. DOI: 10.1117/12.197321.

323. G. E. Miller. "The assessment of clinical skills/competence/performance". *Academic medicine* 65:9, 1990, S63–7.

324. A. Millette and M. J. McGuffin. "DualCAD: Integrating Augmented Reality with a Desktop GUI and Smartphone Interaction". In: *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*. 2016, pp. 21–26. DOI: 10.1109/ISMAR-Adjunct.2016.0030.

325. Á. Montero, T. Zarraonandia, P. Díaz, and I. Aedo. "An End-User Tool for Creating Augmented Reality Experiences". *IS-EUD 2017 6th international symposium on end-user development*, 2017, p. 63.

326. J. M. Mota, I. Ruiz-Rube, J. M. Dodero, and I. Arnedillo-Sánchez. "Augmented reality mobile app development for all". *Computers & Electrical Engineering* 65, 2018, pp. 250–260. ISSN: 0045-7906. DOI: 10.1016/j.compeleceng.2017.08.025.

327. E. Mourão, J. F. Pimentel, L. Murta, M. Kalinowski, E. Mendes, and C. Wohlin. "On the performance of hybrid search strategies for systematic literature reviews in software engineering". *Information and Software Technology* 123, 2020, p. 106294. ISSN: 0950-5849. DOI: 10.1016/j.infsof.2020.106294.

328. N. Mourkoussis, F. Liarokapis, A. Basu, M. White, and P. Lister. "Using XML Technologies to Present Digital Content with Augmented Reality". In: *Eurographics Ireland Chapter Workshop Proceedings*. Vol. 3. 2004.

329. D. Mourtzis, J. Angelopoulos, and V. Zogopoulos. "Integrated and adaptive AR maintenance and shop-floor rescheduling". *Computers in Industry* 125, 2021, p. 103383. ISSN: 0166-3615. DOI: 10.1016/j.compind.2020.103383.

330. D. Mourtzis, V. Siatras, J. Angelopoulos, and N. Panopoulos. "An Augmented Reality Collaborative Product Design Cloud-Based Platform in the Context of Learning Factory". *Procedia Manufacturing* 45, 2020. Learning Factories across the value chain – from innovation to service – The 10th Conference on Learning Factories 2020, pp. 546–551. ISSN: 2351-9789. DOI: 10.1016/j.promfg.2020.04.076.

331. D. Mourtzis, F. Xanthi, and V. Zogopoulos. "An Adaptive Framework for Augmented Reality Instructions Considering Workforce Skill". *Procedia CIRP* 81, 2019. 52nd CIRP Conference on Manufacturing Systems (CMS), Ljubljana, Slovenia, June 12-14, 2019, pp. 363–368. ISSN: 2212-8271. DOI: 10.1016/j.procir.2019.03.063.

332. S. S. Muhammad Nizam, R. Zainal Abidin, N. Che Hashim, M. C. Lam, H. Arshad, and N. A. Abd Majid. "A review of multimodal interaction technique in augmented reality environment". *Int. J. Adv. Sci. Eng. Inf. Technol* 8:4-2, 2018, p. 1460. ISSN: 2460-6952, 2088-5334. DOI: 10.18517/ijaseit.8.4-2.6824.

333. L. Müller, I. Aslan, and L. Krüßen. "GuideMe: A Mobile Augmented Reality System to Display User Manuals for Home Appliances". In: *Advances in Computer Entertainment*. Ed. by D. Reidsma, H. Katayose, and A. Nijholt. Springer International Publishing, Cham, 2013, pp. 152–167. ISBN: 978-3-319-03161-3.

334. T. Müller and T. Rieger. "[POSTER] ARPML: The Augmented Reality Process Modeling Language". In: *2015 IEEE International Symposium on Mixed and Augmented Reality*. 2015, pp. 160–163. DOI: 10.1109/ISMAR.2015.46.

335. J. A. Muñoz-Cristóbal, V. Gallego-Lema, H. F. Arribas-Cubero, J. I. Asensio-Pérez, and A. Martínez-Monés. "Game of Blazons: Helping Teachers Conduct Learning Situations That Integrate Web Tools and Multiple Types of Augmented Reality". *IEEE Transactions on Learning Technologies* 11:4, 2018, pp. 506–519. DOI: 10.1109/TLT.2018.2808491.

336. Y. Nam. "Designing interactive narratives for mobile augmented reality". *Cluster Computing* 18:1, 2015, pp. 309–320. DOI: 10.1007/s10586-014-0354-3.

337. B. Napoleão, K. R. Felizardo, É. F. de Souza, and N. L. Vijaykumar. "Practical similarities and differences between Systematic Literature Reviews and Systematic Mappings: a tertiary study." In: *SEKE*. Vol. 2017. 2017, pp. 85–90. DOI: 10.18293/seke2017-069.

338. A. Nassani, H. Bai, G. Lee, and M. Billinghurst. “Tag It! AR Annotation Using Wearable Sensors”. In: *SIGGRAPH Asia 2015 Mobile Graphics and Interactive Applications*. SA ’15. Association for Computing Machinery, Kobe, Japan, 2015. ISBN: 9781450339285. DOI: 10.1145/2818427.2818438.

339. M. Nebeling. “Playing the Tricky Game of Toolkits Research”. In: *Workshop on HCI. Tools at CHI 2017*. 2017.

340. M. Nebeling and M. Speicher. “The trouble with augmented reality/virtual reality authoring tools”. In: *2018 IEEE international symposium on mixed and augmented reality adjunct (ISMAR-Adjunct)*. IEEE. 2018, pp. 333–337. DOI: 10.1109/ISMAR-Adjunct.2018.00098.

341. M. Neges and C. Koch. “Augmented reality supported work instructions for onsite facility maintenance”. In: *23rd International Workshop of the European Group for Intelligent Computing in Engineering*. 2016.

342. M. Neges, M. Wolf, and M. Abramovici. “Secure Access Augmented Reality Solution for Mobile Maintenance Support Utilizing Condition-Oriented Work Instructions”. *Procedia CIRP* 38, 2015. Proceedings of the 4th International Conference on Through-life Engineering Services, pp. 58–62. ISSN: 2212-8271. DOI: 10.1016/j.procir.2015.08.036.

343. G. Ng, J. G. Shin, A. Plopski, C. Sandor, and D. Saakes. “Situated Game Level Editing in Augmented Reality”. In: *Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction*. TEI ’18. Association for Computing Machinery, Stockholm, Sweden, 2018, pp. 409–418. ISBN: 9781450355681. DOI: 10.1145/3173225.3173230.

344. T. V. Nguyen, B. Mirza, D. Tan, and J. Sepulveda. “ASMIM: Augmented Reality Authoring System for Mobile Interactive Manuals”. In: *Proceedings of the 12th International Conference on Ubiquitous Information Management and Communication*. IMCOM ’18. Association for Computing Machinery, Langkawi, Malaysia, 2018. ISBN: 9781450363853. DOI: 10.1145/3164541.3164592.

345. T. V. Nguyen, D. Tan, B. Mirza, and J. Sepulveda. “MARIM: Mobile Augmented Reality for Interactive Manuals”. In: *Proceedings of the 24th ACM International Conference on Multimedia*. MM ’16. Association for Computing Machinery, Amsterdam, The Netherlands, 2016, pp. 689–690. ISBN: 9781450336031. DOI: 10.1145/2964284.2973807.

346. V. T. Nguyen, K. Jung, and T. Dang. “BlocklyAR: A Visual Programming Interface for Creating Augmented Reality Experiences”. *Electronics* 9:8, 2020. ISSN: 2079-9292. DOI: 10.3390/electronics9081205.

347. J. Nielsen. *Top 10 Mistakes in Web Design - Violating Design Conventions*. https://www.nngroup.com/articles/top-10-mistakes-web-design/. Accessed: 18.05.2023. 2023.

348. C. Nolêto, M. Lima, L. F. Maia, W. Viana, and F. Trinta. "An Authoring Tool for Location-Based Mobile Games with Augmented Reality Features". In: *2015 14th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames)*. 2015, pp. 99–108. DOI: 10.1109/SBGames.2015.12.

349. E. Not and D. Petrelli. "Empowering cultural heritage professionals with tools for authoring and deploying personalised visitor experiences". *User Modeling and User-Adapted Interaction* 29:1, 2019, pp. 67–120. DOI: 10.1007/s11257-019-09224-9.

350. T. Novotny, I. Lindt, and W. Broll. "A Multi Modal Table-Top 3D Modeling Tool in Augmented Environments". In: *Proceedings of the 12th Eurographics Conference on Virtual Environments*. EGVE'06. Eurographics Association, Lisbon, Portugal, 2006, pp. 45–52. ISBN: 3905673339.

351. B. Nuernberger, K.-C. Lien, T. Höllerer, and M. Turk. "Interpreting 2D gesture annotations in 3D augmented reality". In: *2016 IEEE Symposium on 3D User Interfaces (3DUI)*. 2016, pp. 149–158. DOI: 10.1109/3DUI.2016.7460046.

352. M. O'Connor and C. E. Hughes. "Authoring and delivering mixed reality experiences". In: *Proceedings of 2005 International Conference on Human-Computer Interface Advances in Modeling and Simulation (SIMCHI'05)*. 2005, pp. 23–27.

353. *OGC® Augmented Reality Markup Language 2.0 (ARML 2.0)*. https://www.ogc.org/standard/arml/. Accessed: 17.04.2023.

354. A. Oliveira and R. B. Araujo. "Creation and Visualization of Context Aware Augmented Reality Interfaces". In: *Proceedings of the International Working Conference on Advanced Visual Interfaces*. AVI '12. Association for Computing Machinery, Capri Island, Italy, 2012, pp. 324–327. ISBN: 9781450312875. DOI: 10.1145/2254556.2254618.

355. S. Ong and J. Zhu. "A novel maintenance system for equipment serviceability improvement". *CIRP Annals* 62:1, 2013, pp. 39–42. ISSN: 0007-8506. DOI: 10.1016/j.cirp.2013.03.091.

356. G. Osorio-Gómez, R. Viganò, and J. C. Arbeláez. "An augmented reality tool to validate the assembly sequence of a discrete product". *International Journal of Computer Aided Engineering and Technology* 8:1-2, 2016, pp. 164–178. DOI: 10.1504/IJCAET.2016.073263.

357. P. Oswald, J. Tost, and R. Wettach. "The Real Augmented Reality: Real-Time Game Editor in a Spatial Augmented Environment". In: *Proceedings of the 11th Conference on Advances in Computer Entertainment Technology*. ACE '14. Association for Computing Machinery, Funchal, Portugal, 2014. ISBN: 9781450329453. DOI: 10.1145/2663806.2663853.

358. M. Otto, M. Prieur, P. Agethen, and E. Rukzio. "Dual Reality for Production Verification Workshops: A Comprehensive Set of Virtual Methods". *Procedia CIRP* 44, 2016. 6th CIRP Conference on Assembly Technologies and Systems (CATS), pp. 38–43. ISSN: 2212-8271. DOI: 10.1016/j.procir.2016.02.140.

359. M. Ozdemir, C. Sahin, S. Arcagok, and M. K. Demir. “The effect of augmented reality applications in the learning process: A meta-analysis study”. *Eurasian Journal of Educational Research* 18:74, 2018, pp. 165–186. DOI: 10.14689/ejer.2018.74.9.

360. M. J. Page, D. Moher, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, R. Chou, J. Glanville, J. M. Grimshaw, A. Hróbjartsson, M. M. Lalu, T. Li, E. W. Loder, E. Mayo-Wilson, S. McDonald, L. A. McGuinness, L. A. Stewart, J. Thomas, A. C. Tricco, V. A. Welch, P. Whiting, and J. E. McKenzie. “PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews”. *BMJ* 372, 2021. DOI: 10.1136/bmj.n160.

361. M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, R. Chou, J. Glanville, J. M. Grimshaw, A. Hróbjartsson, M. M. Lalu, T. Li, E. W. Loder, E. Mayo-Wilson, S. McDonald, L. A. McGuinness, L. A. Stewart, J. Thomas, A. C. Tricco, V. A. Welch, P. Whiting, and D. Moher. “The PRISMA 2020 statement: An updated guideline for reporting systematic reviews”. *International Journal of Surgery* 88, 2021, p. 105906. ISSN: 1743-9191. DOI: 10.1016/j.ijsu.2021.105906.

362. R. Palmarini, J. A. Erkoyuncu, R. Roy, and H. Torabmostaedi. “A systematic review of augmented reality applications in maintenance”. *Robotics and Computer-Integrated Manufacturing* 49, 2018, pp. 215–228. ISSN: 0736-5845. DOI: https://doi.org/10.1016/j.rcim.2017.06.002.

363. H. Park and W. Woo. “Metadata Design for Location-Based Film Experience in Augmented Places”. In: *2015 IEEE International Symposium on Mixed and Augmented Reality - Media, Art, Social Science, Humanities and Design*. 2015, pp. 40–45. DOI: 10.1109/ISMAR-MASHD.2015.12.

364. J. Park, B. Son, J. Seo, and C. Park. “Development of Authoring Tool for Extended Interaction in Mixed Reality Environments”. In: *HCI International 2018 – Posters’ Extended Abstracts*. Ed. by C. Stephanidis. Springer International Publishing, Cham, 2018, pp. 304–310. ISBN: 978-3-319-92279-9. DOI: 10.1007/978-3-319-92279-9_41.

365. J. Park, S. Sil Kim, H. Park, and W. Woo. “"DreamHouse": NUI-Based Photo-Realistic AR Authoring System for Interior Design”. In: *Proceedings of the 7th Augmented Human International Conference 2016*. AH ’16. Association for Computing Machinery, Geneva, Switzerland, 2016. ISBN: 9781450336802. DOI: 10.1145/2875194.2875221.

366. J.-S. Park. “AR-Room: a rapid prototyping framework for augmented reality applications”. *Multimedia tools and applications* 55:3, 2011, pp. 725–746. DOI: 10.1007/s11042-010-0592-1.

367. J. Park and W. Woo. "Multi-layer Based Authoring Tool for Digilog Book". In: *Entertainment Computing – ICEC 2009*. Ed. by S. Natkin and J. Dupire. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, pp. 234–239. ISBN: 978-3-642-04052-8. DOI: 10.1007/978-3-642-04052-8_26.

368. J. Park and C. Lim. "Mixed Reality Based Interactive 3D Story Composition Tool". In: *Technologies for Interactive Digital Storytelling and Entertainment*. Ed. by S. Göbel, R. Malkewitz, and I. Iurgel. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006, pp. 181–186. ISBN: 978-3-540-49935-0. DOI: 10.1007/11944577_18.

369. S. Y. Park. "An Analysis of the Technology Acceptance Model in Understanding University Students' Behavioral Intention to Use e-Learning". *Journal of Educational Technology & Society* 12:3, 2009, pp. 150–162. ISSN: 11763647, 14364522.

370. Y. J. Park, Y. Yang, S. Chae, I. Kim, and T.-d. Han. "DesignAR: Portable projection-based AR system specialized in interior design". In: *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 2017, pp. 2879–2884. DOI: 10.1109/SMC.2017.8123064.

371. L. Pastoor. "Cloud Anbindung von Augmented Reality Inhalten". Bachelors's Thesis. University of Applied Sciences Emden/Leer, 2020.

372. K. Pentenrieder, C. Bade, F. Doil, and P. Meier. "Augmented Reality-based factory planning - an application tailored to industrial needs". In: *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*. 2007, pp. 31–42. DOI: 10.1109/ISMAR.2007.4538822.

373. S. A. Pessoa, G. de S. Moura, J. P. S. do M. Lima, V. Teichrieb, and J. Kelner. "RPR-SORS: Real-time photorealistic rendering of synthetic objects into real scenes". *Computers & Graphics* 36:2, 2012, pp. 50–69. ISSN: 0097-8493. DOI: 10.1016/j.cag.2011.12.003.

374. K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson. "Systematic mapping studies in software engineering". In: *12th International Conference on Evaluation and Assessment in Software Engineering (EASE) 12*. 2008, pp. 1–10. DOI: 10.14236/ewic/EASE2008.8.

375. N. Petersen, A. Pagani, and D. Stricker. "Real-time modeling and tracking manual workflows from first-person vision". In: *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 2013, pp. 117–124. DOI: 10.1109/ISMAR.2013.6671771.

376. N. Petersen and D. Stricker. "Learning task structure from video examples for workflow tracking and authoring". In: *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 2012, pp. 237–246. DOI: 10.1109/ISMAR.2012.6402562.

377. K. Pietroszek and C. Moore. "AHMED: Toolset for Ad-Hoc Mixed-Reality Exhibition Design". In: *25th ACM Symposium on Virtual Reality Software and Technology*. VRST '19. Association for Computing Machinery, Parramatta, NSW, Australia, 2019. ISBN: 9781450370011. DOI: 10.1145/3359996.3364729.

378. A. Plopski, V. Fuvattanasilp, J. Poldi, T. Taketomi, C. Sandor, and H. Kato. "Efficient In-Situ Creation of Augmented Reality Tutorials". In: *2018 Workshop on Metrology for Industry 4.0 and IoT*. 2018, pp. 7–11. DOI: 10.1109/METROI4.2018.8428320.

379. J. Polvi, T. Taketomi, G. Yamamoto, H. Kato, and M. Billinghurst. "Exploring the Use of Handheld Augmented Reality Authoring and Guidance in an Unprepared Indoor Environment".

380. I. Poupyrev, D. Tan, M. Billinghurst, H. Kato, H. Regenbrecht, and N. Tetsutani. "Tiles: A Mixed Reality Authoring Interface". In: *(2001) INTERACT 2001 Conference on Human Computer Interaction*. 2001.

381. F. B. Pressestelle. *Projekt Heb@AR: Hebammen trainieren Notfälle an der FH Bielefeld in „Augmented-Reality"*. https://www.fh-bielefeld.de/presse/pressemitteilungen/projekt-heb-ar-hebammen-trainieren-notfaelle-an-der-fh-bielefeld-in-augmented-reality [Accessed: 06.04.2023]. 2022.

382. A. Prouzeau, Y. Wang, B. Ens, W. Willett, and T. Dwyer. "Corsican Twin: Authoring In Situ Augmented Reality Visualisations in Virtual Reality". In: *Proceedings of the International Conference on Advanced Visual Interfaces*. AVI '20. Association for Computing Machinery, Salerno, Italy, 2020. ISBN: 9781450375351. DOI: 10.1145/3399715.3399743.

383. C. Putnam, M. Puthenmadom, M. A. Cuerdo, W. Wang, and N. Paul. "Adaptation of the System Usability Scale for User Testing with Children". In: *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI EA '20. Association for Computing Machinery, Honolulu, HI, USA, 2020, pp. 1–7. ISBN: 9781450368193. DOI: 10.1145/3334480.3382840.

384. P. Qu and J. Tanaka. "Spatial Note System Using Virtual Agent to Enhance Family Connection". Note: Full bibliographic details could not be retrieved (29.05.2023).

385. F. Quint, F. Loch, M. Orfgen, and D. Zuehlke. "A system architecture for assistance in manual tasks". In: *Intelligent Environments 2016*. IOS Press, 2016, pp. 43–52. DOI: 10.3233/978-1-61499-690-3-43.

386. E. Quintana and J. Favela. "Augmented reality annotations to assist persons with Alzheimers and their caregivers". *Personal and ubiquitous computing* 17:6, 2013, pp. 1105–1116. DOI: 10.1007/s00779-012-0558-6.

387. J. Quintero, S. Baldiris, R. Rubira, J. Cerón, and G. Velez. "Augmented reality in educational inclusion. A systematic review on the last decade". *Frontiers in psychology* 10, 2019, p. 1835. DOI: 10.3389/fpsyg.2019.01835.

388. I. Radu and B. MacIntyre. "Augmented-Reality Scratch: A Children's Authoring Environment for Augmented-Reality Experiences". In: *Proceedings of the 8th International Conference on Interaction Design and Children*. IDC '09. Association for Computing Machinery, Como, Italy, 2009, pp. 210–213. ISBN: 9781605583952. DOI: 10.1145/1551788.1551831.

389. A. S. M. M. Rahman, J. Cha, and A. El Saddik. "Authoring edutainment content through video annotations and 3D model augmentation". In: *2009 IEEE International Conference on Virtual Environments, Human-Computer Interfaces and Measurements Systems*. 2009, pp. 370–374. DOI: 10.1109/VECIMS.2009.5068927.

390. P. Ralph. "The illusion of requirements in software development". *Requirements Engineering* 18, 2013, pp. 293–296. DOI: 10.1007/s00766-012-0161-4.

391. H. Ramirez, E. G. Mendivil, P. R. Flores, and M. C. Gonzalez. "Authoring Software for Augmented Reality Applications for the Use of Maintenance and Training Process". *Procedia Computer Science* 25, 2013. 2013 International Conference on Virtual and Augmented Reality in Education, pp. 189–193. ISSN: 1877-0509. DOI: 10.1016/j.procs.2013.11.023.

392. I. M. Ratner and J. Harvey. "Vertical Slicing: Smaller is Better". In: *2011 Agile Conference*. 2011, pp. 240–245. DOI: 10.1109/AGILE.2011.46.

393. S. Rattanarungrot, M. White, and B. Jackson. "The application of service orientation on a mobile AR platform — a museum scenario". In: *2015 Digital Heritage*. Vol. 1. 2015, pp. 329–332. DOI: 10.1109/DigitalHeritage.2015.7413894.

394. G. Reinmann. "Innovation ohne Forschung? Ein Plädoyer für den Design-Based Research-Ansatz in der Lehr-Lernforschung". *Unterrichtswissenschaft* 33:1, 2005, pp. 52–69.

395. P. Reipschlager, T. Flemisch, and R. Dachselt. "Personal Augmented Reality for Information Visualization on Large Interactive Displays". *IEEE Transactions on Visualization and Computer Graphics* 27:2, 2021, pp. 1182–1192. DOI: 10.1109/TVCG.2020.3030460.

396. G. Reitmayr, E. Eade, and T. W. Drummond. "Semi-automatic Annotations in Unknown Environments". In: *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*. 2007, pp. 67–70. DOI: 10.1109/ISMAR.2007.4538827.

397. A. Riedl and A. Schelten. *Kompetenzentwicklung in Lernfeldern im Unterricht gewerblich-technischer Schulen*. Peter Lang, 2013. DOI: 10.3726/978-3-653-00180-8.

398. M. Rieser and S. Clark. "Locative media and situated learning". *Digital Creativity* 24:3, 2013, pp. 208–221. DOI: 10.1080/14626268.2013.813381.

399. M. Rimol DeLisi and C. Howley. *Gartner Forecasts Worldwide Low-Code Development Technologies Market to Grow 20% in 2023*. https://www.gartner.com/en/newsroom/press-releases/2022-12-13-gartner-forecasts-worldwide-low-code-development-technologies-market-to-grow-20-percent-in-2023. Accessed: 19.04.2023.

400. R. A. Roberto, J. P. Lima, R. C. Mota, and V. Teichrieb. "Authoring Tools for Augmented Reality: An Analysis and Classification of Content Design Tools". In: *Design, User Experience, and Usability: Technological Contexts*. Ed. by A. Marcus. Springer International Publishing, Cham, 2016, pp. 237–248. ISBN: 978-3-319-40406-6.

401. M. Rossi, G. D'Avenio, S. Morelli, and M. Grigioni. "CogAR: an augmented reality App to improve quality of life of the people with cognitive impairment". In: *2020 IEEE 20th Mediterranean Electrotechnical Conference (MELECON)*. 2020, pp. 339–343. DOI: 10.1109/MELECON48756.2020.9140554.

402. D. Rumiński and K. Walczak. "CARL: A Language for Modelling Contextual Augmented Reality Environments". In: *Technological Innovation for Collective Awareness Systems*. Ed. by L. M. Camarinha-Matos, N. S. Barrento, and R. Mendonça. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014, pp. 183–190. ISBN: 978-3-642-54734-8.

403. D. Rumiński and K. Walczak. "Creation of Interactive AR Content on Mobile Devices". In: *Business Information Systems Workshops*. Ed. by W. Abramowicz. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 258–269. ISBN: 978-3-642-41687-3. DOI: 10.1007/978-3-642-41687-3_24.

404. D. Rumiński and K. Walczak. "Dynamic composition of interactive AR scenes with the CARL language". In: *IISA 2014, The 5th International Conference on Information, Intelligence, Systems and Applications*. 2014, pp. 329–334. DOI: 10.1109/IISA.2014.6878808.

405. D. Rumiński and K. Walczak. "Semantic Model for Distributed Augmented Reality Services". In: *Proceedings of the 22nd International Conference on 3D Web Technology*. Web3D '17. Association for Computing Machinery, Brisbane, Queensland, Australia, 2017. ISBN: 9781450349550. DOI: 10.1145/3055624.3077121.

406. R. M. Ryan. *Intrinsic motivation inventory (IMI)*. https://selfdeterminationtheory.org/intrinsic-motivation-inventory/. Accessed: 15.08.2022.

407. R. M. Ryan and E. L. Deci. "Self-Determination Theory". In: *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)*. Ed. by J. D. Wright. 2nd ed. Elsevier, Oxford, 2015, pp. 486–491. ISBN: 978-0-08-097087-5. DOI: 10.1016/B978-0-08-097086-8.26036-4.

408. N. Saforrudin, H. Badioze Zaman, and A. Ahmad. "Technical Skills in Developing Augmented Reality Application: Teachers' Readiness". In: *Visual Informatics: Sustaining Research and Innovations*. Ed. by H. B. Zaman, P. Robinson, M. Petrou, P. Olivier, T. K. Shih, S. Velastin, and I. Nyström. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 360–370. ISBN: 978-3-642-25200-6. DOI: 10.1007/978-3-642-25200-6_34.

409. G. F. D. Salvador, P. J. Bota, V. Vinayagamoorthy, H. Plácido da Silva, and A. Fred. "Smartphone-Based Content Annotation for Ground Truth Collection in Affective Computing". In: *ACM International Conference on Interactive Media Experiences*. IMX '21. Association for Computing Machinery, Virtual Event, USA, 2021, pp. 199–204. ISBN: 9781450383899. DOI: 10.1145/3452918.3465505.

410. L. Sambrooks and B. Wilkinson. "Designing HARATIO: A Novice AR Authoring Tool". In: *Proceedings of the 28th Australian Conference on Computer-Human Interaction*. OzCHI '16. Association for Computing Machinery, Launceston, Tasmania, Australia, 2016, pp. 175–179. ISBN: 9781450346184. DOI: 10.1145/3010915.3011005.

411. A. Sanna, F. Manuri, G. Piumatti, G. Paravati, F. Lamberti, and P. Pezzolla. "A Flexible AR-based Training System for Industrial Maintenance". In: *Augmented and Virtual Reality*. Ed. by L. T. De Paolis and A. Mongelli. Springer International Publishing, Cham, 2015, pp. 314–331. ISBN: 978-3-319-22888-4. DOI: 10.1007/978-3-319-22888-4_23.

412. M. E. C. Santos, A. I. W. Luebke, M. M. T. Rodrigo, T. Taketomi, G. Yamamoto, C. Sandor, K. Hirokazu, et al. "Authoring augmented reality as situated multimedia". English. In: *Proceedings of the 22nd International Conference on Computers in Education (ICCE 2014)*. 22nd International Conference on Computers in Education (ICCE 2014) ; Conference date: 30-11-2014 Through 04-12-2014. Asia-Pacific Society for Computers in Education, 2014, pp. 554–556. ISBN: 978-4-9908014-1-0.

413. M. E. C. Santos, T. Taketomi, G. Yamamoto, M. M. T. Rodrigo, C. Sandor, H. Kato, et al. "Augmented reality as multimedia: the case for situated vocabulary learning". *Research and Practice in Technology Enhanced Learning* 11:1, 2016, pp. 1–23. DOI: 10.1186/s41039-016-0028-2.

414. M. E. C. Santos, T. Taketomi, C. Sandor, J. Polvi, G. Yamamoto, and H. Kato. "A Usability Scale for Handheld Augmented Reality". In: *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology*. VRST '14. Association for Computing Machinery, Edinburgh, Scotland, 2014, pp. 167–176. ISBN: 9781450332538. DOI: 10.1145/2671015.2671019.

415. J. Sauro. *Predicting Net Promoter Scores from System Usability Scale Scores*. 2012. URL: https://measuringu.com/nps-sus/ (visited on 01/12/2022).

416. J. Sauro and J. R. Lewis. *Quantifying the user experience: Practical statistics for user research*. Morgan Kaufmann, 2016.

417. D. Schmalstieg, A. Fuhrmann, G. Hesina, Z. Szalavári, L. M. Encarnaçao, M. Gervautz, and W. Purgathofer. "The studierstube augmented reality project". *Presence: Teleoperators & Virtual Environments* 11:1, 2002, pp. 33–54.

418. D. Schmalstieg, T. Langlotz, and M. Billinghurst. "Augmented Reality 2.0". In: *Virtual Realities: Dagstuhl Seminar 2008*. Ed. by G. Brunnett, S. Coquillart, and G. Welch. Springer Vienna, Vienna, 2011, pp. 13–37. ISBN: 978-3-211-99178-7. DOI: 10.1007/978-3-211-99178-7_2.

419. M. Schrepp, A. Hinderks, and J. Thomaschewski. "Construction of a Benchmark for the User Experience Questionnaire (UEQ)." *IJIMAI* 4:4, 2017, pp. 40–44. DOI: 10.9781/ijimai.2017.445.

420. M. Schrepp, J. Kollmorgen, A.-L. Meiners, A. Hinderks, D. Winter, H. B. Santoso, and J. Thomaschewski. "On the Importance of UX Quality Aspects for Different Product Categories". *International Journal of Interactive Multimedia and Artificial Intelligence* In Press:In Press, 2023, pp. 1–15. ISSN: 1989-1660. DOI: 10.9781ijjimai.2023.03.001.

421. G. Schubert, D. Schattel, M. Tönnis, G. Klinker, and F. Petzold. "Tangible Mixed Reality On-Site: Interactive Augmented Visualisations from Architectural Working Models in Urban Design". In: *Computer-Aided Architectural Design Futures. The Next City - New Technologies and the Future of the Built Environment*. Ed. by G. Celani, D. M. Sperling, and J. M. S. Franco. Springer Berlin Heidelberg, Berlin, Heidelberg, 2015, pp. 55–74. ISBN: 978-3-662-47386-3. DOI: 10.1007/978-3-662-47386-3_4.

422. H. Seichter, J. Looser, and M. Billinghurst. "ComposAR: An intuitive tool for authoring AR applications". In: *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*. 2008, pp. 177–178. DOI: 10.1109/ISMAR.2008.4637354.

423. D. Seo and B. Yoo. "Interoperable information model for geovisualization and interaction in XR environments". *International Journal of Geographical Information Science* 34:7, 2020, pp. 1323–1352. DOI: 10.1080/13658816.2019.1706739.

424. D. Seo, B. Yoo, and H. Ko. "Webizing Mixed Reality for Cooperative Augmentation of Life Experience". In: *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*. CSCW '16 Companion. Association for Computing Machinery, San Francisco, California, USA, 2016, pp. 401–404. ISBN: 9781450339506. DOI: 10.1145/2818052.2869078.

425. S. Seuring, M. Müller, M. Westhaus, and R. Morana. "Conducting a Literature Review — The Example of Sustainability in Supply Chains". In: *Research Methodologies in Supply Chain Management: In Collaboration with Magnus Westhaus*. Ed. by H. Kotzab, S. Seuring, M. Müller, and G. Reiner. Physica-Verlag HD, Heidelberg, 2005, pp. 91–106. ISBN: 978-3-7908-1636-5. DOI: 10.1007/3-7908-1636-1_7.

426. S. Z. Shariff, S. A. Bejaimal, J. M. Sontrop, A. V. Iansavichus, R. B. Haynes, M. A. Weir, and A. X. Garg. "Retrieving Clinical Evidence: A Comparison of PubMed and Google Scholar for Quick Clinical Searches". *J Med Internet Res* 15:8, 2013, e164. ISSN: 14388871. DOI: 10.2196/jmir.2624.

427. S. Shekhar, P. Maheshwari, M. J, A. Singhal, K. K. Singh, and K. Krishna. "ARComposer: Authoring Augmented Reality Experiences through Text". In: *The Adjunct Publication of the 32nd Annual ACM Symposium on User Interface Software and Technology*. UIST '19. Association for Computing Machinery, New Orleans, LA, USA, 2019, pp. 63–65. ISBN: 9781450368179. DOI: 10.1145/3332167.3357116.

428. J. Shim, M. Kong, Y. Yang, J. Seo, and T.-D. Han. "Interactive features based augmented reality authoring tool". In: *2014 IEEE International Conference on Consumer Electronics (ICCE)*. 2014, pp. 47–50. DOI: 10.1109/ICCE.2014.6775902.

429. J. Shim, J. Seo, and T.-d. Han. "MSL_AR Toolkit: AR Authoring Tool with Interactive Features". In: *Virtual and Mixed Reality - New Trends*. Ed. by R. Shumaker. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 105–112. ISBN: 978-3-642-22021-0. DOI: 10.1007/978-3-642-22021-0_13.

430. J. Shim, Y. Yang, N. Kang, J. Seo, and T.-D. Han. "Gesture-based interactive augmented reality content authoring system using HMD". *Virtual Reality* 20:1, 2016, pp. 57–69. DOI: 10.1007/s10055-016-0282-z.

431. C. Shin, S.-H. Hong, and H. Yoon. "Enriching Natural Monument with User-Generated Mobile Augmented Reality Mashup". *Journal of Multimedia Information System* 7:1, 2020, pp. 25–32. DOI: 10.33851/JMIS.2020.7.1.25.

432. C. Shin, H. Kim, C. Kang, Y. Jang, A. Choi, and W. Woo. "Unified Context-Aware Augmented Reality Application Framework for User-Driven Tour Guides". In: *2010 International Symposium on Ubiquitous Virtual Reality*. 2010, pp. 52–55. DOI: 10.1109/ISUVR.2010.22.

433. J. G. Shin, G. Ng, and D. Saakes. "Couples Designing Their Living Room Together: A Study with Collaborative Handheld Augmented Reality". In: *Proceedings of the 9th Augmented Human International Conference*. AH '18. Association for Computing Machinery, Seoul, Republic of Korea, 2018. ISBN: 9781450354158. DOI: 10.1145/3174910.3174930.

434. M. Shin, B.-s. Kim, and J. Park. "AR storyboard: an augmented reality based interactive storyboard authoring tool". In: *Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'05)*. 2005, pp. 198–199. DOI: 10.1109/ISMAR.2005.12.

435. B. Shneiderman, C. Plaisant, M. S. Cohen, S. Jacobs, N. Elmqvist, and N. Diakopoulos. *Designing the user interface: strategies for effective human-computer interaction*. Pearson, 2016.

436. R. Sicat, J. Li, J. Choi, M. Cordeil, W.-K. Jeong, B. Bach, and H. Pfister. "DXR: A Toolkit for Building Immersive Data Visualizations". *IEEE Transactions on Visualization and Computer Graphics* 25:1, 2019, pp. 715–725. DOI: 10.1109/TVCG.2018.2865152.

437. J. L. Siewert, M. Wolf, B. Böhm, and S. Thienhaus. "Usability Study for an Augmented Reality Content Management System". In: *Cross Reality and Data Science in Engineering*. Ed. by M. E. Auer and D. May. Springer International Publishing, Cham, 2021, pp. 274–287. ISBN: 978-3-030-52575-0. DOI: 10.1007/978-3-030-52575-0_22.

438. M. M. da Silva, J. M. X. Teixeira, P. S. Cavalcante, and V. Teichrieb. "Perspectives on how to evaluate augmented reality technology tools for education: a systematic review". *Journal of the Brazilian Computer Society* 25:1, 2019, pp. 1–18. DOI: 10.1186/s13173-019-0084-8.

439. A. M. Soaita, B. Serin, and J. Preece. "A methodological quest for systematic literature mapping". *International Journal of Housing Policy* 20:3, 2020, pp. 320–343.

440. B. Soedji, J. Lacoche, and E. Villain. "Creating AR Applications for the IOT: A New Pipeline". In: *26th ACM Symposium on Virtual Reality Software and Technology*. VRST '20. Association for Computing Machinery, Virtual Event, Canada, 2020. ISBN: 9781450376198. DOI: 10.1145/3385956.3422088.

441. F. Sorrentino and L. D. Spano. "Post-it notes: supporting teachers in authoring vocabulary game contents". *Multimedia Tools and Applications* 78:16, 2019, pp. 23049–23074. DOI: 10.1007/s11042-019-7604-6.

442. M. Speicher, K. Tenhaft, S. Heinen, and H. Handorf. "Enabling industry 4.0 with holobuilder". *INFORMATIK 2015*, 2015.

443. R. Stallman. "Technology should help us share, not constrain us". *The Guardian*, 2012. https://www.theguardian.com/technology/2012/apr/17/technology-help-us-share-not-constrain (Accessed: 28.06.2023).

444. J. W. Stigler and J. Hiebert. *The teaching gap: Best ideas from the world's teachers for improving education in the classroom*. New York: The Free Press, 1999.

445. V. Stirbu, Y. You, K. Roimela, and V.-V. Mattila. "A Lightweight Platform for Web Mashups in Immersive Mirror Worlds". *IEEE Pervasive Computing* 12:1, 2013, pp. 34–41. DOI: 10.1109/MPRV.2012.60.

446. I. Stock and M. Weber. "Authoring Technical Documentation Using a Generic Document Model". In: *Proceedings of the 24th Annual ACM International Conference on Design of Communication*. SIGDOC '06. Association for Computing Machinery, Myrtle Beach, SC, USA, 2006, pp. 172–179. ISBN: 1595935231. DOI: 10.1145/1166324.1166364.

447. I. Stock, M. Weber, and E. Steinmeier. "Metadata Based Authoring for Technical Documentation". In: *Proceedings of the 23rd Annual International Conference on Design of Communication: Documenting & Designing for Pervasive Information*. SIGDOC '05. Association for Computing Machinery, Coventry, United Kingdom, 2005, pp. 60–67. ISBN: 1595931759. DOI: 10.1145/1085313.1085330.

448. G. A. Straka and G. Macke. *Lern-lehr-theoretische Didaktik*. Waxmann Verlag, 2002.

449. I. E. Sutherland. "A Head-Mounted Three Dimensional Display". In: *Seminal Graphics: Pioneering Efforts That Shaped the Field*. Association for Computing Machinery, New York, NY, USA, 1968, pp. 295–302. ISBN: 158113052X. DOI: 10.1145/280811.281016.

450. R. Suzuki, R. H. Kazi, L.-y. Wei, S. DiVerdi, W. Li, and D. Leithinger. "RealitySketch: Embedding Responsive Graphics and Visualizations in AR through Dynamic Sketching". In: *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. UIST '20. Association for Computing Machinery, Virtual Event, USA, 2020, pp. 166–181. ISBN: 9781450375146. DOI: 10.1145/3379337.3415892.

451. R. Tache, H. A. Abeykoon, K. T. Karunanayaka, J. P. Kumarasinghe, G. Roth, O. N. N. Fernando, and A. D. Cheok. "Command Center: Authoring tool to supervise augmented reality session". In: *2012 IEEE Virtual Reality Workshops (VRW)*. 2012, pp. 99–100. DOI: 10.1109/VR.2012.6180901.
452. K. Tainaka, Y. Fujimoto, M. Kanbara, H. Kato, A. Moteki, K. Kuraki, K. Osamura, T. Yoshitake, and T. Fukuoka. "Guideline and Tool for Designing an Assembly Task Support System Using Augmented Reality". In: *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 2020, pp. 486–497. DOI: 10.1109/ISMAR50242.2020.00077.
453. K. Tainaka, Y. Fujimoto, T. Sawabe, M. Kanbara, and H. Kato. "Selection framework of visualization methods in designing AR industrial task-support systems". *Computers in Industry* 145, 2023, p. 103828. ISSN: 0166-3615. DOI: 10.1016/j.compind.2022.103828.
454. F. Tang, D. R. Tretter, and Q. Lin. "TouchPaper: An Augmented Reality Application with Cloud-Based Image Recognition Service". In: *Proceedings of the 4th International Conference on Internet Multimedia Computing and Service*. ICIMCS '12. Association for Computing Machinery, Wuhan, China, 2012, pp. 7–10. ISBN: 9781450316002. DOI: 10.1145/2382336.2382339.
455. J. K. T. Tang, T.-Y. A. Duong, Y.-W. Ng, and H.-K. Luk. "Learning to create 3D models via an augmented reality smartphone interface". In: *2015 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*. 2015, pp. 236–241. DOI: 10.1109/TALE.2015.7386050.
456. J. K. Tang, W.-M. Lau, K.-K. Chan, and K.-H. To. "AR interior designer: Automatic furniture arrangement using spatial and functional relationships". In: *2014 International Conference on Virtual Systems & Multimedia (VSMM)*. 2014, pp. 345–352. DOI: 10.1109/VSMM.2014.7136652.
457. D. Tarasienko. *Mesh Combiner for Unity*. https://github.com/dawid-t/Mesh-Combiner. Accessed: 02.05.2023, GitHub repository. 2019.
458. A. Terenzi and G. Terenzi. "Towards Augmented Reality Design: the case for the ARplugins". Note: Full bibliographic details could not be retrieved (29.05.2023).
459. H. Theil. *Statistical Decomposition Analysis: With Applications in the Social and Administrative Sciences*. English. BibTeX package created from National Library of Australia Catalogue. North-Holland Pub. Co.; American Elsevier Pub. Co, Amsterdam, New York, 1972, pp. xvi, 337. ISBN: 0444103783.
460. M. Tomlein and K. Grønbæk. "Augmented Reality Supported Modeling of Industrial Systems to Infer Software Configuration". *Proc. ACM Hum.-Comput. Interact.* 2:EICS, 2018. DOI: 10.1145/3229087.

461. A. C. Tricco, E. Lillie, W. Zarin, K. K. O'Brien, H. Colquhoun, D. Levac, D. Moher, M. D. Peters, T. Horsley, L. Weeks, et al. "PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation". *Annals of Internal Medicine* 169:7, 2018. PMID: 30178033, pp. 467–473. DOI: 10.7326/M18-0850.

462. P. Y. Tseng, H. Haraldsson, and S. Belongie. "Annotate All! A Perspective Preserved Asynchronous Annotation System for Collaborative Augmented Reality", 2019.

463. T. S. Tullis and J. N. Stetson. "A comparison of questionnaires for assessing website usability". In: *Usability professional association conference*. Vol. 1. Minneapolis, USA. 2004, pp. 1–12.

464. M. Turner, B. Kitchenham, P. Brereton, S. Charters, and D. Budgen. "Does the technology acceptance model predict actual use? A systematic literature review". *Information and Software Technology* 52:5, 2010, pp. 463–479. ISSN: 0950-5849. DOI: 10.1016/j.infsof.2009.11.005.

465. J. F. Ty, M. M. T. Rodrigo, and M. E. C. Santos. "A Mobile Authoring Tool for Augmented Reality Content Generation Using Images as Annotations". Note: Full bibliographic details could not be retrieved (29.05.2023).

466. S. Tzima, G. Styliaras, and A. Bassounas. "Augmented Reality Applications in Education: Teachers Point of View". *Education Sciences* 9:2, 2019. ISSN: 2227-7102. DOI: 10.3390/educsci9020099.

467. Unity Technologies. *Unity Visual Scripting Package*. https://docs.unity3d.com/Packages/com.unity.visualscripting@1.8/manual/index.html. Accessed: 2023-02-13.

468. Unity Technologies. *Visual Scripting: Graph Types*. Unity Online Documentation. https://docs.unity3d.com/Packages/com.unity.visualscripting@1.8/manual/vs-graph-types.html. Accessed: 02.05.2023. Unity Technologies, 2023.

469. D. H. (e. V. *Zahlenspiegel zur Situation der Hebammen 11/2021*. https://www.unsere-hebammen.de/w/files/kampagnenmaterial/202111-zahlenspiegel-zur-situation-der-hebammen.pdf (Accessed: 25.06.2023). 2021.

470. K. van Lopik, M. Sinclair, R. Sharpe, P. Conway, and A. West. "Developing augmented reality capabilities for industry 4.0 small enterprises: Lessons learnt from a content authoring case study". *Computers in Industry* 117, 2020, p. 103208. ISSN: 0166-3615. DOI: 10.1016/j.compind.2020.103208.

471. J. J. Van Merriënboer, R. E. Clark, and M. B. De Croock. "Blueprints for complex learning: The 4C/ID-model". *Educational technology research and development* 50:2, 2002, pp. 39–61. DOI: 10.1007/BF02504993.

472. A. Vargas González, S. Koh, K. Kapalo, R. Sottilare, P. Garrity, M. Billinghurst, and J. LaViola. "A Comparison of Desktop and Augmented Reality Scenario Based Training Authoring Tools". In: *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 2019, pp. 339–350. DOI: 10.1109/ISMAR.2019.00032.

473. F. Vera and J. A. Sánchez. "A Model for In-Situ Augmented Reality Content Creation Based on Storytelling and Gamification". In: *Proceedings of the 6th Mexican Conference on Human-Computer Interaction*. MexIHC'16. Association for Computing Machinery, Colima, Mexico, 2016, pp. 39–42. ISBN: 9781450347822. DOI: 10.1145/2967175.2967385.

474. F. Vera and J. A. Sánchez. "SITUAR: A platform for in-situ augmented reality content creation". *Avances en Interacción Humano-Computadora* 0:1, 2016, pp. 90–92. ISSN: 2594-2352. DOI: 10.47756/aihc.y1i1.28.

475. A. Villanueva, Z. Zhu, Z. Liu, K. Peppler, T. Redick, and K. Ramani. "Meta-AR-App: An Authoring Platform for Collaborative Augmented Reality in STEM Classrooms". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI '20. Association for Computing Machinery, Honolulu, HI, USA, 2020, pp. 1–14. ISBN: 9781450367080. DOI: 10.1145/3313831.3376146.

476. A. Visser. "Survey of XML Languages for Augmented Reality Content". In: *Proc. of AR Standardization Forum, Barcelona*. 2011.

477. A. Vitzthum. "SSIML/AR: A Visual Language for the Abstract Specification of Augmented Reality User Interfaces". In: *3D User Interfaces (3DUI'06)*. 2006, pp. 135–142. DOI: 10.1109/VR.2006.123.

478. A. Vitzthum. "Entwicklungsunterstützung für interaktive 3D-Anwendungen". PhD thesis. Ludwig Maximilian University of Munich, 2008.

479. A. Vitzthum and H. Hussmann. "Modeling Augmented Reality User Interfaces with SSIML/AR." *Journal of Multimedia* 1:3, 2006, pp. 13–22.

480. J. Vogel and O. Thomas. "Generating Smart Glasses-based Information Systems with BPMN4SGA: A BPMN Extension for Smart Glasses Applications", 2019. Note: Full bibliographic details could not be retrieved (29.05.2023).

481. K. Vogel, A. Bernloehr, C. Lewa, J. Blattgerste, M. Joswig, T. Schäfer, T. Pfeiffer, and N. Bauer. "Augmented Reality based training for student midwives (Heb@AR) – what kind of support do teachers need?" In: *6th International Conference of the German Association of Midwifery Science (DGHWi)*. Winterthur, Switzerland, 28, 2022. DOI: 10.3205/22dghwi25.

482. S. Waffenschmidt, M. Knelangen, W. Sieben, S. Bühn, and D. Pieper. "Single screening versus conventional double screening for study selection in systematic reviews: a methodological systematic review". *BMC medical research methodology* 19:1, 2019, pp. 1–9. DOI: 10.1186/s12874-019-0782-0.

483. K. Walczak and W. Cellary. "X-VRML for advanced virtual reality applications". *Computer* 36:3, 2003, pp. 89–92. DOI: 10.1109/MC.2003.1185226.

484. K. Walczak and W. Cellary. "Building Database Applications of Virtual Reality with X-VRML". In: *Proceedings of the Seventh International Conference on 3D Web Technology*. Web3D '02. Association for Computing Machinery, Tempe, Arizona, USA, 2002, pp. 111–120. ISBN: 1581134681. DOI: 10.1145/504502.504521.

485. K. Walczak, W. Wiza, R. Wojciechowski, A. Wójtowicz, D. Rumiński, and W. Cellary. "Building Augmented Reality Presentations with Web 2.0 Tools". In: *International Joint Conference*. Ed. by Á. Herrero, B. Baruque, J. Sedano, H. Quintián, and E. Corchado. Springer International Publishing, Cham, 2015, pp. 595–605. ISBN: 978-3-319-19713-5. DOI: 10.1007/978-3-319-19713-5_52.

486. K. Walczak and R. Wojciechowski. "Dynamic Creation of Interactive Mixed Reality Presentations". In: *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*. VRST '05. Association for Computing Machinery, Monterey, CA, USA, 2005, pp. 167–176. ISBN: 1595930981. DOI: 10.1145/1101616.1101653.

487. W. H. Walters. "Google Scholar coverage of a multidisciplinary field". *Information Processing & Management* 43:4, 2007, pp. 1121–1132. ISSN: 0306-4573. DOI: 10.1016/j.ipm.2006.08.006.

488. J. Wang, C. Zeng, Y. Liu, and S. Li. "Integrated content authoring for augmented reality based product manual assembly process instruction". In: *Proceedings of International Conference on Computers and Industrial Engineering, CIE*. Vol. 1. 2013, pp. 242–251.

489. M.-J. Wang, C.-H. Tseng, and C.-Y. Shen. "An Easy to Use Augmented Reality Authoring Tool for Use in Examination Purpose". In: *Human-Computer Interaction*. Ed. by P. Forbrig, F. Paternó, and A. Mark Pejtersen. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, pp. 285–288. ISBN: 978-3-642-15231-3. DOI: 10.1007/978-3-642-15231-3_31.

490. Y. Wang, A. Anne, and T. Ropp. "Applying the technology acceptance model to understand aviation students' perceptions toward augmented reality maintenance training instruction". *International Journal of Aviation, Aeronautics, and Aerospace* 3:4, 2016, p. 3. DOI: 10.15394/ijaaa.2016.1144.

491. Y. Wang, T. Langlotz, M. Billinghurst, and T. Bell. "An authoring tool for mobile phone AR environments". In: *Proceedings of New Zealand Computer Science Research Student Conference*. Vol. 9. 2009, pp. 1–4.

492. Z. Wang, S. Qiu, Q. Chen, N. Trayan, A. Ringlein, J. Dorsey, and H. Rushmeier. "AniCode: authoring coded artifacts for network-free personalized animations". *The Visual Computer* 35:6, 2019, pp. 885–897. DOI: 10.1007/s00371-019-01681-y.

493. Z. Wang, T. Nayfeh, J. Tetzlaff, P. O'Blenis, and M. H. Murad. "Error rates of human reviewers during abstract screening in systematic reviews". *PLOS ONE* 15:1, 2020, pp. 1–8. DOI: 10.1371/journal.pone.0227742.

494. S. Webel, U. Bockholt, T. Engelke, N. Gavish, M. Olbrich, and C. Preusche. "An augmented reality training platform for assembly and maintenance skills". *Robotics and Autonomous Systems* 61:4, 2013. Models and Technologies for Multi-modal Skill Training, pp. 398–403. ISSN: 0921-8890. DOI: 10.1016/j.robot.2012.09.013.

495. C. Weßel. "Semi-strukturierte Interviews im Software-Engineering: Indikationsstellung, Vorbereitung, Durchführung und Auswertung–Ein Fall-basiertes Tutorium". *INFORMATIK 2010. Service Science–Neue Perspektiven für die Informatik. Band 2*, 2010, pp. 927–937. ISSN: 1617-5468.

496. J. Westlin and T. H. Laine. "Short paper: Calory Battle AR: An extensible mobile augmented reality exergame platform". In: *2014 IEEE World Forum on Internet of Things (WF-IoT)*. 2014, pp. 171–172. DOI: 10.1109/WF-IoT.2014.6803144.

497. R. Wetzel, I. Lindt, A. Waern, and S. Johnson. "The Magic Lens Box: Simplifying the Development of Mixed Reality Games". In: *Proceedings of the 3rd International Conference on Digital Interactive Media in Entertainment and Arts*. DIMEA '08. Association for Computing Machinery, Athens, Greece, 2008, pp. 479–486. ISBN: 9781605582481. DOI: 10.1145/1413634.1413719.

498. M. White, F. Liarokapis, N. Mourkoussis, A. Basu, J. Darcy, P. Petridis, M. Sifniotis, and P. Lister. "ARCOLite-an XML based system for building and presenting virtual museum exhibitions using Web3D and augmented reality". In: *Proceedings Theory and Practice of Computer Graphics, 2004.* 2004, pp. 94–101. DOI: 10.1109/TPCG.2004.1314458.

499. S. White, D. Feng, and S. Feiner. "Interaction and presentation techniques for shake menus in tangible augmented reality". In: *2009 8th IEEE International Symposium on Mixed and Augmented Reality*. 2009, pp. 39–48. DOI: 10.1109/ISMAR.2009.5336500.

500. M. Whitlock, G. Fitzmaurice, T. Grossman, and J. Matejka. "AuthAR: concurrent authoring of tutorials for AR assembly guidance", 2020.

501. M. Whitlock, J. Mitchell, N. Pfeufer, B. Arnot, R. Craig, B. Wilson, B. Chung, and D. A. Szafir. "MRCAT: In Situ Prototyping of Interactive AR Environments". In: *Virtual, Augmented and Mixed Reality. Design and Interaction*. Ed. by J. Y. C. Chen and G. Fragomeni. Springer International Publishing, Cham, 2020, pp. 235–255. ISBN: 978-3-030-49695-1. DOI: 10.1007/978-3-030-49695-1_16.

502. G. Wiegand, C. Mai, K. Holländer, and H. Hussmann. "InCarAR: A Design Space Towards 3D Augmented Reality Applications in Vehicles". In: *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. AutomotiveUI '19. Association for Computing Machinery, Utrecht, Netherlands, 2019, pp. 1–13. ISBN: 9781450368841. DOI: 10.1145/3342197.3344539.

503. J. Willis. "A Recursive, Reflective Instructional Design Model Based on Constructivist-Interpretivist Theory". *Educational Technology* 35:6, 1995, pp. 5–23. ISSN: 00131962.

504. C. Wohlin. "Guidelines for Snowballing in Systematic Literature Studies and a Replication in Software Engineering". In: *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*. EASE '14. Association for Computing Machinery, London, England, United Kingdom, 2014. ISBN: 9781450324762. DOI: 10.1145/2601248.2601268.

505. C. Wohlin. "Second-Generation Systematic Literature Studies Using Snowballing". In: *Proceedings of the 20th International Conference on Evaluation and Assessment in Software Engineering*. EASE '16. Association for Computing Machinery, Limerick, Ireland, 2016. ISBN: 9781450336918. DOI: 10.1145/2915970.2916006.

506. C. Wohlin, M. Kalinowski, K. Romero Felizardo, and E. Mendes. "Successful combination of database search and snowballing for identification of primary studies in systematic literature studies". *Information and Software Technology* 147, 2022, p. 106908. ISSN: 0950-5849. DOI: 10.1016/j.infsof.2022.106908.

507. R. Wojciechowski and W. Cellary. "Evaluation of learners' attitude toward learning in ARIES augmented reality environments". *Computers & Education* 68, 2013, pp. 570–585. ISSN: 0360-1315. DOI: 10.1016/j.compedu.2013.02.014.

508. R. Wojciechowski, K. Walczak, M. White, and W. Cellary. "Building Virtual and Augmented Reality Museum Exhibitions". In: *Proceedings of the Ninth International Conference on 3D Web Technology*. Web3D '04. Association for Computing Machinery, Monterey, California, 2004, pp. 135–144. ISBN: 1581138458. DOI: 10.1145/985040.985060.

509. D. Wolf, D. Besserer, K. Sejunaite, M. Riepe, and E. Rukzio. "CARe: An Augmented Reality Support System for Dementia Patients". In: *The 31st Annual ACM Symposium on User Interface Software and Technology Adjunct Proceedings*. UIST '18 Adjunct. Association for Computing Machinery, Berlin, Germany, 2018, pp. 42–44. ISBN: 9781450359498. DOI: 10.1145/3266037.3266095.

510. H.-K. Wu, S. W.-Y. Lee, H.-Y. Chang, and J.-C. Liang. "Current status, opportunities and challenges of augmented reality in education". *Computers & Education* 62, 2013, pp. 41–49. ISSN: 0360-1315. DOI: https://doi.org/10.1016/j.compedu.2012.10.024.

511. Y. Xiao and M. Watson. "Guidance on Conducting a Systematic Literature Review". *Journal of Planning Education and Research* 39:1, 2019, pp. 93–112. DOI: 10.1177/0739456X17723971.

512. H. S. Yang. "Realistic E-Learning System Based on Mixed Reality". In: *Proceedings of the 8th International Conference on Virtual Reality Continuum and Its Applications in Industry*. VRCAI '09. Association for Computing Machinery, Yokohama, Japan, 2009, p. 10. ISBN: 9781605589121. DOI: 10.1145/1670252.1670255.

513. Y. Yang, J. Shim, S. Chae, and T.-D. Han. "Interactive Augmented Reality Authoring System using mobile device as input method". In: *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 2016, pp. 001429–001432. DOI: 10.1109/SMC.2016.7844437.

514. Y. Yang, J. Shim, S. Chae, and T.-D. Han. "Mobile Augmented Reality Authoring Tool". In: *2016 IEEE Tenth International Conference on Semantic Computing (ICSC)*. 2016, pp. 358–361. DOI: 10.1109/ICSC.2016.42.

515. H. Ye, K. C. Kwan, W. Su, and H. Fu. "ARAnimator: In-Situ Character Animation in Mobile AR with User-Defined Motion Gestures". *ACM Trans. Graph.* 39:4, 2020. ISSN: 0730-0301. DOI: 10.1145/3386569.3392404.

516. R. M. Yılmaz and Y. Göktaş. "Using augmented reality technology in education". *Çukurova Üniversitesi Eğitim Fakültesi Dergisi* 47:2, 2018, pp. 510–537. ISSN: 1302-9967. DOI: 10.14812/cuefd.376066.

517. H. K. Yoo and J. W. Lee. "Mobile augmented reality system for in-situ 3D modeling and authoring". In: *2014 International Conference on Big Data and Smart Computing (BIGCOMP)*. 2014, pp. 282–285. DOI: 10.1109/BIGCOMP.2014.6741453.

518. Y. You and V.-V. Mattila. "Visualizing Web Mash-Ups for in-Situ Vision-Based Mobile AR Applications". In: *Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication*. UbiComp '13 Adjunct. Association for Computing Machinery, Zurich, Switzerland, 2013, pp. 271–274. ISBN: 9781450322157. DOI: 10.1145/2494091.2494177.

519. J. Yu, J. Jeon, J. Park, G. Park, H.-i. Kim, and W. Woo. "Geometry-Aware Interactive AR Authoring Using a Smartphone in a Wearable AR Environment". In: *Distributed, Ambient and Pervasive Interactions*. Ed. by N. Streitz and P. Markopoulos. Springer International Publishing, Cham, 2017, pp. 416–424. ISBN: 978-3-319-58697-7. DOI: 10.1007/978-3-319-58697-7_31.

520. Y.-T. Yue, Y.-L. Yang, G. Ren, and W. Wang. "SceneCtrl: Mixed Reality Enhancement via Efficient Scene Editing". In: *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*. UIST '17. Association for Computing Machinery, Québec City, QC, Canada, 2017, pp. 427–436. ISBN: 9781450349819. DOI: 10.1145/3126594.3126601.

521. M. Ez-zaouia, I. Marfisi-Schottman, M. Oueslati, C. Mercier, A. Karoui, and S. George. "A Design Space of Educational Authoring Tools for Augmented Reality". In: *International Conference on Games and Learning Alliance*. Ed. by K. Kiili, K. Antti, F. de Rosa, M. Dindar, M. Kickmeier-Rust, and F. Bellotti. Springer International Publishing, Cham, 2022, pp. 258–268. ISBN: 978-3-031-22124-8. DOI: 10.1007/978-3-031-22124-8_25.

522. J. Zauner, M. Haller, A. Brandl, and W. Hartman. "Authoring of a mixed reality assembly instructor for hierarchical structures". In: *The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003. Proceedings.* 2003, pp. 237–246. DOI: 10.1109/ISMAR.2003.1240707.

523. J. Zauner and M. Haller. "Authoring of Mixed Reality Applications Including Multi-Marker Calibration for Mobile Devices". In: *Proceedings of the Tenth Eurographics Conference on Virtual Environments*. EGVE'04. Eurographics Association, Grenoble, France, 2004, pp. 87–90. ISBN: 390567310X.

524. J. Zhang, T.-C. Liu, Y.-T. Sung, and K.-E. Chang. "Using Augmented Reality to Promote Homogeneity in Learning Achievement". In: *2015 IEEE International Symposium on Mixed and Augmented Reality - Media, Art, Social Science, Humanities and Design*. 2015, pp. 1–5. DOI: 10.1109/ISMAR-MASHD.2015.17.

525. L. Zheng, X. Liu, Z. An, S. Li, and R. Zhang. "A smart assistance system for cable assembly by combining wearable augmented reality with portable visual inspection". *Virtual Reality & Intelligent Hardware* 2:1, 2020, pp. 12–27. ISSN: 2096-5796. DOI: 10.1016/j.vrih.2019.12.002.

526. F. Zhou, H. B.-L. Duh, and M. Billinghurst. "Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR". In: *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*. 2008, pp. 193–202. DOI: 10.1109/ISMAR.2008.4637362.

527. X. Zhou, Y. Jin, H. Zhang, S. Li, and X. Huang. "A Map of Threats to Validity of Systematic Literature Reviews in Software Engineering". In: *2016 23rd Asia-Pacific Software Engineering Conference (APSEC)*. 2016, pp. 153–160. DOI: 10.1109/APSEC.2016.031.

528. Y. Zhou, H. Zhang, X. Huang, S. Yang, M. A. Babar, and H. Tang. "Quality Assessment of Systematic Reviews in Software Engineering: A Tertiary Study". In: *Proceedings of the 19th International Conference on Evaluation and Assessment in Software Engineering*. EASE '15. Association for Computing Machinery, Nanjing, China, 2015. ISBN: 9781450333504. DOI: 10.1145/2745802.2745815.

529. E. Zhu, A. Lilienthal, L. A. Shluzas, I. Masiello, and N. Zary. "Design of Mobile Augmented Reality in Health Care Education: A Theory-Driven Framework". *JMIR Medical Education* 1:2, 2015, e10. ISSN: 2369-3762. DOI: 10.2196/mededu.4443.

530. J. Zhu, S. Ong, and A. Nee. "A context-aware augmented reality system to assist the maintenance operators". *International Journal on Interactive Design and Manufacturing (IJIDeM)* 8:4, 2014, pp. 293–304. DOI: 10.1007/s12008-013-0199-7.

531. J. Zhu, S. Ong, and A. Nee. "A context-aware augmented reality assisted maintenance system". *International Journal of Computer Integrated Manufacturing* 28:2, 2015, pp. 213–225. DOI: 10.1080/0951192X.2013.874589.

532. J. Zhu, S. K. Ong, and A. Y. Nee. “An authorable context-aware augmented reality system to assist the maintenance technicians”. *The International Journal of Advanced Manufacturing Technology* 66:9, 2013, pp. 1699–1714.

533. S. Zollmann, C. Hoppe, S. Kluckner, C. Poglitsch, H. Bischof, and G. Reitmayr. “Augmented Reality for Construction Site Monitoring and Documentation”. *Proceedings of the IEEE* 102:2, 2014, pp. 137–154. DOI: 10.1109/JPROC.2013.2294314.

534. J. Zubizarreta, I. Aguinaga, and A. Amundarain. “A framework for augmented reality guidance in industry”. *The International Journal of Advanced Manufacturing Technology* 102:9, 2019, pp. 4095–4108. DOI: 10.1007/s00170-019-03527-2.

535. J. Zumbach et al. *Problem-Based Learning in Österreich: eine Bestandsaufnahme*. na, 2012.

536. F. Zünd, M. Ryffel, S. Magnenat, A. Marra, M. Nitti, M. Kapadia, G. Noris, K. Mitchell, M. Gross, and R. W. Sumner. “Augmented Creativity: Bridging the Real and Virtual Worlds to Enhance Creative Play”. In: *SIGGRAPH Asia 2015 Mobile Graphics and Interactive Applications*. SA ’15. Association for Computing Machinery, Kobe, Japan, 2015. ISBN: 9781450339285. DOI: 10.1145/2818427.2818460.

537. S. Zychlinski. *Dython: A set of data tools in Python*. https://github.com/shakedzy/dython (Accessed: 16.06.2023). 2023.

538. S. Zychlinski. *The Search for Categorical Correlation*. https://towardsdatascience.com/the-search-for-categorical-correlation-a1cf7f1888c9 (Accessed: 16.06.2023). 2018.

Appendix

Additional information and supporting materials

The Appendix of this thesis contains the PRISMA 2020 checklist, PRISMA ScR 2018 checklist, the results of the mapping study for each dimension in table format, all SUS results of this thesis from the SUS Analysis Toolkit in table format, and stateflow descriptions used during the TrainAR Authoring Tool evaluation.

1 PRISMA 2020 Checklist

Section	Item	PRISMA Checklist Item	Section
TITLE			
Title	1	Identify the report as a systematic review.	2
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	Abstract
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	2, 2.1
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	2.2
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	2.3.4
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	2.3.1, 2.3.3, 2.3.6
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	2.3, 2.3.2, 2.3.6
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process	2.3.5
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	N/A
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	2.5
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	2.5
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	2.3.4
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	N/A
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	2.3.4, 2.5
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	2.5
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	2.5
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	2.5
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	2.5

continued on the next page...

...continued from the previous page

Section	Item	PRISMA Checklist Item	Section
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	N/A
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	2.4.2, 2.4.3
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	2.4.2, 2.4.3
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	2.3.3, 2.3.7, 2.4.1
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	2.4.3
Study characteristics	17	Cite each included study and present its characteristics.	2.6
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	N/A
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	N/A
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	N/A
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	N/A
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	N/A
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results	N/A
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	N/A
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	2.4.2
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	2.4.3, 2.7, 2.9
	23b	Discuss any limitations of the evidence included in the review.	2.4, 2.8
	23c	Discuss any limitations of the review processes used.	2.4.1
	23d	Discuss implications of the results for practice, policy, and future research.	2.7, 2.9
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	2.4.1
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	2.4.1
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	2.4.1
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	1.2.7
Competing interests	26	Declare any competing interests of review authors.	N/A
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	2.6

Table 1: The official PRISMA checklist, transferred from [361].

2 PRISMA-ScR 2018 Checklist

Section	Item	PRISMA-ScR Checklist Item	Section
TITLE			
Title	1	Identify the report as a scoping review.	2
ABSTRACT			
Structured summary	2	Provide a structured summary that includes (as applicable): background, objectives, eligibility criteria, sources of evidence, charting methods, results, and conclusions that relate to the review questions and objectives.	Abstract
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known. Explain why the review questions/objectives lend themselves to a scoping review approach.	2.2, 2.5
Objectives	4	Provide an explicit statement of the questions and objectives being addressed with reference to their key elements (e.g., population or participants, concepts, and context) or other relevant key elements used to conceptualize the review questions and/or objectives.	2.2
METHODS			
Protocol and registration	5	Indicate whether a review protocol exists; state if and where it can be accessed (e.g., a Web address); and if available, provide registration information, including the registration number.	2.4.1
Eligibility criteria	6	Specify characteristics of the sources of evidence used as eligibility criteria (e.g., years considered, language, and publication status), and provide a rationale.	2.3.4
Information sources	7	Describe all information sources in the search (e.g., databases with dates of coverage and contact with authors to identify additional sources), as well as the date the most recent search was executed.	2.3.1, 2.3.6
Search	8	Present the full electronic search strategy for at least 1 database, including any limits used, such that it could be repeated.	2.3
Selection of sources of evidence	9	State the process for selecting sources of evidence (i.e.,screening and eligibility) included in the scoping review.	2.3.4
Data charting process	10	Describe the methods of charting data from the included sources of evidence (e.g., calibrated forms or forms that have been tested by the team before their use, and whether data charting was done independently or in duplicate) and any processes for obtaining and confirming data from investigators.	2.5
Data items	11	List and define all variables for which data were sought and any assumptions and simplifications made.	2.5
Critical appraisal of individual sources of evidence	12	If done, provide a rationale for conducting a critical appraisal of included sources of evidence; describe the methods used and how this information was used in any data synthesis (if appropriate).	N/A
Synthesis of results	13	Describe the methods of handling and summarizing the data that were charted.	2.6
RESULTS			
Selection of sources of evidence	14	Give numbers of sources of evidence screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally using a flow diagram.	2.3.3, 2.3.7, 2.4.1
Characteristics of sources of evidence	15	For each source of evidence, present characteristics for which data were charted and provide the citations.	2.6
Critical appraisal within sources of evidence	16	If done, present data on critical appraisal of included sources of evidence (see item 12).	N/A
Results of individual sources of evidence	17	For each included source of evidence, present the relevant data that were charted that relate to the review questions and objectives.	2.6
Synthesis of results	18	Summarize and/or present the charting results as they relate to the review questions and objectives.	2.7

continued on the next page...

...continued from the previous page

Section	Item	PRISMA-ScR Checklist Item	Section
DISCUSSION			
Summary of evidence	19	Summarize the main results (including an overview of concepts, themes, and types of evidence available), link to the review questions and objectives, and consider the relevance to key groups.	2.7
Limitations	20	Discuss the limitations of the scoping review process.	2.4.1, 2.4.2. 2.8
Conclusions	21	Provide a general interpretation of the results with respect to the review questions and objectives, as well as potential implications and/or next steps.	2.7, 2.9
FUNDING			
Funding	22	Describe sources of funding for the included sources of evidence, as well as sources of funding for the scoping review. Describe the role of the funders of the scoping review.	1.2.7

Table 2: The official PRISMA-ScR checklist, transferred from [461].

3 Systematic Review — Mapping Result Tables

Search Source	Publications
DB Search (128)	[34], [214], [430], [39], [25], [358], [19], [134], [139], [235], [428], [10], [497], [136], [222], [93], [234], [313], [262], [249], [29], [44], [178], [367], [182], [473], [410], [315], [81], [273], [212], [474], [379], [252], [198], [251], [522], [519], [343], [365], [513], [77], [164], [267], [304], [13], [348], [40], [492], [47], [189], [534], [532], [446], [531], [268], [272], [86], [76], [141], [296], [447], [107], [108], [422], [427], [518], [219], [514], [517], [179], [396], [250], [160], [266], [43], [154], [499], [364], [366], [31], [507], [132], [181], [211], [42], [429], [82], [326], [294], [284], [73], [524], [475], [84], [412], [489], [98], [335], [441], [92], [33], [97], [237], [355], [125], [391], [196], [220], [472], [344], [190], [345], [508], [349], [241], [265], [203], [398], [78], [434], [488], [166], [163], [525], [187], [127], [58]
FS1 (62)	[306], [89], [80], [373], [71], [325], [159], [261], [404], [303], [333], [63], [2], [436], [498], [354], [111], [269], [414], [175], [195], [351], [431], [130], [403], [336], [231], [370], [117], [500], [501], [346], [168], [201], [281], [331], [302], [382], [21], [350], [311], [486], [450], [465], [16], [88], [15], [393], [485], [437], [342], [137], [470], [377], [352], [239], [95], [36], [290], [258], [144], [452]
BS1 (37)	[388], [225], [523], [380], [386], [509], [161], [180], [491], [121], [23], [3], [536], [255], [254], [297], [183], [288], [338], [17], [357], [103], [356], [157], [126], [477], [192], [292], [119], [451], [512], [455], [247], [30], [307], [1], [378]
FS2 (27)	[533], [318], [162], [401], [368], [75], [215], [259], [445], [271], [5], [202], [253], [200], [245], [48], [433], [384], [462], [171], [372], [341], [328], [128], [87], [329], [460]
FS3 (19)	[395], [442], [424], [515], [151], [199], [291], [421], [440], [193], [32], [148], [270], [213], [375], [411], [8], [385], [280]
BS2 (14)	[26], [324], [456], [186], [496], [520], [256], [376], [79], [67], [494], [530], [298], [283]
BS3 (5)	[330], [223], [204], [305], [432]
FS4 (1)	[480]

Table 3: The 293 publications of the literature review, grouped by the search strategy (database search, forward snowballing, backward snowballing) and iteration, they were found with.

Publication Year	Publications
2020 (31)	[395], [89], [330], [401], [515], [431], [130], [151], [199], [291], [501], [440], [164], [346], [168], [171], [382], [82], [450], [475], [437], [329], [137], [470], [78], [163], [36], [290], [258], [280], [452]
2019 (31)	[261], [245], [175], [81], [273], [384], [500], [462], [77], [13], [492], [47], [32], [67], [331], [534], [427], [154], [8], [15], [441], [196], [472], [190], [349], [241], [377], [95], [525], [480], [58]
2016 (27)	[34], [430], [318], [358], [324], [215], [303], [424], [473], [410], [48], [315], [351], [212], [474], [365], [513], [204], [356], [341], [76], [21], [514], [250], [247], [345], [385]
2018 (24)	[214], [225], [509], [19], [136], [121], [436], [433], [198], [343], [117], [17], [193], [189], [281], [364], [294], [88], [335], [92], [344], [460], [378], [144]
2013 (22)	[306], [80], [386], [259], [333], [253], [403], [231], [267], [494], [305], [375], [518], [31], [507], [292], [284], [355], [391], [398], [488], [166]
2015 (21)	[25], [139], [159], [442], [536], [182], [336], [338], [421], [348], [531], [73], [411], [524], [455], [16], [393], [485], [97], [342], [125]
2014 (21)	[533], [428], [456], [75], [404], [5], [269], [414], [496], [379], [357], [148], [530], [213], [517], [311], [128], [465], [412], [98], [30]
2012 (15)	[373], [134], [445], [354], [186], [255], [223], [251], [256], [376], [532], [219], [119], [451], [84]
2017 (15)	[71], [325], [222], [93], [23], [200], [370], [519], [520], [79], [304], [326], [33], [203], [127]
2004 (13)	[523], [313], [2], [498], [3], [522], [296], [266], [43], [328], [283], [508], [1]
2009 (12)	[388], [39], [235], [491], [234], [111], [367], [268], [499], [42], [239], [307]
2005 (10)	[271], [262], [183], [103], [447], [192], [486], [237], [352], [434]
2010 (10)	[195], [254], [40], [302], [126], [86], [160], [181], [489], [432]
2011 (9)	[10], [297], [252], [288], [272], [366], [132], [211], [429]
2006 (8)	[368], [29], [44], [201], [446], [477], [179], [350]
2007 (7)	[180], [202], [372], [396], [512], [87], [265]
2008 (6)	[497], [63], [249], [270], [422], [220]
2003 (5)	[162], [178], [141], [107], [108]
2000 (2)	[26], [161]
2001 (2)	[380], [298]
2002 (2)	[157], [187]

Table 4: The 293 publications of the literature review, grouped by the year they were published in.

Number of Authors	Publications
4 (70)	[34], [89], [26], [318], [225], [162], [358], [330], [401], [161], [456], [325], [75], [497], [442], [445], [303], [491], [313], [23], [253], [515], [223], [254], [351], [288], [151], [522], [370], [520], [338], [365], [513], [79], [500], [171], [372], [305], [76], [296], [219], [514], [266], [154], [364], [211], [42], [292], [326], [119], [524], [455], [16], [247], [92], [30], [97], [87], [437], [237], [391], [220], [344], [345], [508], [488], [166], [290], [1], [258]
3 (65)	[395], [214], [25], [134], [93], [333], [234], [5], [63], [2], [249], [3], [424], [48], [433], [431], [183], [252], [231], [117], [199], [462], [357], [440], [13], [40], [346], [356], [157], [32], [189], [331], [534], [532], [530], [213], [531], [272], [447], [375], [422], [192], [179], [396], [250], [160], [350], [499], [132], [311], [429], [294], [465], [489], [8], [393], [329], [342], [196], [190], [265], [203], [434], [58], [280]
2 (63)	[306], [388], [39], [80], [523], [324], [386], [19], [235], [10], [180], [159], [368], [136], [222], [215], [404], [271], [262], [29], [178], [354], [111], [367], [186], [473], [200], [245], [410], [496], [474], [384], [403], [198], [376], [291], [77], [267], [168], [47], [341], [270], [302], [446], [268], [298], [518], [517], [31], [507], [284], [73], [486], [84], [441], [355], [349], [377], [352], [398], [187], [480], [460]
5 (47)	[430], [373], [509], [139], [428], [261], [202], [269], [182], [255], [175], [195], [297], [130], [379], [343], [17], [256], [204], [421], [348], [103], [201], [148], [382], [21], [107], [108], [427], [43], [328], [82], [128], [512], [88], [15], [335], [33], [137], [470], [95], [78], [385], [525], [36], [127], [144]
6 (25)	[533], [380], [121], [44], [414], [81], [273], [212], [251], [519], [193], [67], [281], [494], [86], [411], [450], [475], [98], [485], [125], [241], [432], [307], [378]
7 (11)	[436], [164], [304], [492], [126], [181], [451], [412], [472], [239], [163]
1 (7)	[71], [259], [315], [336], [477], [141], [366]
8 (3)	[498], [501], [283]
10 (1)	[536]
9 (1)	[452]

Table 5: The 293 publications of the literature review, grouped by the number of authors who contributed to the publication.

Publication Format	Publications
Proceedings (123)	[306], [533], [26], [225], [523], [380], [161], [139], [235], [456], [159], [497], [222], [442], [261], [259], [404], [93], [303], [333], [234], [271], [313], [63], [262], [2], [498], [202], [367], [186], [536], [253], [255], [414], [200], [245], [48], [433], [175], [351], [297], [273], [384], [183], [403], [522], [370], [519], [343], [117], [520], [17], [256], [376], [199], [500], [501], [421], [77], [13], [348], [372], [193], [157], [32], [281], [331], [341], [302], [446], [382], [477], [21], [296], [447], [375], [107], [108], [298], [192], [160], [266], [350], [499], [364], [328], [42], [429], [82], [411], [119], [512], [486], [450], [475], [84], [455], [465], [8], [16], [88], [98], [30], [485], [33], [87], [437], [237], [125], [472], [508], [265], [352], [239], [488], [385], [187], [480], [36], [1], [58], [258], [452]
Academic Journal (78)	[395], [34], [430], [39], [25], [80], [373], [318], [162], [386], [75], [136], [215], [445], [5], [436], [44], [178], [182], [515], [254], [81], [431], [212], [130], [336], [231], [151], [251], [291], [164], [304], [346], [168], [171], [492], [103], [356], [47], [67], [189], [534], [532], [270], [530], [213], [531], [268], [272], [76], [141], [366], [507], [311], [326], [128], [294], [284], [15], [335], [247], [441], [92], [97], [329], [137], [355], [470], [190], [349], [241], [398], [166], [163], [525], [460], [127], [144]
Proceedings (Short) (63)	[89], [388], [358], [330], [324], [401], [19], [134], [71], [428], [10], [325], [368], [491], [249], [29], [23], [3], [354], [269], [473], [410], [223], [315], [288], [198], [338], [365], [513], [79], [204], [462], [357], [267], [40], [148], [494], [305], [126], [514], [517], [396], [154], [132], [181], [211], [292], [73], [283], [524], [489], [393], [342], [391], [220], [344], [95], [203], [432], [78], [307], [378], [280]
Proceedings (Poster) (29)	[214], [509], [180], [121], [111], [424], [496], [195], [474], [379], [252], [440], [201], [86], [422], [427], [518], [219], [179], [250], [43], [31], [451], [412], [196], [345], [377], [434], [290]

Table 6: The 293 publications of the literature review, grouped by publication format, they were published in.

Publisher (>5 Publications)	Publications
IEEE (83)	[395], [533], [324], [401], [19], [235], [428], [10], [456], [261], [215], [259], [404], [93], [445], [63], [262], [436], [249], [498], [496], [315], [195], [254], [351], [297], [81], [273], [522], [370], [376], [199], [513], [79], [204], [77], [13], [348], [40], [372], [201], [281], [126], [477], [86], [375], [107], [108], [298], [422], [514], [517], [179], [396], [266], [499], [132], [181], [211], [42], [292], [128], [73], [451], [524], [84], [455], [393], [98], [335], [237], [220], [472], [203], [432], [78], [434], [187], [290], [1], [378], [258], [452]
ACM (70)	[214], [89], [388], [26], [225], [509], [161], [134], [71], [139], [497], [2], [121], [354], [111], [186], [536], [182], [424], [253], [255], [473], [414], [200], [245], [410], [48], [433], [515], [223], [175], [343], [117], [520], [338], [365], [17], [256], [357], [440], [267], [157], [446], [382], [21], [296], [447], [427], [518], [192], [219], [43], [350], [154], [31], [82], [119], [512], [486], [450], [475], [125], [196], [344], [345], [508], [377], [239], [460], [58]
Springer (53)	[34], [430], [162], [386], [159], [368], [333], [234], [271], [3], [367], [403], [336], [231], [198], [151], [251], [519], [462], [501], [421], [164], [171], [492], [32], [67], [534], [532], [270], [530], [213], [141], [160], [364], [366], [429], [411], [489], [8], [16], [247], [441], [92], [30], [485], [33], [87], [437], [349], [241], [166], [36], [144]
Others (52)	[306], [39], [25], [523], [380], [325], [180], [222], [442], [303], [491], [313], [29], [23], [202], [269], [431], [212], [474], [384], [183], [130], [379], [252], [288], [291], [500], [346], [168], [356], [148], [341], [302], [76], [250], [328], [311], [294], [284], [283], [465], [412], [88], [97], [190], [265], [352], [95], [307], [488], [385], [480]
Elsevier (28)	[80], [373], [358], [330], [136], [304], [103], [193], [47], [189], [331], [494], [305], [268], [272], [507], [326], [15], [329], [342], [137], [355], [391], [470], [163], [525], [127], [280]
Taylor & Francis (7)	[318], [75], [5], [44], [178], [531], [398]

Table 7: The 293 publications of the literature review, grouped by Publisher they were published with. Publishers with less than 5 publications published in them are grouped in "Others".

Contribution	Publications
Main Contribution (252)	[306], [34], [214], [89], [430], [388], [39], [25], [533], [26], [80], [318], [225], [162], [523], [330], [324], [380], [386], [401], [509], [19], [161], [134], [71], [139], [235], [428], [10], [456], [325], [180], [159], [368], [75], [497], [222], [442], [261], [259], [404], [93], [303], [491], [333], [313], [63], [262], [2], [121], [436], [249], [29], [178], [498], [202], [3], [354], [111], [367], [186], [269], [182], [424], [253], [255], [473], [200], [245], [410], [48], [433], [515], [223], [175], [496], [195], [254], [351], [297], [81], [273], [431], [212], [474], [384], [183], [130], [379], [252], [288], [403], [336], [231], [198], [151], [251], [522], [370], [519], [343], [117], [520], [338], [365], [17], [256], [376], [199], [513], [79], [204], [291], [500], [462], [501], [421], [357], [440], [77], [164], [267], [304], [13], [348], [40], [346], [168], [171], [492], [103], [356], [372], [193], [157], [47], [32], [67], [201], [189], [281], [331], [532], [302], [213], [446], [305], [382], [268], [272], [126], [477], [86], [76], [141], [21], [296], [375], [107], [108], [298], [422], [427], [518], [192], [219], [514], [517], [179], [396], [250], [160], [266], [43], [350], [154], [364], [366], [31], [328], [132], [181], [211], [42], [292], [311], [429], [82], [326], [284], [73], [411], [119], [451], [512], [486], [283], [524], [450], [475], [84], [455], [465], [412], [489], [8], [88], [335], [247], [92], [30], [485], [33], [97], [87], [437], [329], [342], [237], [391], [470], [196], [220], [472], [344], [190], [241], [352], [239], [95], [203], [432], [78], [307], [434], [488], [166], [385], [163], [187], [480], [36], [290], [460], [127], [1], [378], [58], [258], [144], [280], [452]
Side Contribution (41)	[395], [373], [358], [136], [215], [445], [234], [271], [5], [23], [44], [536], [414], [315], [148], [494], [341], [534], [270], [530], [531], [447], [499], [507], [128], [294], [16], [15], [393], [98], [441], [137], [355], [125], [345], [508], [349], [377], [265], [398], [525]

Table 8: The 293 publications of the literature review, grouped by whether the AR authoring tool is main contribution, the main or a side contribution of the paper.

Deployment Purpose	Publications
Assistance (69)	[25], [533], [386], [401], [509], [442], [261], [215], [445], [333], [313], [23], [269], [414], [315], [254], [351], [379], [288], [151], [522], [338], [376], [79], [500], [462], [440], [492], [47], [189], [281], [331], [341], [534], [532], [270], [302], [530], [446], [305], [531], [382], [477], [447], [375], [179], [329], [342], [137], [237], [355], [125], [344], [190], [345], [307], [488], [166], [385], [525], [187], [480], [36], [290], [460], [127], [1], [144], [452]
Entertainment (62)	[395], [214], [89], [388], [523], [19], [134], [325], [159], [368], [497], [136], [222], [303], [234], [436], [29], [498], [3], [111], [367], [186], [536], [182], [424], [255], [473], [410], [515], [223], [175], [496], [195], [297], [81], [431], [212], [474], [384], [403], [336], [198], [519], [343], [357], [77], [164], [267], [304], [348], [192], [364], [455], [15], [349], [241], [377], [239], [95], [432], [398], [78]
Multipurpose (58)	[34], [358], [161], [235], [428], [259], [404], [93], [491], [5], [262], [2], [249], [44], [178], [354], [183], [252], [251], [117], [513], [13], [40], [346], [168], [103], [268], [86], [76], [141], [21], [107], [108], [298], [422], [427], [518], [219], [514], [517], [396], [250], [160], [266], [43], [499], [31], [311], [82], [486], [220], [508], [265], [352], [203], [163], [58], [280]
Prototyping (53)	[39], [26], [80], [373], [318], [225], [162], [330], [324], [380], [71], [139], [10], [456], [180], [75], [271], [63], [202], [253], [200], [245], [48], [433], [273], [130], [231], [370], [520], [365], [17], [256], [199], [204], [291], [501], [421], [171], [356], [372], [193], [157], [32], [67], [201], [148], [213], [272], [296], [350], [154], [366], [434]
Learning (51)	[306], [430], [121], [494], [126], [507], [328], [132], [181], [211], [42], [292], [429], [326], [128], [294], [284], [73], [411], [119], [451], [512], [283], [524], [450], [475], [84], [465], [412], [489], [8], [16], [88], [393], [98], [335], [247], [441], [92], [30], [485], [33], [97], [87], [437], [391], [470], [196], [472], [378], [258]

Table 9: The 293 publications of the literature review, grouped by the deployment purpose.

Deployment Context	Publications
Industrial Assembly, Maintenance & Machine Operation (60)	[306], [358], [442], [333], [414], [288], [376], [492], [356], [47], [281], [148], [331], [494], [341], [534], [532], [270], [302], [530], [446], [305], [531], [382], [126], [477], [447], [375], [311], [411], [87], [437], [329], [342], [137], [237], [355], [125], [391], [470], [196], [344], [190], [345], [307], [488], [166], [385], [163], [525], [187], [36], [290], [460], [127], [1], [378], [258], [144], [452]
No specific context (51)	[162], [161], [235], [428], [261], [259], [404], [93], [491], [271], [5], [313], [262], [2], [44], [178], [354], [269], [212], [183], [379], [252], [251], [117], [338], [513], [77], [13], [346], [168], [268], [86], [76], [21], [107], [108], [422], [518], [192], [219], [514], [517], [396], [250], [160], [266], [43], [154], [499], [31], [480]
(All levels of) Education (31)	[430], [507], [328], [132], [211], [42], [429], [326], [294], [284], [73], [119], [512], [283], [524], [450], [475], [84], [465], [412], [489], [8], [88], [98], [335], [247], [441], [92], [30], [33], [97]
Cultural Heritages & Museum Exhibits (25)	[34], [523], [136], [222], [445], [498], [186], [431], [403], [501], [164], [103], [298], [486], [393], [508], [349], [241], [377], [265], [352], [239], [95], [432], [398]
Game Development (23)	[214], [388], [225], [19], [497], [303], [63], [29], [536], [182], [410], [175], [273], [474], [519], [343], [357], [304], [427], [364], [366], [16], [15]
Furniture Design, Planning and Assembly (19)	[373], [318], [71], [139], [456], [75], [200], [433], [231], [522], [370], [520], [365], [256], [204], [291], [500], [171], [280]
Construction, Factory & Urban Planning (15)	[39], [25], [533], [80], [10], [130], [151], [421], [267], [372], [193], [157], [189], [213], [203]
Interactive Books & Stories (13)	[89], [134], [159], [368], [234], [3], [111], [367], [336], [198], [181], [485], [472]
Design Exploration (12)	[26], [330], [380], [180], [253], [245], [17], [199], [32], [67], [272], [296]
Navigation & Object Annotation (12)	[215], [23], [315], [254], [351], [297], [384], [79], [462], [440], [141], [179]
Healthcare, Elderly & Disabled People (8)	[386], [401], [509], [121], [48], [292], [128], [58]
Location-based Ads, Content & Games (7)	[424], [473], [496], [195], [348], [40], [78]
Creative Content Creation (6)	[324], [325], [515], [223], [350], [455]
Film Production (5)	[249], [202], [255], [201], [434]
Data Visualization (4)	[395], [436], [81], [82]
Military Planning (2)	[451], [220]

Table 10: The 293 publications of the literature review, grouped by context they were deployed in.

Usability Evaluation	Publications
None (188)	[395], [34], [214], [89], [430], [388], [39], [25], [533], [26], [162], [523], [330], [324], [380], [401], [509], [19], [161], [134], [71], [139], [235], [10], [325], [180], [368], [497], [222], [442], [261], [259], [404], [445], [491], [234], [271], [5], [313], [63], [2], [436], [249], [29], [23], [44], [178], [498], [202], [3], [354], [111], [367], [536], [269], [424], [473], [410], [223], [496], [195], [351], [297], [273], [431], [212], [474], [183], [379], [252], [288], [403], [231], [198], [251], [522], [519], [343], [365], [376], [204], [500], [462], [421], [440], [77], [164], [267], [40], [168], [171], [492], [103], [356], [372], [193], [157], [47], [32], [201], [281], [148], [331], [341], [534], [532], [270], [302], [530], [213], [446], [305], [272], [126], [477], [86], [141], [21], [296], [447], [375], [107], [298], [422], [518], [192], [219], [517], [179], [396], [250], [160], [43], [499], [364], [366], [31], [328], [181], [211], [42], [292], [311], [429], [294], [73], [451], [512], [486], [283], [450], [412], [16], [393], [30], [485], [33], [87], [329], [342], [237], [355], [391], [196], [220], [190], [345], [508], [377], [265], [352], [239], [203], [432], [398], [78], [307], [434], [488], [385], [163], [525], [187], [480], [290], [1], [378], [280]
Authoring (59)	[80], [373], [318], [225], [428], [456], [159], [75], [262], [186], [253], [414], [200], [245], [48], [433], [515], [175], [81], [130], [151], [370], [117], [520], [338], [17], [256], [199], [513], [79], [291], [501], [304], [13], [348], [346], [67], [531], [268], [76], [108], [427], [514], [266], [350], [154], [132], [82], [84], [455], [465], [88], [92], [472], [349], [36], [460], [258], [144]
Usage (29)	[306], [358], [386], [136], [215], [303], [333], [121], [255], [315], [254], [336], [189], [494], [507], [128], [284], [411], [119], [524], [489], [8], [15], [247], [437], [137], [344], [241], [95]
Both (17)	[93], [182], [384], [357], [382], [326], [475], [98], [335], [441], [97], [125], [470], [166], [127], [58], [452]

Table 11: The 293 publications of the literature review, grouped by which part of the process was evaluated by the authors of the publication in terms of usability.

Availability	Publications
Not available (272)	[306], [34], [214], [89], [430], [39], [25], [533], [26], [80], [373], [318], [162], [358], [523], [330], [324], [380], [386], [401], [509], [19], [161], [134], [71], [235], [428], [10], [456], [325], [180], [159], [368], [75], [497], [136], [222], [442], [261], [215], [404], [93], [445], [303], [491], [333], [234], [271], [5], [313], [63], [121], [249], [29], [23], [44], [178], [498], [202], [3], [354], [111], [367], [536], [269], [182], [424], [253], [255], [473], [414], [200], [245], [410], [48], [433], [515], [223], [175], [496], [315], [195], [254], [351], [297], [81], [273], [431], [212], [474], [384], [130], [379], [252], [288], [403], [336], [231], [198], [151], [251], [522], [370], [519], [343], [117], [520], [338], [365], [17], [256], [376], [199], [513], [79], [204], [291], [500], [462], [501], [421], [357], [440], [77], [164], [267], [304], [348], [40], [168], [103], [356], [372], [193], [157], [47], [32], [67], [201], [281], [148], [331], [494], [341], [534], [532], [270], [302], [530], [213], [446], [305], [531], [382], [268], [272], [126], [477], [86], [76], [141], [21], [447], [375], [107], [108], [298], [422], [427], [518], [192], [219], [514], [517], [179], [396], [250], [160], [266], [43], [350], [154], [499], [364], [366], [31], [507], [132], [181], [211], [42], [292], [311], [429], [128], [284], [73], [411], [119], [451], [512], [486], [283], [524], [450], [475], [84], [455], [465], [412], [489], [8], [16], [88], [15], [393], [98], [335], [247], [441], [92], [30], [485], [33], [97], [87], [437], [329], [342], [137], [237], [355], [125], [391], [470], [220], [472], [344], [190], [345], [508], [349], [241], [377], [265], [352], [239], [95], [203], [432], [398], [78], [307], [434], [488], [166], [385], [163], [525], [187], [480], [36], [290], [460], [127], [1], [378], [58], [258], [144], [280], [452]
Open Source (13)	[395], [388], [225], [139], [436], [186], [13], [346], [171], [492], [189], [82], [196]
Available (8)	[259], [262], [2], [183], [296], [328], [326], [294]

Table 12: The 293 publications of the literature review, grouped by whether they are available as binaries or even open source.

Construct Author	Publications
Endusers (62)	[376], [351], [223], [198], [182], [199], [462], [291], [256], [515], [175], [79], [78], [117], [273], [520], [40], [496], [357], [235], [368], [204], [350], [212], [370], [325], [159], [442], [428], [456], [178], [10], [491], [252], [384], [431], [315], [474], [473], [424], [433], [255], [48], [231], [253], [336], [414], [195], [297], [251], [130], [245], [518], [269], [254], [492], [134], [410], [13], [346], [71], [338]
Not specified (61)	[375], [154], [517], [513], [514], [139], [472], [328], [271], [265], [385], [352], [342], [39], [262], [119], [234], [160], [111], [272], [43], [499], [427], [531], [219], [396], [237], [429], [126], [21], [44], [488], [166], [367], [534], [298], [86], [108], [249], [532], [422], [250], [379], [19], [127], [179], [31], [348], [87], [125], [192], [267], [378], [203], [121], [47], [34], [76], [183], [164], [302]
Designer (52)	[324], [440], [382], [17], [365], [366], [77], [519], [497], [341], [222], [82], [23], [318], [75], [29], [266], [380], [180], [421], [364], [200], [303], [261], [162], [356], [354], [313], [331], [307], [373], [268], [80], [161], [26], [296], [508], [32], [288], [67], [95], [330], [280], [168], [92], [81], [523], [452], [358], [225], [477], [304]
Teachers & Trainer (39)	[73], [42], [58], [181], [88], [283], [507], [220], [284], [412], [306], [344], [451], [211], [292], [512], [132], [294], [128], [465], [8], [455], [16], [489], [524], [441], [485], [335], [33], [98], [15], [97], [247], [437], [475], [196], [326], [411], [30]
Domain Experts (18)	[470], [258], [501], [500], [404], [480], [533], [2], [141], [1], [290], [460], [391], [36], [270], [345], [522], [190]
Engineers (15)	[144], [447], [215], [355], [157], [25], [525], [305], [163], [329], [189], [193], [213], [372], [281]
Curators (11)	[498], [486], [239], [398], [377], [241], [136], [103], [403], [393], [349]
Developers (8)	[93], [445], [259], [3], [107], [436], [171], [63]
Parents & Caregivers (4)	[509], [430], [401], [386]
Writers (4)	[333], [434], [187], [446]
Others (3)	[395], [148], [151]
Enduser & Consumer (3)	[343], [311], [432]
Operators (3)	[530], [494], [137]
Programmers (3)	[5], [186], [536]
Children (3)	[214], [89], [388]
Students (2)	[450], [84]
Directors (2)	[202], [201]

Table 13: The 293 publications of the literature review, grouped by the (described or inferrable) intended author of the AR construct.

Authoring Hardware	Publications
Desktop (137)	[306], [34], [39], [26], [80], [373], [318], [225], [162], [358], [386], [401], [161], [134], [71], [139], [235], [428], [10], [456], [325], [180], [159], [368], [75], [497], [136], [222], [442], [261], [215], [259], [404], [93], [445], [303], [491], [333], [234], [271], [5], [313], [63], [262], [2], [121], [436], [249], [29], [23], [44], [178], [498], [202], [3], [354], [111], [367], [186], [356], [372], [193], [47], [201], [281], [148], [331], [494], [341], [534], [532], [270], [302], [530], [213], [446], [305], [268], [272], [126], [477], [86], [76], [141], [21], [296], [447], [375], [107], [108], [298], [422], [507], [328], [132], [181], [211], [42], [292], [311], [429], [82], [326], [128], [294], [284], [73], [411], [119], [451], [512], [486], [283], [329], [342], [137], [237], [391], [470], [196], [508], [349], [241], [377], [265], [352], [239], [307], [434], [488], [166], [385], [163], [525], [187], [480], [36], [182], [330], [509], [355], [472], [523], [536], [269], [524], [220], [95], [290]
Handheld (74)	[214], [89], [25], [533], [19], [424], [253], [255], [473], [414], [200], [245], [410], [48], [433], [515], [223], [175], [496], [315], [195], [254], [351], [297], [81], [273], [431], [212], [474], [384], [183], [130], [379], [252], [288], [403], [336], [231], [198], [151], [251], [522], [32], [67], [189], [427], [518], [192], [219], [514], [517], [179], [396], [250], [160], [450], [475], [84], [455], [465], [412], [489], [8], [16], [344], [190], [345], [203], [432], [398], [460], [127], [1], [378], [437], [88], [370], [182], [330], [523], [536], [269], [524], [220], [95], [290]
Head-mounted (31)	[395], [430], [324], [380], [519], [343], [117], [520], [338], [365], [17], [256], [376], [199], [513], [79], [204], [291], [500], [462], [501], [157], [531], [266], [43], [350], [154], [499], [364], [58], [258], [88], [182], [330], [509], [355], [472]
Web (27)	[388], [267], [304], [13], [348], [40], [346], [168], [171], [492], [103], [31], [15], [393], [98], [335], [247], [441], [92], [30], [485], [33], [97], [87], [125], [280], [452], [437]
VR (6)	[440], [77], [164], [382], [366], [78]
Projector (3)	[421], [357], [144], [370]

Table 14: The 293 publications of the literature review, grouped by the hardware supported for the authoring. As some AR authoring tools support multiple hardware platforms, this includes overlap.

Markup Notation	Publications
None / Not Specified (183)	[395], [214], [89], [430], [388], [25], [533], [26], [373], [318], [225], [162], [358], [523], [330], [324], [380], [386], [401], [509], [19], [161], [71], [235], [428], [10], [456], [180], [159], [368], [75], [261], [215], [303], [121], [249], [29], [23], [44], [3], [111], [186], [536], [182], [253], [255], [473], [414], [200], [245], [48], [433], [515], [223], [175], [496], [315], [254], [351], [81], [273], [431], [212], [474], [384], [130], [379], [252], [288], [336], [198], [151], [370], [519], [343], [117], [520], [338], [365], [17], [256], [376], [199], [513], [79], [204], [291], [462], [421], [357], [77], [164], [304], [348], [346], [168], [492], [356], [193], [157], [47], [32], [148], [494], [534], [532], [270], [213], [382], [272], [76], [141], [296], [375], [298], [422], [427], [518], [192], [219], [514], [517], [179], [396], [160], [266], [43], [350], [154], [499], [364], [366], [31], [132], [211], [42], [311], [326], [128], [284], [73], [524], [450], [84], [455], [465], [412], [489], [88], [15], [393], [98], [335], [247], [97], [87], [355], [125], [391], [196], [220], [472], [344], [345], [349], [241], [377], [239], [203], [432], [398], [78], [434], [488], [290], [460], [127], [378], [58], [258], [144], [280], [452]
XML (90)	[306], [34], [134], [139], [325], [497], [136], [222], [442], [259], [404], [93], [445], [491], [333], [234], [271], [5], [313], [63], [262], [2], [178], [498], [202], [354], [367], [424], [195], [297], [183], [403], [231], [251], [522], [500], [501], [440], [40], [103], [372], [201], [281], [331], [341], [302], [530], [446], [305], [531], [268], [126], [477], [86], [21], [447], [107], [108], [250], [507], [328], [181], [292], [411], [451], [512], [486], [283], [475], [16], [441], [92], [30], [485], [33], [329], [237], [190], [508], [265], [352], [95], [307], [166], [385], [163], [525], [187], [480], [1]
JSON (12)	[436], [269], [267], [13], [171], [189], [82], [294], [8], [437], [470], [36]
Propriatary Configuration Format (7)	[39], [80], [410], [429], [119], [342], [137]
CSV (1)	[67]

Table 15: The 293 publications of the literature review, grouped by the markup notation used.

Modularity	Publications
Standalone (268)	[395], [306], [34], [214], [89], [430], [388], [39], [25], [533], [80], [373], [318], [225], [358], [523], [330], [324], [380], [386], [509], [19], [161], [71], [139], [235], [428], [10], [456], [325], [180], [159], [368], [75], [497], [136], [222], [442], [261], [215], [259], [404], [93], [445], [303], [491], [333], [234], [271], [5], [313], [63], [262], [2], [121], [29], [23], [178], [498], [202], [354], [111], [367], [186], [536], [269], [182], [424], [253], [255], [414], [200], [245], [410], [48], [433], [515], [223], [175], [496], [315], [195], [254], [351], [297], [81], [273], [431], [212], [384], [183], [130], [379], [252], [288], [403], [336], [231], [198], [151], [251], [522], [370], [519], [343], [117], [520], [338], [365], [17], [256], [376], [199], [513], [79], [204], [291], [500], [462], [501], [421], [357], [440], [267], [304], [13], [348], [40], [346], [168], [171], [492], [103], [356], [193], [157], [47], [32], [67], [201], [189], [281], [148], [494], [341], [534], [532], [270], [302], [530], [213], [446], [305], [531], [382], [268], [272], [477], [86], [141], [21], [447], [375], [107], [108], [298], [422], [427], [518], [192], [219], [514], [517], [179], [396], [250], [160], [266], [350], [154], [499], [364], [366], [31], [507], [328], [132], [181], [211], [42], [292], [311], [429], [128], [294], [284], [73], [411], [119], [451], [512], [486], [283], [524], [450], [475], [84], [455], [465], [412], [489], [8], [16], [88], [15], [393], [98], [335], [247], [441], [92], [30], [485], [33], [97], [437], [329], [342], [137], [237], [355], [125], [391], [196], [220], [472], [344], [190], [345], [508], [349], [241], [265], [352], [239], [95], [203], [432], [398], [78], [307], [434], [488], [166], [385], [163], [525], [36], [290], [460], [127], [1], [378], [58], [258], [144], [280], [452]
Plugin (24)	[26], [162], [401], [134], [436], [249], [44], [3], [473], [474], [164], [372], [331], [126], [76], [296], [43], [82], [326], [87], [470], [377], [187], [480]
Both (1)	[77]

Table 16: The 293 publications of the literature review, grouped by whether they are standalone applications or plugins for host software.

Scene Preview	Publications
AR Scene Preview (223)	[395], [306], [214], [89], [430], [388], [25], [533], [26], [80], [373], [318], [225], [162], [358], [523], [330], [324], [380], [386], [401], [509], [19], [161], [134], [71], [139], [235], [428], [10], [456], [325], [180], [159], [368], [75], [136], [442], [261], [445], [303], [491], [313], [2], [121], [436], [249], [29], [44], [178], [202], [3], [354], [111], [367], [186], [536], [269], [182], [424], [253], [255], [473], [414], [200], [245], [410], [48], [433], [515], [223], [175], [315], [254], [351], [297], [81], [273], [431], [212], [474], [384], [183], [130], [379], [252], [288], [403], [336], [231], [198], [151], [251], [522], [370], [519], [343], [117], [520], [338], [365], [17], [256], [199], [513], [79], [204], [291], [500], [462], [501], [421], [357], [440], [77], [164], [346], [168], [492], [103], [356], [372], [193], [157], [47], [32], [67], [201], [189], [281], [148], [331], [494], [534], [532], [270], [213], [305], [531], [382], [268], [272], [126], [86], [76], [296], [108], [298], [422], [427], [518], [192], [219], [514], [517], [179], [396], [250], [160], [266], [43], [350], [154], [499], [364], [366], [31], [132], [181], [211], [42], [292], [311], [82], [128], [73], [451], [524], [450], [475], [84], [455], [465], [412], [489], [16], [88], [92], [30], [33], [437], [137], [237], [355], [470], [196], [220], [472], [344], [190], [345], [508], [377], [239], [95], [432], [398], [78], [307], [434], [488], [166], [187], [36], [290], [460], [127], [1], [378], [58], [258], [144], [280]
No Preview (70)	[34], [39], [497], [222], [215], [259], [404], [93], [333], [234], [271], [5], [63], [262], [23], [498], [496], [195], [376], [267], [304], [13], [348], [40], [171], [341], [302], [530], [446], [477], [141], [21], [447], [375], [107], [507], [328], [429], [326], [294], [284], [411], [119], [512], [486], [283], [8], [15], [393], [98], [335], [247], [441], [485], [97], [87], [329], [342], [125], [391], [349], [241], [265], [352], [203], [385], [163], [525], [480], [452]

Table 17: The 293 publications of the literature review, grouped by whether the AR authoring tool provides preview functionality of the AR content.

In-situ Authoring	**Publications**
In-situ Authoring (149)	[395], [306], [214], [89], [430], [388], [25], [533], [373], [318], [225], [523], [330], [324], [380], [509], [19], [71], [139], [235], [428], [10], [456], [180], [159], [368], [75], [491], [2], [29], [3], [111], [367], [536], [182], [424], [253], [255], [473], [414], [200], [245], [410], [48], [433], [515], [223], [175], [315], [351], [297], [81], [273], [431], [212], [474], [384], [183], [130], [379], [252], [288], [403], [336], [231], [151], [251], [522], [370], [519], [343], [117], [520], [338], [365], [17], [256], [376], [199], [513], [79], [204], [291], [500], [462], [501], [357], [346], [356], [157], [47], [32], [67], [189], [148], [531], [268], [272], [375], [422], [518], [192], [219], [514], [517], [179], [396], [250], [160], [266], [43], [350], [154], [499], [364], [132], [181], [42], [292], [73], [524], [450], [84], [455], [465], [412], [489], [88], [15], [92], [30], [33], [220], [344], [190], [345], [508], [239], [432], [398], [434], [290], [460], [127], [1], [378], [58], [258], [144]
Decontextualized (138)	[34], [39], [26], [80], [162], [358], [386], [401], [161], [134], [325], [497], [136], [222], [442], [261], [215], [259], [404], [93], [445], [303], [333], [234], [271], [5], [313], [63], [262], [121], [436], [249], [23], [44], [178], [498], [202], [354], [186], [496], [195], [254], [198], [421], [440], [77], [164], [267], [304], [13], [348], [40], [168], [171], [492], [103], [372], [193], [201], [281], [331], [494], [341], [534], [532], [270], [302], [530], [213], [446], [305], [382], [126], [477], [86], [76], [141], [21], [296], [447], [107], [108], [298], [366], [31], [507], [328], [211], [311], [429], [82], [326], [128], [294], [284], [411], [119], [451], [512], [486], [283], [475], [8], [16], [393], [98], [335], [247], [441], [485], [97], [87], [329], [342], [137], [237], [125], [391], [470], [196], [349], [241], [377], [265], [352], [203], [78], [307], [488], [166], [385], [163], [525], [187], [480], [36], [280], [452]
Partially In-situ (6)	[269], [427], [437], [355], [472], [95]

Table 18: The 293 publications of the literature review, grouped by whether content is authored in-situ or decontextualized.

Authoring Interaction	Publications
Markup Language (28)	[497], [404], [447], [530], [328], [271], [265], [385], [5], [498], [341], [222], [82], [93], [445], [259], [480], [40], [352], [342], [39], [262], [119], [234], [283], [507], [486], [333]
Tangible Marker & Objects & Interfaces (18)	[29], [239], [266], [357], [111], [235], [214], [398], [380], [180], [368], [355], [220], [272], [43], [499], [157], [434]
Visual Scripting (18)	[388], [349], [358], [225], [477], [304], [63], [134], [410], [13], [346], [34], [76], [183], [164], [326], [411], [30]
Gestures/Handtracking (13)	[515], [175], [395], [365], [366], [79], [514], [139], [472], [343], [144], [88], [77]
External Controllers (9)	[324], [440], [382], [199], [462], [291], [513], [58], [181]
Drawing/Sketching (5)	[351], [223], [198], [182], [42]
Video & Picture Annotation (5)	[269], [254], [492], [121], [47]
Not Specified (4)	[23], [496], [215], [160]
Robots & Drones (3)	[533], [318], [75]
Tangible Marker & Objects & Interfaces, Drawing/Sketching (3)	[450], [89], [421]
Voice (3)	[302], [71], [338]
Automatic Task Segmentation (2)	[376], [375]
Gestures/Handtracking, External Controllers/Sensors (2)	[78], [519]
Gestures/Handtracking, Gaze (2)	[509], [117]
Gestures/Handtracking, Tangible Marker & Objects & Interfaces (2)	[273], [430]
Gestures/Handtracking, Voice (2)	[520], [258]
Drawing/Sketching, External Controllers (1)	[154]
Drawing/Sketching, Gestures/Handtracking (1)	[517]
External Controllers, Gestures/Handtracking (1)	[17]
External Controllers, Tangible Marker & Objects & Interfaces (1)	[256]
Gaze (1)	[470]
Gestures/Handtracking, Voice, Gaze (1)	[501]
Gestures/Handtracking, Voice, Tangible Marker & Objects & Interfaces (1)	[500]
Tangible Marker & Objects & Interfaces, Visual Scripting (1)	[364]
Tangible Marker & Objects & Interfaces, Voice (1)	[204]
Tangible Marker & Objects & Interfaces, Voice, Gestures/Handtracking (1)	[350]
Textual Description (1)	[427]

Table 19: The publications of the literature review, grouped by primary authoring interaction concept. Only publications which are using concepts beyond the "traditional interaction concepts" (163 publications) of the authoring hardware utilized are mapped in this table.

App Relationship	Publications
External (145)	[306], [34], [39], [80], [162], [358], [386], [401], [161], [325], [159], [497], [136], [222], [442], [261], [215], [259], [404], [93], [445], [303], [491], [333], [234], [271], [5], [63], [262], [2], [121], [436], [249], [29], [23], [44], [178], [498], [354], [111], [367], [496], [183], [403], [522], [421], [440], [77], [164], [267], [304], [13], [348], [40], [346], [168], [171], [103], [372], [193], [47], [201], [281], [331], [494], [341], [534], [532], [270], [302], [530], [213], [446], [305], [382], [268], [126], [477], [86], [76], [141], [21], [296], [447], [375], [107], [108], [298], [422], [396], [366], [31], [328], [311], [82], [326], [128], [294], [284], [73], [411], [119], [451], [512], [486], [283], [16], [88], [15], [98], [335], [247], [441], [30], [485], [33], [97], [87], [329], [342], [137], [237], [125], [391], [470], [349], [241], [377], [265], [352], [239], [78], [307], [488], [166], [385], [163], [525], [187], [480], [36], [290], [1], [280], [452]
Internal (134)	[395], [214], [89], [430], [388], [25], [533], [26], [373], [318], [225], [330], [324], [380], [19], [134], [71], [235], [428], [10], [456], [180], [368], [75], [313], [202], [3], [186], [182], [424], [253], [255], [473], [414], [200], [245], [410], [48], [433], [515], [223], [175], [315], [195], [254], [351], [297], [81], [273], [431], [212], [474], [384], [130], [379], [252], [288], [336], [231], [198], [151], [251], [519], [343], [117], [520], [338], [365], [17], [256], [376], [199], [513], [79], [204], [291], [500], [462], [501], [357], [356], [157], [32], [67], [189], [148], [531], [272], [427], [518], [192], [219], [514], [517], [179], [250], [160], [266], [43], [350], [154], [499], [364], [507], [132], [181], [211], [42], [292], [429], [450], [475], [84], [455], [465], [412], [489], [8], [393], [196], [344], [190], [345], [508], [203], [432], [398], [434], [460], [127], [378], [58], [258], [144]
Split (11)	[523], [509], [139], [536], [269], [492], [524], [437], [355], [220], [95]
Both (3)	[370], [92], [472]

Table 20: The 293 publications of the literature review, grouped by whether the application that is used to author the AR construct is also the application for its usage (internal), or one application authors content to be used with a second application (external).

Construct Distribution	Publications
Local (176)	[214], [89], [430], [388], [39], [533], [26], [80], [373], [225], [162], [523], [324], [380], [401], [509], [161], [134], [71], [139], [235], [428], [10], [456], [325], [180], [368], [75], [497], [222], [261], [93], [303], [491], [333], [234], [63], [262], [2], [121], [436], [29], [23], [44], [178], [202], [3], [354], [111], [367], [186], [536], [182], [424], [253], [414], [200], [245], [410], [48], [433], [515], [351], [81], [273], [183], [130], [379], [288], [231], [198], [522], [370], [519], [343], [117], [520], [338], [365], [17], [256], [376], [199], [79], [204], [291], [462], [501], [357], [440], [77], [164], [492], [356], [372], [193], [157], [201], [281], [148], [534], [532], [302], [446], [305], [531], [382], [268], [126], [477], [86], [76], [141], [21], [296], [447], [375], [107], [108], [298], [422], [427], [192], [219], [514], [517], [179], [396], [250], [160], [266], [43], [350], [154], [499], [507], [132], [181], [211], [42], [292], [311], [429], [326], [73], [119], [450], [84], [455], [465], [412], [489], [88], [125], [470], [196], [345], [508], [377], [265], [352], [239], [307], [434], [488], [166], [163], [525], [187], [460], [127], [1], [378], [58], [258], [144]
Server (117)	[395], [306], [34], [25], [318], [358], [330], [386], [19], [159], [136], [442], [215], [259], [404], [445], [271], [5], [313], [249], [498], [269], [255], [473], [223], [175], [496], [315], [195], [254], [297], [431], [212], [474], [384], [252], [403], [336], [151], [251], [513], [500], [421], [267], [304], [13], [348], [40], [346], [168], [171], [103], [47], [32], [67], [189], [331], [494], [341], [270], [530], [213], [272], [518], [364], [366], [31], [328], [82], [128], [294], [284], [411], [451], [512], [486], [283], [524], [475], [8], [16], [15], [393], [98], [335], [247], [441], [92], [30], [485], [33], [97], [87], [437], [329], [342], [137], [237], [355], [391], [220], [472], [344], [190], [349], [241], [95], [203], [432], [398], [78], [385], [480], [36], [290], [280], [452]

Table 21: The 293 publications of the literature review, grouped whether the AR constructs are distributed locally or through a server.

Construct User	Publications
End user & Consumer (120)	[161], [134], [71], [139], [235], [428], [10], [456], [325], [180], [159], [368], [75], [497], [136], [222], [442], [261], [215], [259], [404], [93], [445], [303], [491], [333], [234], [271], [5], [313], [63], [262], [2], [121], [436], [249], [29], [23], [44], [178], [498], [202], [3], [354], [111], [367], [186], [536], [269], [182], [424], [253], [255], [473], [414], [200], [245], [410], [48], [433], [515], [223], [175], [496], [315], [195], [254], [351], [297], [81], [273], [431], [212], [474], [384], [183], [130], [379], [252], [288], [403], [336], [231], [198], [151], [251], [522], [370], [519], [343], [117], [520], [338], [365], [17], [256], [376], [199], [513], [79], [204], [291], [500], [462], [501], [421], [357], [440], [77], [164], [267], [304], [13], [348], [40], [346], [168], [171], [492], [103]
Students (45)	[507], [328], [132], [181], [211], [42], [292], [311], [429], [82], [326], [128], [294], [284], [73], [411], [119], [451], [512], [486], [283], [524], [450], [475], [84], [455], [465], [412], [489], [8], [16], [88], [15], [393], [98], [335], [247], [441], [92], [30], [485], [33], [97], [87], [437]
Not specified (33)	[268], [272], [126], [477], [86], [76], [141], [21], [296], [447], [375], [107], [108], [298], [422], [427], [518], [192], [219], [514], [517], [179], [396], [250], [160], [266], [43], [350], [154], [499], [364], [366], [31]
Workers (20)	[307], [434], [488], [166], [385], [163], [525], [187], [480], [36], [290], [460], [127], [1], [378], [58], [258], [144], [280], [452]
Maintenance & Assembly Worker (15)	[281], [148], [331], [494], [341], [534], [532], [270], [302], [530], [213], [446], [305], [531], [382]
Visitors (12)	[508], [349], [241], [377], [265], [352], [239], [95], [203], [432], [398], [78]
Designers (11)	[26], [80], [373], [318], [225], [162], [358], [523], [330], [324], [380]
Trainees (8)	[391], [470], [196], [220], [472], [344], [190], [345]
Technicians (6)	[329], [342], [137], [237], [355], [125]
Children (5)	[34], [214], [89], [430], [388]
Engineers (5)	[356], [372], [193], [157], [189]
Construction Personal (3)	[39], [25], [533]
Elderly People (3)	[386], [401], [509]
Factory Workers (3)	[47], [32], [67]
Analysts (1)	[395]
Astronauts (1)	[306]
Employees (1)	[19]
Film Directors (1)	[201]

Table 22: The 293 publications of the literature review, grouped by the (described or inferrable) intended user of the AR construct.

Usage Hardware	**Publications**
Handheld (141)	[34], [214], [89], [25], [533], [523], [386], [19], [497], [136], [222], [442], [261], [215], [259], [404], [93], [445], [303], [491], [333], [234], [536], [269], [424], [253], [255], [473], [414], [200], [245], [410], [48], [433], [515], [223], [175], [496], [315], [195], [254], [351], [297], [81], [273], [431], [212], [474], [384], [183], [130], [379], [252], [288], [403], [336], [231], [198], [151], [251], [519], [421], [440], [77], [267], [304], [13], [348], [40], [346], [168], [171], [492], [32], [67], [189], [331], [494], [341], [534], [532], [86], [76], [141], [427], [518], [192], [219], [514], [517], [179], [396], [250], [160], [31], [82], [326], [128], [294], [284], [524], [450], [475], [84], [455], [465], [412], [489], [8], [88], [15], [393], [98], [335], [247], [441], [92], [30], [485], [33], [97], [437], [329], [342], [125], [391], [220], [344], [190], [349], [241], [95], [203], [432], [398], [166], [290], [460], [127], [1], [378], [385], [330], [182], [225], [162], [271], [5], [270], [21], [73], [87], [313], [63], [262], [103], [372], [47], [302], [411], [345], [78], [163]
Head-mounted (73)	[395], [306], [430], [39], [358], [324], [380], [401], [509], [2], [121], [436], [249], [29], [23], [44], [178], [522], [343], [117], [520], [338], [365], [17], [256], [376], [199], [513], [79], [204], [291], [500], [462], [501], [164], [193], [157], [201], [530], [213], [531], [382], [296], [447], [375], [107], [108], [298], [43], [350], [154], [499], [364], [366], [119], [451], [16], [137], [237], [355], [470], [472], [377], [265], [352], [525], [187], [480], [36], [58], [258], [280], [452], [330], [182], [498], [202], [3], [512], [313], [63], [262], [103], [372], [47], [302], [411], [345], [78], [163]
Desktop (39)	[388], [26], [80], [373], [318], [161], [134], [71], [139], [235], [428], [10], [456], [325], [180], [159], [368], [75], [356], [281], [148], [268], [272], [126], [477], [266], [507], [328], [132], [181], [211], [42], [292], [311], [429], [508], [307], [434], [488], [385], [330], [182], [498], [202], [3], [512], [225], [162], [271], [5], [270], [21], [73], [87]
Not specified (6)	[354], [111], [367], [446], [305], [422]
Projector (6)	[186], [370], [357], [196], [239], [144], [385]
Web (2)	[486], [283]

Table 23: The 293 publications of the literature review, grouped by the hardware supported for the usage of the authored AR constructs. As some authored AR constructs can be used on multiple hardware platforms, this includes overlap.

User Interactions	Publications
Viewing (125)	[395], [89], [533], [373], [318], [162], [358], [523], [330], [324], [380], [401], [71], [139], [10], [456], [325], [180], [368], [75], [234], [271], [313], [121], [249], [23], [178], [498], [202], [354], [111], [367], [269], [473], [414], [200], [245], [410], [48], [433], [515], [223], [175], [195], [254], [351], [297], [81], [273], [252], [288], [198], [151], [251], [522], [370], [117], [520], [338], [365], [17], [256], [376], [199], [513], [79], [204], [291], [77], [40], [492], [356], [372], [193], [157], [201], [148], [302], [213], [446], [305], [477], [86], [21], [447], [298], [422], [427], [518], [192], [219], [517], [179], [396], [250], [160], [43], [350], [154], [499], [31], [328], [42], [292], [82], [73], [451], [512], [455], [465], [412], [8], [16], [485], [33], [87], [237], [355], [508], [398], [78], [434], [488], [36], [290]
Traditional Interaction Techniques (94)	[34], [25], [26], [80], [19], [136], [222], [442], [215], [259], [404], [93], [445], [303], [491], [333], [5], [182], [424], [253], [255], [496], [315], [431], [212], [474], [384], [183], [130], [379], [336], [231], [343], [421], [440], [267], [304], [13], [348], [346], [168], [171], [47], [32], [67], [189], [281], [331], [494], [341], [534], [532], [270], [126], [141], [211], [311], [326], [128], [284], [411], [283], [524], [475], [84], [489], [88], [15], [393], [98], [335], [441], [92], [30], [97], [437], [329], [342], [125], [391], [344], [190], [345], [349], [241], [352], [95], [203], [432], [307], [460], [127], [1], [378]
Tangible Markers (31)	[430], [388], [225], [134], [235], [159], [497], [261], [262], [2], [29], [44], [536], [357], [268], [272], [76], [514], [266], [364], [507], [132], [181], [119], [486], [450], [247], [196], [220], [265], [239]
Combinatory Approaches (19)	[214], [428], [63], [436], [3], [501], [103], [382], [296], [108], [366], [137], [470], [472], [525], [58], [258], [280], [452]
Gestures/Handtracking (9)	[186], [403], [519], [164], [530], [531], [107], [377], [144]
External Controllers/Sensors (8)	[39], [161], [462], [375], [166], [385], [163], [187]
Voice (5)	[306], [509], [500], [294], [480]
Head/Eye Gaze (1)	[386]
Tangible markers (1)	[429]

Table 24: The 293 publications of the literature review, grouped by main interaction concept used to interact with the authored AR constructs on the usage hardware.

Content	Publications
3D Model (227)	[306], [34], [89], [430], [39], [26], [80], [373], [318], [225], [162], [358], [523], [330], [324], [380], [509], [19], [161], [134], [71], [139], [235], [428], [10], [456], [325], [180], [159], [368], [75], [497], [136], [442], [261], [215], [259], [404], [93], [445], [491], [234], [271], [5], [313], [63], [262], [2], [121], [436], [249], [29], [44], [178], [498], [202], [3], [354], [111], [536], [182], [424], [473], [414], [200], [245], [410], [48], [433], [515], [175], [496], [297], [81], [273], [212], [474], [384], [183], [130], [252], [403], [336], [231], [151], [251], [522], [519], [343], [520], [365], [256], [513], [204], [291], [500], [462], [501], [421], [357], [440], [77], [164], [267], [304], [13], [348], [346], [168], [171], [103], [356], [372], [193], [157], [47], [32], [67], [201], [189], [281], [148], [331], [494], [341], [534], [532], [270], [530], [446], [305], [531], [382], [268], [272], [126], [86], [76], [296], [447], [107], [108], [422], [427], [192], [219], [514], [517], [396], [250], [266], [43], [350], [499], [364], [366], [31], [507], [328], [132], [181], [211], [42], [292], [311], [429], [326], [294], [284], [73], [411], [119], [451], [512], [486], [283], [475], [84], [455], [489], [8], [16], [88], [98], [247], [92], [30], [485], [33], [87], [437], [329], [342], [137], [237], [355], [125], [391], [220], [472], [344], [345], [508], [349], [377], [265], [239], [203], [78], [307], [434], [488], [166], [385], [163], [525], [187], [36], [290], [460], [127], [1], [378], [58], [258], [280], [452]
Text (125)	[306], [34], [225], [523], [330], [380], [509], [19], [159], [497], [136], [222], [442], [215], [259], [93], [445], [303], [333], [234], [5], [313], [2], [436], [23], [178], [3], [354], [367], [186], [424], [473], [414], [223], [496], [315], [195], [254], [297], [431], [474], [384], [379], [252], [288], [403], [336], [231], [198], [151], [338], [500], [164], [267], [304], [346], [168], [47], [32], [281], [331], [494], [341], [534], [270], [530], [531], [382], [126], [477], [86], [21], [447], [518], [179], [396], [364], [311], [326], [128], [512], [486], [283], [524], [450], [8], [88], [15], [393], [98], [441], [97], [87], [437], [342], [137], [237], [355], [125], [391], [470], [472], [190], [345], [508], [349], [241], [377], [95], [203], [432], [398], [307], [488], [385], [163], [525], [480], [36], [127], [1], [378], [58], [280], [452]
2D Images / Sprites (103)	[395], [306], [34], [214], [388], [225], [162], [380], [386], [401], [509], [159], [136], [442], [259], [404], [303], [333], [5], [2], [436], [178], [3], [354], [111], [367], [186], [269], [424], [253], [473], [223], [496], [315], [195], [297], [81], [431], [212], [474], [336], [198], [151], [251], [370], [376], [199], [267], [13], [40], [168], [67], [281], [331], [494], [530], [382], [141], [21], [375], [518], [250], [366], [292], [311], [82], [326], [294], [283], [524], [465], [412], [8], [15], [393], [98], [335], [247], [441], [485], [33], [97], [87], [437], [329], [355], [125], [470], [196], [508], [265], [352], [239], [95], [203], [432], [398], [385], [525], [36], [1], [144], [452]
3D Animations (63)	[89], [39], [26], [80], [162], [161], [134], [368], [313], [63], [2], [121], [436], [249], [29], [44], [202], [536], [182], [515], [273], [212], [183], [403], [519], [343], [500], [501], [357], [346], [168], [193], [201], [281], [494], [534], [446], [268], [76], [427], [266], [43], [211], [42], [119], [486], [283], [475], [84], [16], [30], [485], [237], [355], [344], [345], [377], [434], [163], [36], [460], [58], [452]
Highlights / Arrows (52)	[25], [533], [358], [386], [442], [215], [445], [333], [23], [269], [253], [414], [254], [252], [288], [336], [231], [251], [522], [117], [47], [32], [189], [494], [341], [530], [213], [531], [477], [179], [396], [160], [311], [128], [411], [451], [450], [92], [342], [137], [237], [470], [196], [190], [345], [307], [488], [385], [460], [1], [378], [144]
Audio (43)	[225], [161], [159], [497], [136], [234], [2], [44], [178], [3], [367], [536], [253], [473], [496], [474], [252], [403], [251], [513], [462], [501], [304], [13], [348], [346], [168], [179], [292], [128], [119], [512], [283], [247], [485], [33], [87], [190], [265], [352], [239], [398], [258]
Video / Animations (32)	[306], [34], [222], [5], [2], [178], [3], [111], [186], [255], [473], [212], [403], [500], [13], [168], [492], [281], [534], [21], [298], [292], [119], [512], [8], [485], [87], [391], [265], [398], [36], [258]
Drawings (21)	[395], [324], [200], [245], [223], [351], [81], [273], [151], [117], [17], [79], [462], [421], [341], [296], [250], [154], [450], [342], [290]
Photos (7)	[234], [500], [8], [342], [137], [163], [58]

Table 25: The 293 publications of the literature reviewed, grouped by the AR content type they support. As some authored AR constructs utilize multiple content types, this includes overlap.

Sequentiality	Publications
Constant (184)	[395], [89], [25], [533], [26], [80], [373], [318], [162], [523], [330], [324], [380], [386], [401], [19], [71], [139], [428], [10], [456], [325], [180], [368], [75], [136], [222], [215], [259], [404], [445], [491], [234], [271], [5], [2], [121], [436], [249], [29], [23], [44], [178], [498], [202], [111], [367], [186], [536], [269], [424], [253], [255], [414], [200], [245], [48], [433], [515], [223], [175], [496], [315], [195], [254], [351], [297], [81], [273], [431], [212], [474], [384], [130], [252], [288], [403], [336], [231], [198], [151], [251], [370], [117], [520], [338], [365], [17], [256], [199], [513], [79], [204], [291], [501], [421], [357], [440], [77], [164], [267], [13], [40], [171], [103], [372], [193], [157], [32], [67], [201], [189], [148], [213], [382], [86], [141], [21], [107], [108], [298], [427], [518], [192], [219], [514], [517], [179], [396], [250], [160], [43], [350], [154], [499], [366], [31], [328], [132], [42], [292], [429], [82], [128], [294], [73], [119], [451], [512], [486], [283], [524], [450], [455], [465], [412], [489], [8], [16], [88], [15], [393], [98], [335], [247], [441], [485], [33], [97], [87], [137], [391], [220], [508], [377], [352], [239], [95], [203], [432], [398], [78], [480], [1]
Sequential (109)	[306], [34], [214], [430], [388], [39], [225], [358], [509], [161], [134], [235], [159], [497], [442], [261], [93], [303], [333], [313], [63], [262], [3], [354], [182], [473], [410], [183], [379], [522], [519], [343], [376], [500], [462], [304], [348], [346], [168], [492], [356], [47], [281], [331], [494], [341], [534], [532], [270], [302], [530], [446], [305], [531], [268], [272], [126], [477], [76], [296], [447], [375], [422], [266], [364], [507], [181], [211], [311], [326], [284], [411], [475], [84], [92], [30], [437], [329], [342], [237], [355], [125], [470], [196], [472], [344], [190], [345], [349], [241], [265], [307], [434], [488], [166], [385], [163], [525], [187], [36], [290], [460], [127], [378], [58], [258], [144], [280], [452]

Table 26: The 293 publications of the literature review, grouped by the complexity of provided content; therefore, whether it is sequential or constant.

Tracking Type	Publications
Marker (158)	[306], [34], [214], [89], [430], [388], [80], [373], [318], [225], [162], [523], [380], [401], [161], [134], [71], [139], [235], [428], [10], [180], [159], [368], [75], [497], [136], [442], [404], [93], [303], [491], [333], [234], [271], [5], [63], [262], [2], [249], [29], [44], [498], [3], [111], [367], [536], [182], [414], [410], [496], [81], [273], [212], [183], [130], [403], [336], [198], [522], [365], [256], [513], [204], [13], [346], [356], [372], [157], [32], [67], [281], [148], [331], [532], [270], [302], [530], [446], [305], [531], [382], [268], [126], [477], [76], [21], [447], [107], [108], [422], [518], [192], [219], [514], [517], [250], [266], [43], [350], [499], [366], [31], [507], [328], [132], [181], [211], [42], [292], [311], [429], [82], [294], [284], [73], [119], [512], [486], [283], [524], [455], [465], [412], [489], [8], [16], [88], [15], [393], [98], [247], [441], [92], [30], [485], [33], [97], [87], [437], [329], [342], [237], [220], [345], [508], [349], [265], [239], [95], [307], [434], [166], [187], [36], [290], [127], [1]
Markerless (57)	[395], [25], [533], [26], [358], [330], [324], [509], [19], [456], [313], [121], [436], [23], [178], [253], [200], [245], [48], [433], [515], [384], [379], [519], [343], [117], [520], [338], [17], [199], [79], [291], [500], [462], [501], [440], [164], [171], [193], [213], [141], [375], [298], [427], [179], [396], [154], [137], [125], [470], [472], [344], [190], [377], [460], [58], [258]
Object (14)	[386], [325], [261], [223], [288], [376], [494], [341], [128], [411], [488], [163], [280], [452]
GPS (11)	[39], [222], [473], [315], [195], [254], [267], [304], [348], [40], [241]
GPS, Markerless (11)	[445], [269], [424], [255], [151], [421], [189], [160], [451], [84], [203]
Marker, Markerless (11)	[354], [175], [168], [103], [534], [296], [364], [432], [78], [525], [378]
Marker, GPS (5)	[259], [474], [77], [86], [391]
No Tracking (5)	[186], [370], [357], [196], [144]
Marker, Object (3)	[215], [351], [47]
Marker, GPS, Markerless (3)	[297], [252], [251]
QR Code (2)	[492], [480]
External Sensors (2)	[201], [352]
Internal Sensors (1)	[202]
Marker, QR Code (1)	[431]
Markerless, Internal Sensors (1)	[231]
Marker, RFID (1)	[272]
Marker, Object, Text Recognition (1)	[326]
Markerless, Object (1)	[450]
Markerless, QR Code (1)	[475]
Marker, GPS, QR Code (1)	[335]
marker (1)	[355]
GPS, QR Code (1)	[398]
Marker, Static (1)	[385]

Table 27: The 293 publications of the literature review, grouped by combination of tracking techniques utilized to contextualize the AR content.

4 System Usability Scale Results in Table Format

SUS Metrics	Usability Study 1	Usability Study 2 (Implicit Interaction)	Usability Study 2 (Explicit Interaction)	Usability Study 3
SUS Score (mean)	64.58	63	81	80
SD	7.13	7.97	3	7.32
Min	52.5	50	77.5	67.5
Max	72.5	70	85	90
1. Quartile	58.125	53.75	77.5	72.5
Median	65	67.5	82.5	82.5
3. Quartile	72.5	70	83.75	85
Item 1	6.25 (3.15)	6.5 (2.55)	8.0 (1.87)	6.43 (2.26)
Item 2	7.5 (2.04)	9.0 (1.22)	9.0 (1.22)	8.57 (1.24)
Item 3	5.42 (1.72)	4.5 (1.87)	8.5 (1.22)	7.5 (1.34)
Item 4	7.5 (2.5)	5.5 (1.87)	9.0 (1.22)	9.64 (0.87)
Item 5	7.08 (0.93)	7.0 (1.0)	8.0 (1.0)	7.5 (1.34)
Item 6	7.5 (1.44)	7.5 (1.58)	8.0 (2.92)	8.57 (1.24)
Item 7	5.42 (1.72)	6.5 (2.55)	8.0 (1.0)	7.86 (2.08)
Item 8	6.25 (2.8)	6.0 (2.0)	8.5 (2.0)	8.57 (1.82)
Item 9	4.17 (1.86)	3.5 (2.0)	6.0 (1.22)	7.14 (1.6)
Item 10	7.5 (1.44)	7.0 (2.92)	8.0 (1.0)	8.21 (2.2)
Questionnaires	6	5	5	7
Conclusiveness [463]	35%	0%	0%	55%
Adjective Scale [27]	OK	OK	Excellent	Good
Grade Scale [416]	C	C	A	A
Quartile Scale [27]	2nd	2nd	4th	4th
Acceptability [28]	Marginal	Marginal	Acceptable	Acceptable
NPS Scale [415]	Passive	Passive	Promoter	Promoter
Industry Benchm. [278]	Below Average	Below Average	Above Industry Benchmark	Above Industry Benchmark

Table 28: All System Usability Scale metrics for the formative usability evaluations for the design of the TrainAR interaction concept reported in Section 6.3. Calculated with the SUS Analysis Toolkit [50].

SUS Metrics	Training 1	Training 2	Training 3	Training 3 (Multi-user)
SUS Score (mean)	83.11	69.81	80.29	60.08
SD	12.9	15.69	12.75	14.33
Min	47.5	27.5	52.5	25.0
Max	100.0	92.5	100.0	95.0
1. Quartile	72.5	62.5	72.5	50.0
Median	87.5	72.5	80.0	60.0
3. Quartile	93.75	80.0	90.625	72.5
Item 1	7.73 (2.25)	6.03 (2.87)	7.5 (2.59)	3.64 (2.31)
Item 2	8.41 (2.29)	7.18 (2.61)	8.46 (1.97)	4.39 (2.46)
Item 3	8.11 (1.95)	7.31 (2.68)	8.08 (1.87)	6.44 (3.02)
Item 4	9.24 (1.44)	6.99 (3.11)	8.37 (2.4)	8.48 (2.68)
Item 5	8.33 (2.01)	6.79 (2.26)	8.08 (1.87)	5.91 (1.83)
Item 6	7.05 (2.26)	6.15 (2.52)	7.79 (2.43)	5.0 (2.68)
Item 7	8.56 (2.22)	6.86 (2.1)	7.79 (2.12)	7.27 (2.71)
Item 8	8.18 (2.33)	6.86 (2.58)	8.37 (2.19)	4.77 (3.28)
Item 9	8.26 (1.9)	7.18 (2.28)	7.69 (2.07)	6.14 (2.54)
Item 10	9.24 (1.15)	8.46 (2.24)	8.17 (2.14)	8.03 (1.92)
Questionnaires	33	39	26	33
Conclusiveness [463]	100%	100%	100%	100%
Adjective Scale [27]	Excellent	OK	Good	OK
Grade Scale [416]	A	C	A	D
Quartile Scale [27]	4th	2nd	4th	1st
Acceptability [28]	Acceptable	Marginal	Acceptable	Marginal
NPS Scale [415]	Promoter	Passive	Promoter	Detractor
Industry Benchm. [278]	Above Industry Benchmark	Above Average	Above Industry Benchmark	Below Average

Table 29: All System Usability Scale metrics for Training 1 (Tocolysis), Training 2 (Sectio), Training 3 (Reanimation) and the multi-user version of Training 3, which are reported in Section 4.6.2, 4.6.3, and 4.6.4. Calculated with the SUS Analysis Toolkit [50].

SUS Metrics	Training 4	Training 5
SUS Score (mean)	73.0	84.79
SD	15.6	13.51
Min	40.0	55.0
Max	92.5	100.0
1. Quartile	63.125	75.625
Median	71.25	90.0
3. Quartile	88.125	96.875
Item 1	7.0 (1.87)	8.47 (2.38)
Item 2	8.5 (1.66)	8.75 (2.08)
Item 3	7.75 (2.84)	8.47 (1.98)
Item 4	7.75 (1.75)	8.96 (1.71)
Item 5	7.0 (2.18)	8.4 (1.68)
Item 6	6.5 (2.29)	8.61 (1.71)
Item 7	7.0 (3.84)	8.68 (1.81)
Item 8	8.25 (2.25)	8.33 (2.36)
Item 9	5.25 (2.36)	8.26 (1.94)
Item 10	8.0 (2.45)	7.85 (3.01)
Questionnaires	10	36
Conclusiveness [463]	80%	100%
Adjective Scale [27]	Good	Best Imaginable
Grade Scale [416]	B	A
Quartile Scale [27]	3rd	4th
Acceptability Scale [28]	Acceptable	Acceptable
NPS Scale [415]	Passive	Promoter
Industry Benchmark[278]	Above Average	Above Industry Benchmark

Table 30: All System Usability Scale metrics for Training 4 (Virtual reanimation), Training 5 (Pelvis Termini), reported in Sections 4.6.6 and 4.6.5. Calculated with the SUS Analysis Toolkit [50].

SUS Metrics	Training 2 (Lecturers)	Training 3 (Lecturers)
SUS Score (mean)	75.42	75.71
SD	10.25	6.64
Min	60	65
Max	92.5	82.5
1. Quartile	67.5	70
Median	75	80
3. Quartile	83.125	82.5
Item 1	8.75 (1.25)	8.93 (2.62)
Item 2	5.83 (2.76)	10.0 (0.0)
Item 3	7.08 (1.72)	5.36 (2.81)
Item 4	6.67 (2.36)	7.86 (1.6)
Item 5	7.92 (0.93)	7.5 (2.31)
Item 6	7.92 (0.93)	8.57 (1.24)
Item 7	8.75 (1.25)	7.14 (2.81)
Item 8	7.08 (2.67)	7.5 (3.27)
Item 9	7.08 (1.72)	6.79 (2.9)
Item 10	8.33 (1.86)	6.07 (2.62)
Questionnaires	6	7
Conclusiveness [463]	35%	55%
Adjective Scale [27]	Good	Good
Grade Scale [416]	B	B
Quartile Scale [27]	3rd	3rd
Acceptability Scale [28]	Acceptable	Acceptable
NPS Scale [415]	Passive	Passive
Industry Benchmark[278]	Above Average	Above Average

Table 31: All System Usability Scale metrics for Training 2 and Training 3 from the lecturers' perspective after completing the trainings during workshops (see Section 4.6.7). Calculated with the SUS Analysis Toolkit [50].

SUS Metrics	Computer Science Students	Media Technology Students	Non-technical Students
SUS Score (mean)	85.75	79.25	60.25
SD	9.88	9.36	15.14
Min	70	67.5	35
Max	97.5	97.5	82.5
1. Quartile	76.25	70	45
Median	86.25	76.25	65
3. Quartile	95.625	88.125	72.5
Item 1	8.0 (1.87)	8.0 (1.0)	5.5 (2.45)
Item 2	9.25 (1.15)	8.25 (2.51)	7.25 (2.36)
Item 3	8.5 (1.66)	7.75 (2.08)	4.25 (2.25)
Item 4	8.25 (1.95)	7.25 (2.84)	4.5 (2.69)
Item 5	8.75 (1.25)	8.5 (1.22)	8.25 (1.6)
Item 6	9.25 (1.6)	8.25 (2.25)	8.25 (1.95)
Item 7	8.25 (1.6)	7.75 (2.08)	6.5 (2.0)
Item 8	9.0 (1.66)	7.75 (3.05)	7.0 (2.18)
Item 9	8.25 (1.6)	7.75 (1.75)	4.75 (2.36)
Item 10	8.25 (1.95)	8.0 (1.87)	4.0 (2.78)
Questionnaires	10	10	10
Conclusiveness [463]	80%	80%	80%
Adjective Scale [27]	Best Imaginable	Good	OK
Grade Scale [416]	A	A	D
Quartile Scale [27]	4th	4th	1st
Acceptability [28]	Acceptable	Acceptable	Marginal
NPS Scale [415]	Promoter	Promoter	Detractor
Industry Benchm. [278]	Above Industry Benchmark	Above Average	Below Average

Table 32: All System Usability Scale metrics for the TrainAR Usability Study evaluation reported in Section 7.2.2. Calculated with the SUS Analysis Toolkit [50].

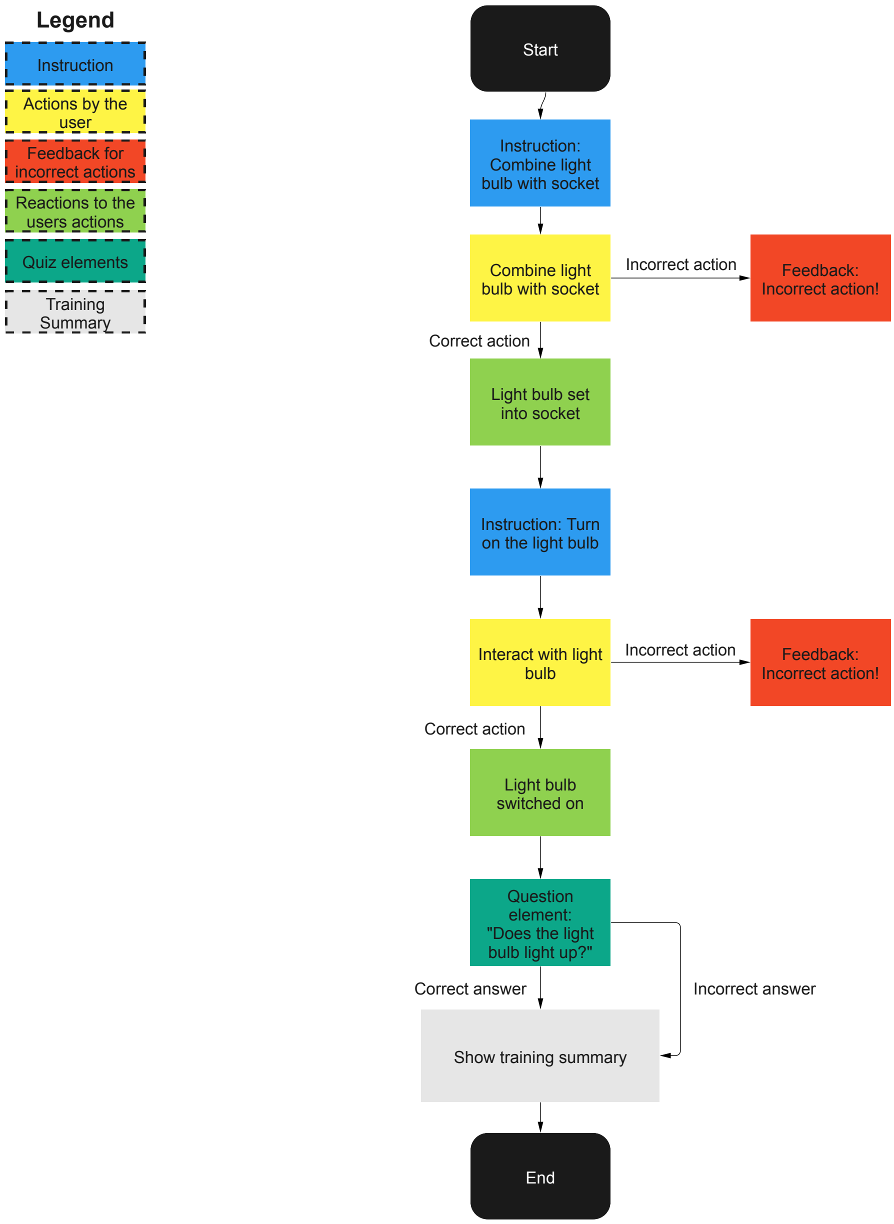
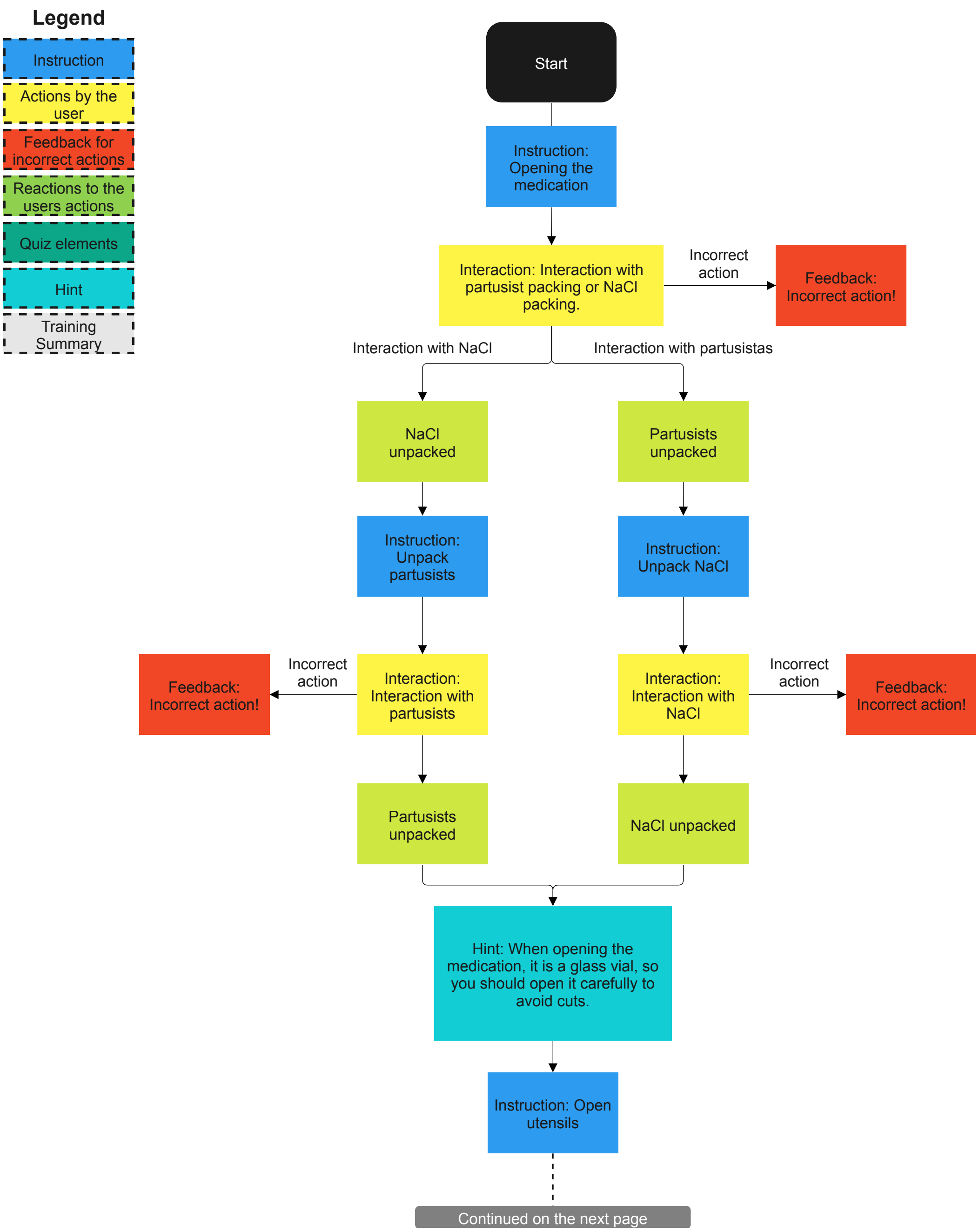


Figure 1: The first authoring task, the construction of a lamp setup, that was used in the systematic TrainAR Authoring Tool usability evaluation study. Translated and adapted from [37].

Figure 2: The second authoring task, the traditional East Frisian tea ceremony, that was used in the systematic TrainAR Authoring Tool usability evaluation study. Translated and adapted from [37].


Legend
Instruction
Actions by the user
Feedback for incorrect actions
Reactions to the users actions
Quiz elements
Hint
Training Summary
Start
Instruction: Opening the medication
Interaction: Interaction with partusist packing or NaCl packing.
Incorrect action
Feedback: Incorrect action!
Interaction with NaCl
Interaction with partusistas
NaCl unpacked
Partusists unpacked
Instruction: Unpack partusists
Instruction: Unpack NaCl
Feedback: Incorrect action!
Incorrect action
Interaction: Interaction with partusists
Interaction: Interaction with NaCl
Incorrect action
Feedback: Incorrect action!
Partusists unpacked
NaCl unpacked
Hint: When opening the medication, it is a glass vial, so you should open it carefully to avoid cuts.
Instruction: Open utensils
Continued on the next page

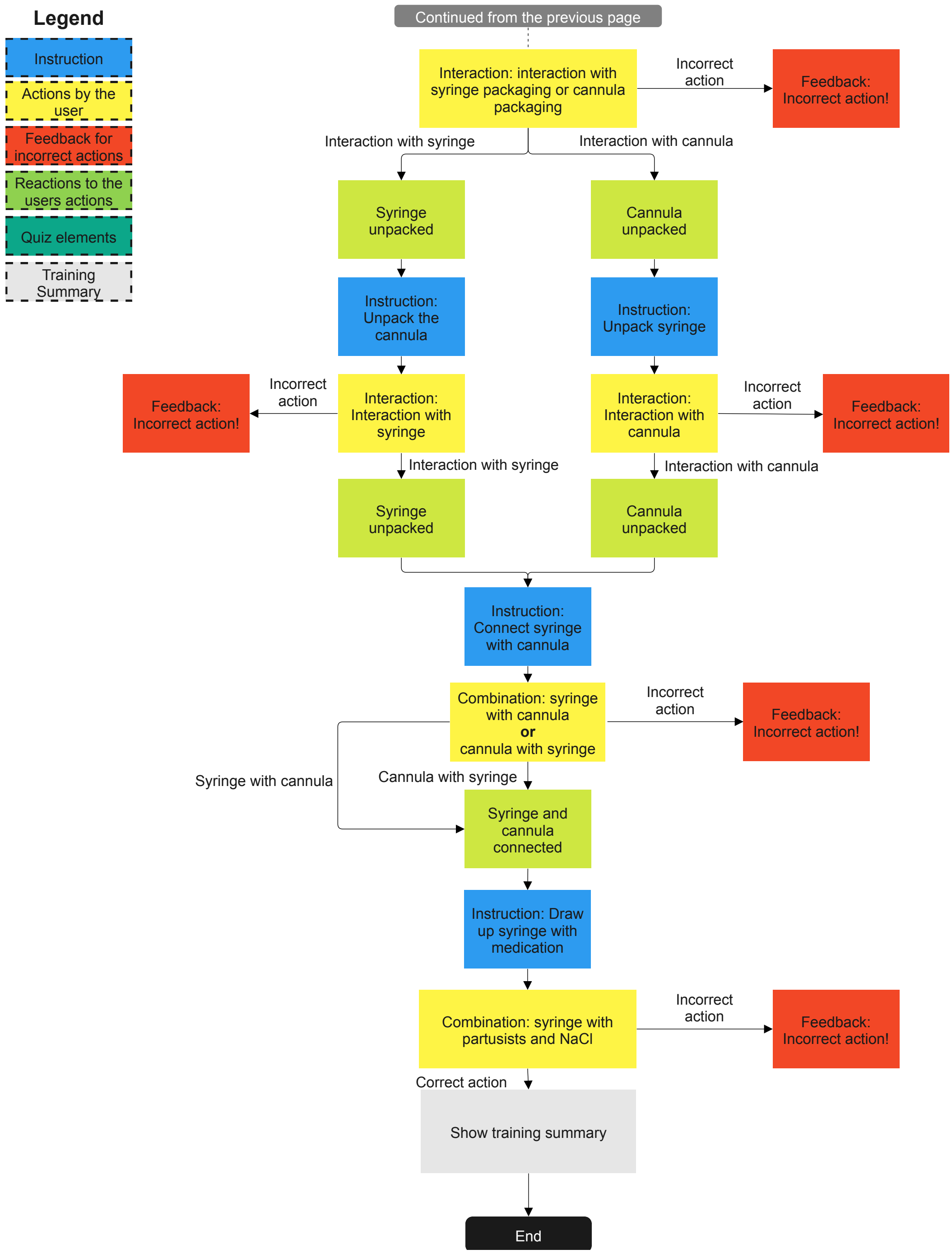


Figure 3: The third authoring task, the preparation of an injection, that was used in the systematic TrainAR Authoring Tool usability evaluation study. Translated and adapted from [37].