Augmented Reality in Informal Learning Environments

Design and Evaluation of Mobile Applications

Augmented Reality in Informal Learning Environments

Design and Evaluation of Mobile Applications

PhD Dissertation
by

Peter Sommerauer

Submitted at the IT University of Copenhagen, Denmark

Supervisor:
Oliver Müller, IT University of Copenhagen, Denmark

Committee:
Oliver Krancher, IT University of Copenhagen, Denmark
Arisa Shollo, Copenhagen Business School, Denmark
Matthias Trier, University Paderborn, Germany

2019

IT UNIVERSITY OF COPENHAGEN

Declaration

Declaration of Authorship

I, Peter SOMMERAUER, declare that this thesis titled, “Augmented Reality in Informal Learning Environments – Design and Evaluation of Mobile Applications” and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at the IT University of Copenhagen.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at the IT University of Copenhagen or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:
August 29, 2019

I would like to dedicate this thesis to my wonderful kids, Maximilian, Julia and Lukas.

Acknowledgements

Throughout my time as a Ph.D. student and the writing of this dissertation I have received a great deal of support and assistance. I would first like to thank my supervisor, Dr. Oliver Müller, who agreed to supervise my Ph.D. and whose expertise was invaluable in all the research and project work. You are holding the special ability to bring out the best in me!

I would like to acknowledge my colleagues from the IT University of Copenhagen for their wonderful collaboration. I would particularly like to thank the Ph.D. school who approved my application for becoming an external, independent and tuition fee-free Ph.D. student, and the Ph.D. support team Julie and Vibe. Special thanks go to Christoph, Tiemo, Leo, Niels and Hissu for their support in discussing particular topics and finding solutions for problems arising. You supported me greatly and were always willing to help me. I would also like to thank the members of the mid-term evaluation committee, Lene Nielsen, Arisa Shollo and Dr. Alexander Richter, for their valuable advice and feedback. Especially, I owe a debt to my former student fellow Mr. "K" for his inspiration towards the field of Augmented Reality at early stages.

Appreciation is due to the Liechtenstein National Museum, in particular Dr. Rainer Vollkommer. He made it possible to carry out the first AR based research project for this thesis. Special thanks to Georg Schierscher, the curator of the math exhibition "Matheliebe" who supported the AR application development as the main actor in the prepared video demonstration and in terms of sharing the comprehensive knowledge and evaluating the accuracy in the presentations of the learning content. I further owe an important debt to the project members of the "LAAR" project, including my colleagues from the Performance Augmentation Lab at the Oxford Brookes University, Dr. Fridolin Wild, Will Guest, and Joshua Secretan, Chris van Goetham from STEPP for challenging me in our unending discussions on European skills, qualifications and competencies, and Tommy Neumann from VPLT, the main association for the Event Technology industry in Germany, who supported the field research at the fair in Frankfurt.

In terms of funding, I am grateful to the European Commission for funding the research project LAAR – Principles for effective Learning Analytics in Augmented Reality learning applications for professional education (grant no. 2017-1-LI01-KA202-000087). I would like to thank Dr. Stefan Sohler and Clarissa Frommelt from AIBA – the Agency for international education affairs in Liechtenstein (EASMUS+ National Agency) for their administrative support of the research project.

Lastly, I could not have got through this without the support of my friends and family. My biggest gratitude goes to you, Angelika, for your understanding and motivation, inspiration and love.

Mauren, August 2019 Peter Sommerauer

Abstract

The digitization of the education sector is already in progress, and the new concepts and technologies are continuously explored. In particular, Augmented Reality (AR) is receiving increasing focus, and many studies are investigating its effectiveness for teaching and training. However, only a few studies have been conducted outside laboratory settings, and most of them have used questionnaires to evaluate learning outcomes in terms of concepts aspects like perceived learning, perceived usefulness, and students' motivation. What's more, the development of such AR applications is rarely based on design guidelines that support a well-founded evaluation and measurement of learning effects based on significant metrics.

Against this background, this cumulative thesis and the related research project introduce the study of AR outside laboratory settings and in informal learning environments. It covers the research (i.e., steps, processes, tasks, and results) undertaken to develop a framework that provides design guidelines for the development of AR applications that support teaching and training. Following a well-defined design science research process, this thesis covers the design and implementation of AR applications as IT artifacts and their applicability and utility in empirical terms. Therefore, the design framework acts as a toolbox that is structured in layers and considers design elements derived from learning theories to support learning and the development of efficient applications. The thesis also provides metrics for the measurement of user and task performance and discusses the requirements for research rigor and validity. The application of the design framework was successfully tested in several implementations in a variety of learning environments. The results, the impact, the practical benefits, and the findings from the field studies make valuable contributions to the knowledge base for mobile AR in education.

This thesis covers most of the research work in the educational research and development project "LAAR - Principles for effective Learning Analytics in AR learning applications for professional education," which was conducted between 2017 and 2019 and is listed in the ERASMUS+ Program of the European Commission under the number 2017-1-LI01-KA202-000087.

Resume

Digitaliseringen af uddannelsessektoren er allerede i gang, og nye koncepter og teknologier udforskes løbende. Især Augmented Reality (AR) får stadig større opmærksomhed, og mange undersøgelser peger på dens virkning inden for undervisning og uddannelse. Der er dog kun blevet gennemført nogle få studier uden for laboratorieomgivelser, og i de fleste anvender de spørgeskemaer til at vurdere læringsresultater med henblik på konceptuelle aspekter som oplevet læring, oplevet udbytte og de studerendes motivation. Derudover beror udviklingen af sådanne AR-applikationer sjældent på designretningslinjer, der støtter en velfunderet bedømmelse og måling af læringsresultater på grundlag af signifikante vurderingskriterier.

På denne baggrund introducerer denne kumulative afhandling og det forbundne forskningsprojekt en undersøgelse af AR, der går ud over laboratorieparametre og fokuserer på uformelle læringsomgivelser. Dette omfatter den forskning (dvs. de skridt, processer, opgaver og resultater), der er blevet gennemført for at udvikle en ramme, der giver designretningslinjer til udvikling af AR-applikationer, som understøtter undervisning og uddannelse. Denne afhandling følger en veldefineret udviklingsvidenskabelig forskningsproces og behandler design og implementering af AR-applikationer som IT-produkter og deres anvendelighed og nytte i empirisk henseende. Derfor fungerer designrammen som en værktøjskasse, der er struktureret i lag og tager hensyn til designelementer fra læringsteorier for at støtte læring og udviklingen af effektive applikationer. Denne afhandling leverer også vurderingskriterier til måling af brugerpræstationer og opgaveudførelse og diskuterer kravene for forskningens kvalitet og gyldighed. Designrammen blev afprøvet med succes i forskellige implementeringer i en række læringsomgivelser. Resultaterne, virkningerne, de praktiske fordele og erkendelserne fra undersøgelserne i marken giver værdifulde bidrag til vidensbasen for mobil AR i uddannelsessektoren.

Denne afhandling omfatter det meste af forskningsarbejdet i uddannelsesforsknings- og udviklingsprojektet “LAAR - Principles for effective Learning Analytics in Augmented Reality learning applications for professional education” (principper for effektiv læringsanalyse i Augmented Reality-læringsapplikationer for erhvervsuddannelser). Projektet blev gennemført mellem 2017 og 2019 og er opført i Den Europæiske Kommissions ERASMUS+ program med nummeret 2017-1-LI01-KA202-000087.

Contents

List of Figures

List of Tables

Nomenclature

2D	Two-dimensional
3D	Three-dimensional
AI	Artificial intelligence
APA	American Psychological Association
App	Application
AR	Augmented Reality
ARLE	Augmented reality learning environment
CSCL	Computer supported collaborative learning
CTML	Cognitive theory of multimedia learning
DIFT	Distinctive image features
DSR	Design science research
EC	European Commission
ECIS	European conference of information systems
ELI	EDUCAUSE Learning Initiative
GPS	Global positioning system
HCI	Human controller interface
HICSS	Hawaii international conference on system sciences
HMD	Head-mounted displays
HOG	Histogram of gradient orientations
IO	Input-output

IS	Information systems
IT	Information technology
LMS	Learning management system
LTS	Long-term storage
LRS	Learning record store
M	Mean
Mdn	Median
MR	Mixed Reality
NMC	New Media Consortium
QR	Quick response
RG	Research gap
RQ	Research question
SD	Standard deviation
SLT	Situated learning theory
STEM	Science, technology, engineering, mathematics
STS	Short-term storage
SUS	System usability scale
TAM	Technology acceptance model
TUI	Tangible user interface
VET	Vocational education and training
VLE	Virtual learning environment
VR	Virtual Reality
WI	Wirtschaftsinformatik
xAPI	Experience application programming interface

Part I

Dissertation Framework

"Imagine, if you could have a superpower, what would that be?"

A question that leads to a sophisticated view of Augmented Reality

and breaking the boundaries of our physical existence.

1 Introduction

1.1 Augmented Reality in Informal Learning Environments

Augmented Reality (AR) refers to a technology-based system that enriches the real-world environment with computer-generated information and that appears to coexist in the same space as the real world (Azuma et al., 2001). Such a system can be used for many everyday situations in nearly every industry (e.g., healthcare, retail, sports, tourism, construction, maintenance, education), adding value, solving problems and enhancing the user's experience (Azuma et al., 2001). Especially in the education sector it offers new opportunities and possibilities for every age group and education level by creating new learning and experiential spaces that help learners to interact with the learning content in playful and diverse ways (Billinghurst & Dünser, 2012). Moreover, mobile AR supports learning both inside and outside classrooms to support learning in formal and informal learning environments by shaping new everyday learning environments (Bacca et al, 2015; Kumpulainen et al., 2009; Lee, 2012).

What makes AR so special is that it is a variation of virtual reality (VR) that creates digitally enriched experiences in the real world with which users can interact in real time. VR is defined as a real and a simulated environment in which a perceiver experiences telepresence (Steuer, 1992) and as computer-generated virtual environments and the associated hardware that provide the user with the illusion of physical presence within those environments (Jayaram et al., 1997). The difference between AR and VR is that VR completely immerses a user inside a synthetic environment such that the user cannot see the real world around him, while AR allows the user to see the real world, with virtual objects superimposed on or composited within the real world. Such information can be overlaid (e.g., on recognized objects) using smartphones, tablets or AR goggles as user interfaces between the real and the virtual world. AR also allows a full three-dimensional (3D) view of virtual objects and enables users to interact with them in the real environment. In that way, AR supplements reality, rather than replacing it (Azuma, 1997).

The use of digital technology to create, enhance and support hybrid environments (physical and virtual), as provided in VR and AR environments, has emerged in many sectors of education (e.g., school and higher education, vocational education and training). Already in 2016, the Goldman Sachs Group, Inc.'s Global Investment Research Report mentioned the high potential of VR and AR in the coming decade[1] and predicted that VR and AR would create new markets and disrupt existing ones. Their forecast for the education sector showed numbers growing to 15 million users and a software revenue of 700 million USD in 2025. AR has gained momentum and attention, triggered by the presentation of the first Google Glass prototype in 2012 and the furor around Pokémon GO in 2016, which had more than 21 million users six days after it was first published.

AR's application has been studied in many use cases and in settings like schools and universities, workplaces, museums and natural environments (Akçayır & Akçayır, 2017; Bacca et al., 2014; Billinghurst, Clark & Lee, 2015). The increasing number of studies in the past ten years shows that education is one of the most promising application areas for Mixed Reality (MR), that is, AR and VR (Wu et al., 2013). Educational business analysts have reported on the rise of AR and its introduction in formal (e.g., school) and informal (e.g., workplaces) learning environments. For example, the German Computer Society dedicated their DeLFI Workshops in 2018 to answering such questions as "Teaching and learning with VR and AR – what are the expectations and what works?" (Zender et al., 2018). These trends show that AR is a serious technology that has already established itself in the market, supporting teaching and learning in a variety of learning environments.

Although AR research can be traced back to the 1950s, only the recent development of pervasive, mobile technologies have made AR systems, especially mobile phone AR, affordable for the broader public (Wagner & Schmalstieg, 2009). Today, mobile AR applications leverage mobile phones' built-in cameras, GPS sensors, and Internet-based access to provide access to just-in-time information at any time and anywhere and to overlay real-world environments with dynamic, context-based, and interactive digital content.

The research process on which this dissertation is based takes place in two research environments, which function as subject areas for the artifact development. The first is the math exhibition "Matheliebe," a touring exhibition that was presented at first in 2013 at the Liechtenstein National Museum and is still touring throughout Europe. The opportunity to use such an innovative exhibition for an early AR research study conducted in a preferred informal learning environment justified the research series for this thesis. The second research environment is embedded in the context of two European research projects that began in 2016 and 2017, respectively. Both projects aim to introduce AR training in informal learning environments and in vocational education and training (VET) activities in the event technology industry.

---

[1] Goldman Sachs Global Investment Research, January 13, 2016; retrieved from https://www.goldmansachs.com/insights/pages/technology-driving-innovation-folder/virtual-and-augmented-reality/report.pdf

1.2 Motivation

The NMC Horizon Report is an annually published key source that identifies and describes emerging technologies and predicts their impact in the education sector. The venture between the New Media Consortium (NMC) and the EDUCAUSE Learning Initiative (ELI) lies behind this report, which is grounded in comprehensive research and is well recognized internationally. Already in 2012 the experts behind this report had identified AR as an emerging technology that would be highly relevant to teaching, learning, and creative inquiry and had predicted its broad adoption by 2015 (Johnson et al., 2012), a prediction amended in 2016 to the predict the next two or three years. Such mistaken predictions from experts show that there is still a need for research before AR can be fully applied in education. In the meantime, in 2017 the NMC assessment showed that AR had introduced new interfaces and was already changing how organizations were fostering innovation and doing business to increase their efficiency. The additional development of AR-related technologies like head-mounted displays (HMD) (e.g., Microsoft Hololens and Meta 2 in 2016) requires additional and research into the application of such new technologies.

Besides AR's technology aspects and expert predictions, previous research on AR implemented in education has primarily covered its qualitative aspects. Dunleavy and Dede (2014) observed in their literature review on AR teaching and learning that little research had been conducted that actively explored how mobile, context-aware AR could be used to enhance teaching and learning. Most of the existing empirical research is of a qualitative nature (e.g., observations, interviews, focus groups), conducted in formal learning environments like schools and universities, and concentrates on the usability, added value and constraints of AR in education. Many studies have been conducted in laboratory settings and have not involved pilot testing (Dey et al., 2018). Therefore, this Ph.D. projects uses both qualitative and quantitative research and implementation in both laboratory environments and field studies.

In addition, only a few quantitative studies (e.g., experiments) have set out to measure the effect of AR on learning outcomes (Bacca et al., 2014; Dunleavy and Dede, 2014). The success factors that have been most widely investigated in previous AR studies are student motivation, perceived learning and perceived learning success. While perceived learning refers to learners' having the perception that they have learned something during the activity, perceived learning success goes beyond that to describe a learners' perceptions of having successfully completed the learning activity and achieved the learning objectives. Both measures are valuable in showing the applicability of AR in the learning context, they reflect only perceptions and so stand in opposition to metrics that measure learning impact and effectiveness based on test scores and expert assessments. Only a few studies have been based on quantitative research that measures success beyond motivation, perceived learning and perceived learning success by applying valid metrics and, thus, providing instructions on how to design AR apps and design guidelines for effective app design. Against this background, this Ph.D. project focuses on a comprehensive evaluation of learning success, including both the ascertainment of the learner's perceived success factors and other metrics for the factual measurement of learning success.

Considering the growing attention from the increasing number of research studies on AR in education and the consistent references to AR in the annual Horizon Reports, it is evident that AR is the next step in technology for teaching and training. However, only a few AR case studies have been well-grounded in learning theories, rather than being creativity-based (Billinghurst et al., 2015; Wu et al., 2013). Moreover, the design of most AR applications that have been developed to support learning does not comprehensibly follow any design guidelines derived from learning theory. Although some studies consider single aspects of, for example, multimedia and mobile learning (e.g., Shukri et al., 2017) in their applications of interface design, they often fail to consider these elements from a wider perspective, such as those proposed in learning theory for collaborated, situated and experiential learning. Therefore, a central aspect of this thesis is the provision of a comprehensive design framework for AR learning application design that is grounded in learning theory.

From a research methods perspective, there are two fundamentally different views on DSR: one that concentrates on design theory and one that focuses on a pragmatic design of IT artifacts (Gregor & Hevner, 2013; Iivari, 2015). Adherents to the design theory aim to develop generalized artifacts based on theory that are later adapted and tested to solve a concrete problem from the real world. Here, the focus is on creating generic knowledge that can then be applied in other areas to gain validity. In contrast, supporters of the pragmatic design of IT artifacts place a specific problem from the real world in the foreground and develop an artifact that offers a solution to that problem, later extracting a more generally valid statement from the insights gained that can then be applied to similar tasks beyond the initial problem. With this in mind, the central research approach that this thesis follows is more theory-driven.

To summarize, the main motivation for this thesis is to contribute to the existing body of knowledge of theory-based AR application design using a mixed-method approach, including quantitative field research, and to provide theory-based findings for introducing AR in informal learning environments.

1.3 Research Objectives

Many researchers have examined the affordances and constraints of AR in learning environments (e.g., Akçayır & Akçayır, 2017; Bacca et al., 2014; Chen et al., 2016; Dunleavy & Dede, 2014; Radu, 2014; Wu et al., 2013). Most previous research has used such techniques as observations, interviews and focus groups and has concentrated on added value and limitations, thus focusing on qualitative aspects. Many researchers have seen the greatest potential of using AR in informal learning environments by facilitating voluntary and self-directed learning outside the classroom (OECD, nd), such as at exhibitions (Screven, 1993), particularly because of the technological potential for interactivity and high context awareness (Dede, 2009; Greenfield, 2009). Only a few studies have been conducted as pilot studies and to explore quantitatively the effect of AR on learning outcomes (Dey et al., 2016). Hence, this thesis focusses on the research gap where there is still little quantitative evidence for the effectiveness of AR in teaching and learning (Dey et al., 2016; Dunleavy & Dede, 2014; Wu et al., 2015).

In addition, there has been little quantitative research that has tried to measure the effect of AR on learning outcomes (e.g., using experiments) and in workplace training. However, since AR applications support context awareness and interactivity, the greatest potential for leveraging AR is in informal learning environments (Dede, 2009; Greenfield, 2009). Therefore, this Ph.D. project applies both qualitative and quantitative methods whose analyses of the application of AR in informal learning environments use both subjective and objective parameters.

The central question in this Ph.D. project, "Augmented Reality in informal learning environments – Design and evaluation of mobile applications" asks how mobile AR applications can be designed to support learning in informal learning environments. The question may be answered by proposing a design multi-level framework that addresses various design aspects that relate to the distinctive features of AR learning and that is grounded in learning theory. The research objectives for this Ph.D. thesis, its main achievements, and its contributions to the research field are summarized and illustrated in Figure 1.

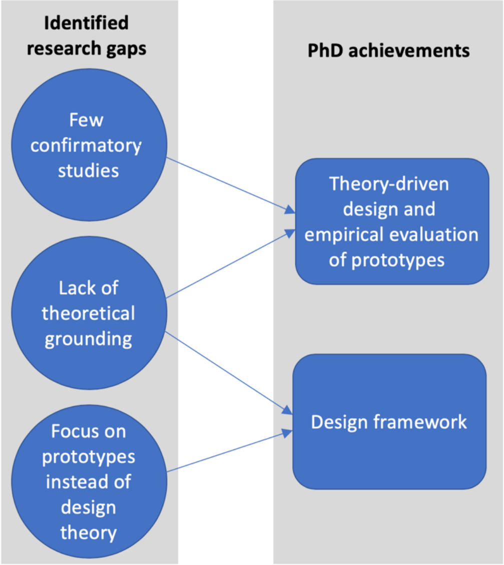


Figure 1.1: Research gaps and Ph.D. achievements.

Derived from the literature, the first research gap (RG1) is the dearth of confirmatory studies that show AR's effectiveness (or ineffectiveness) in teaching and learning. The second research gap (RG2) refers to the many studies that have lacked theoretical grounding that considers learning theory and empirical evaluation of prototypes (qualitative and quantitative). Finally, the third research gap (RG3) addresses the fact that most studies have focused on the prototypes, their development, and their application instead of following a well-founded design theory and presenting guidelines for building AR learning applications in general. To address RG2 and RG3 for AR introduced in informal learning environments, the Ph.D. project originates a design framework, following a theory-driven research approach. To address RG1 and RG2, a series of AR applications are designed and developed as artifacts to support learning in informal learning environments (e.g., museums and workplaces). Elements from the design framework in the app design are integrated and the AR applications are tested in several research activities, including field and laboratory experiments, to demonstrate their impact and outcomes. Therefore, mixed (qualitative and quantitative) methods are applied for the empirical evaluation of the prototypes, performance metrics and self-reported metrics are investigated and convincing documentation of the research results is provided.

1.4 Thesis Structure

The thesis is organized in two main parts. The first part covers the thesis framework: General insights into the research background and important findings, concepts, and discussions are presented to explain the case for this research. Then the thesis' theoretical framework is built, followed by a description of the methodology applied to achieve the research aims, including the tools used in the research activities. Finally, the research results and findings are presented.

The second part of the thesis introduces the research papers and research results from research conducted over the last three years. The order of the research papers included in this thesis is chronological and follows the stages of development toward comprehensive answers to the research questions. All papers have been published individually in English in various academic publications. The papers were formatted according to the thesis format and slightly adjusted for terminology, spelling and language, as well as the layout of figures and tables. All references were revised and presented in the citation format of the American Psychological Association (APA, 2009). The thesis also contains indexes covering content, tables, figures, and abbreviations that occur throughout the work.

2 Research Background

This chapter covers the research background related to the definition of AR, its aspects and the findings, scope and directions of research. Since the thesis focuses on AR support for the education domain, the second subsection examines the application of AR to learning in general, followed by the disaggregation of AR applied in various learning environments. The background section closes with an overview of current challenges for AR in educational settings, as proposed in the literature.

2.1 Augmented Reality

AR is commonly understood as a technology-aided extension of the perception of reality that includes all human senses (Azuma, 1997; Milgram & Kishino). AR is also often understood to mean the visual presentation of computer-generated information, such as text, pictures or videos. AR is a supplement to the real world that allows users to experience virtual and real layers in the same space by supplying information that users cannot grasp directly with their own senses, and that helps users perform tasks in the real world. In short, AR systems combine the real and the virtual worlds, support real-time interactivity and use three dimensions (Azuma, 1997). Milgram and Kishino's (1994) taxonomy of Mixed Reality (MR) illustrates the boundaries of AR between reality and virtuality (Figure 2).

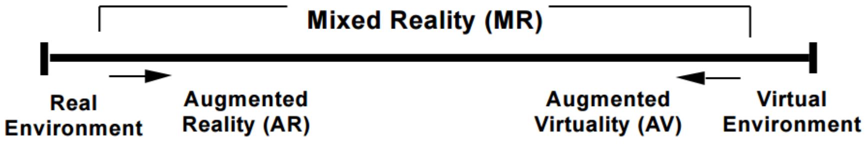


Figure 2.1: A Taxonomy of Mixed Reality Visual Displays (Milgram & Kishino, 1994).

Going back into AR history, the first technology-supported application of a jet fighter heads-up display was tested in 1958. Ten years later, Ivan Sutherland developed the first head-mounted display system to show simple, computer-generated wireframe drawings, which was used in the 1970s for commercial aviation. In a first "artificial reality laboratory," created in 1974, users were surrounded by onscreen silhouettes in an interactive environment with integrated projectors and video cameras to simulate and to create a situational experience. However, the term "Augmented Reality" was coined first in 1990 by the researcher Tom Caudell, from Boeing industries. Louis Rosenberg developed as one of the earliest functioning AR systems in 1992 a full upper-body exoskeleton that allowed the military to control virtually guided machinery from a remote operating space. Applications in theater production (1994), sports (a virtual down marker during a live National Football League game in 1998) and the military (wearable units for soldiers) and for the NASA X-38 spacecraft to enhance visual navigation by overlaying map data followed in 1999. ARToolKit, the first open-source software library for AR, was created by Hirokazu Kato in 2000, which took nine years to bring to web browsers. In 2012, the development race for new AR hardware like interfaces and devices started, and in 2014 Google announced the shipment of Google Glass devices for consumers. In the same year, vendors like Magic Leap started before Microsoft (Hololens) and Meta 2 released their developer kits[2].

In the last decade, researchers proposed shortened definitions of AR that are based on the single feature of superimposing virtual information onto real objects. For example, El Sayed, Zayed, and Sharawy (2011) argued that AR only adds virtual objects to real scenes, Chen and Tsai (2012) supported indicated that AR just allows interactions with virtual objects in two-dimensions and three dimensions in a real-world environment, and Cuendet, Bonnard, Do-Lenh, and Dillenbourg (2013) stated that AR is simply a projection of digital elements on real-world objects. However, AR is not limited to the sense of sight or restricted to special display technologies like HMDs, smartphones and tablets; it comprises also the ability to address other senses (e.g., hearing, touch and haptics, smell) and has the ability to remove real objects by overlaying virtual ones, which is known as mediated or diminished reality (Van Krevelen, 2010). AR is an extension of VR (Wojciechowski & Cellary, 2013).

Types of AR:

For the presentation of the digital content superimposed in the real-world environment, AR uses several types of displays and hardware. Figure 3 shows the types of AR that are commonly accepted by the research community.

To interact with AR systems, some kind of display is necessary. Basically, there are three types of displays: head-worn glasses (e.g., HMDs and goggles), hand-held devices (e.g., smartphones and tablets) and spatial projection display systems (Syberfeldt et al., 2016). A general distinction is made between whether a video-based system (for example) merges the real and virtual world into a completely digitally generated view or just overlays virtually generated views onto the real-world using transparency and look-through technologies.

AR also requires a reference point in a real environment that is used for positioning, orientation and navigation. Such anchors can be based on markers like images or special codes (e.g., QR-codes) that are placed on objects in the real world. In contrast, marker-less AR functions with a dedicated image, surface or object in the real world that the system recognizes and uses as a

---

[2] Augment News, “The Lengthy History of Augmented Reality,” infographic, accessed June 6, 2018, http://www.augment.com/blog/infographic-lengthy-history-augmented-reality/

trigger to navigate and/or interact with the AR system. A third type of AR is location-based, supported either by indoor tracking systems or global positioning systems (GPS), where a user's position is used as the activator. Most spatial AR is implemented in connection with projection-based systems. Spatial AR is frequently used in combination with a fighter pilot's or a car's head-up display system. Finally, AR uses two- and three-dimensional object recognition to gather information about the user's or system's environment, as implemented in, for example, a car to recognize boundaries and lines in the traffic lane and to assist in precarious situations. The types of AR that are mostly applied in education are marker-based AR (59.38%), location-based AR (21.88%), and marker-less AR (12.5%) (Bacca et al., 2014).

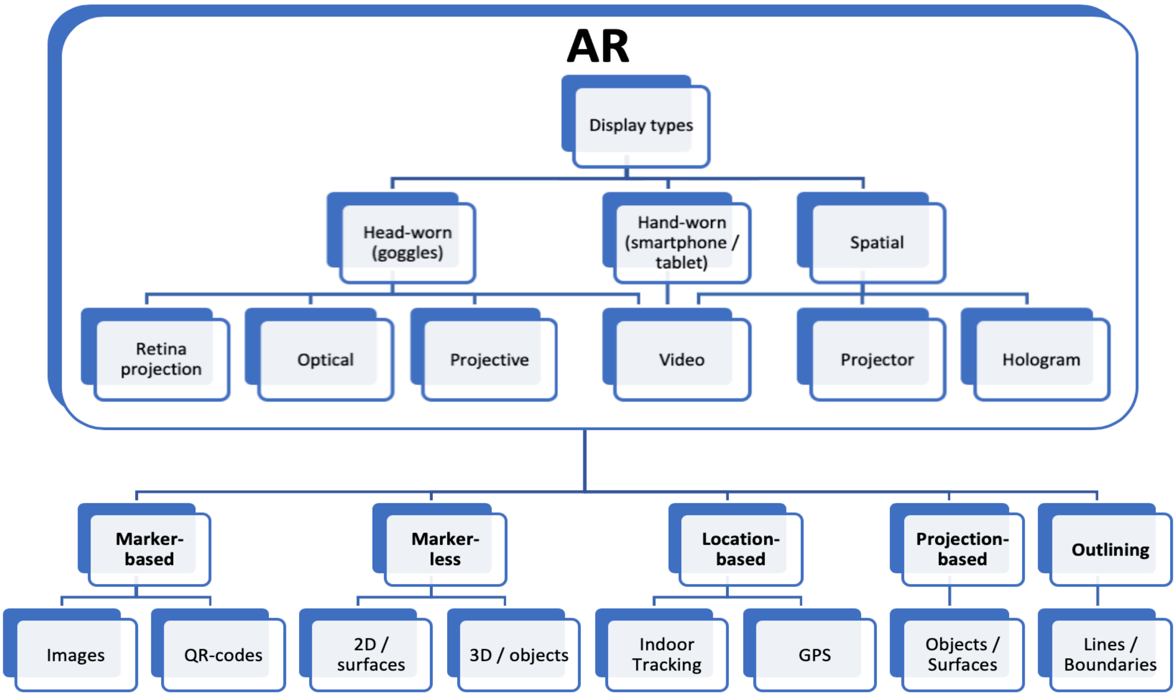


Figure 2.2: Combination of types and systems of AR (sources: Cheng et al. (2013), Syberfeldt et al. (2016))

AR's technological maturity and potential enable developers to design increasingly convincing experiences. The hardware and software available today make it easy for anyone to create appealing applications (Billinghurst, 2014). Research and development in the field of AR continues to grow in terms of technology (e.g., HMDs, goggles, wearables) and its application (e.g., software, use cases). In their review, Papagiannakis et al. (2008) emphasized that the merging of wearable computing, wireless networking and mobile AR interfaces lead to a new type of computing, "augmented ubiquitous computing," which facilitates new experiences that cannot easily be observed with the naked eye (Wu et al., 2013) or met with any other real-world sense. Moreover, the rise of AR, together with advances in complementary technologies like tactile networks, artificial intelligence, cybernetics and ubiquitous computing, will have a significant impact on the development and interactions of future computers (Van Krevelen, 2010).

2.2 AR and Learning

A large and growing body of literature has examined the use of AR-based technologies for the education sector, especially for teaching and learning in natural science, medicine, engineering, languages, history and the arts and in various learning environments like kindergartens, schools, universities, laboratories, museums, parks, and zoos (Dunleavy & Dede, 2014; Wu et al., 2013). The most frequently applied research methods are mixed methods, followed by qualitative-exploratory-case study, quantitative-descriptive research, qualitative-exploratory research, pilot study, quantitative-explanatory research and causal research. Most studies have used questionnaires, interviews, surveys and case observations for their data collection. More recent attention has focused on the implementation of AR training in vocational education and training (VET) and in workplaces (Palmarini et al., 2018).

A Google Scholar-based search for "augmented reality learning" provides almost 700,000 results[3]. Although this search (and, thus, not all findings) does not fulfill the requirements for a systematic review as described by Gough et al. (2017), it demonstrates the vast amount of available information and research articles in the research field. Mark Billinghurst, one of the leading scientists in AR research, posted in mid-2018, that nearly 20,000 research papers can be found on Scopus that use the term "Augmented Reality" in their titles[4]. However, a number of valuable literature reviews have been published that provide a comprehensive overview of the central research topics. Worth mentioning at this point are Bacca et al.'s (2014) review and Billinghurst et al.'s (2015) survey of AR, which summarizes almost fifty years of research and development.

Dunleavy and Dede (2009) provide insights into AR teaching and learning, focusing on AR that uses mobile, context-aware technologies (e.g., smartphones, tablets), thus enabling AR users to interact with digital information that is embedded in physical environments and in both formal and informal learning environments. Focusing on publications that compare student learning with AR versus non-AR apps, Radu (2014) elaborated on the factors that influence learning in AR, such as content and representations that appear at appropriate times and in appropriate spaces, learners' interactions with 3D simulations, collaboration, and applied educational concepts. Diegmann et al. (2015) identified fourteen benefits clustered into six groups: state of mind (increased motivation, increased attention, increased concentration, increased satisfaction), teaching concepts (student-centered learning, collaborative learning), presentation (increased details, information accessibility, interactivity), learning type (improved learning curve, increased creativity), content understanding (improved development of spatial abilities, memory), and reduction of costs. Akçayır and Akçayır (2017) described the current advances and challenges of AR education, emphasizing challenges in terms of cognitive

---

[3] Google Scholar search conducted on 14 August 2019 with the search term "augmented reality learning."

[4] Source: https://medium.com/@marknb00/where-in-the-world-is-ar-vr-research-happening-ddebbdc6436b retrieved on 14 August 2019.

overload and usability and that AR designers should consequently implement empirically proven design principles that focus on AR use and educational outcomes. As already stated, education is one of the most promising application areas for AR, and many researchers have examined its affordances and constraints in various learning environments (e.g., Akçayır & Akçayır, 2017; Bacca et al., 2014; Chen et al., 2016; Dunleavy & Dede, 2014; Radu, 2014; Wu et al., 2013). In their 2005 annual Horizon Report[5], the NMC provided predictions for the application of AR in educational settings based on expert opinions. While AR's initial application was seen as a way to visualize large data sets in educational settings (2006), aspects like collaborative experiences (2007), integration of smart objects (2008), gesture-based computing (2009) and simple AR with easy accessibility and decreased requirements of specialized equipment (2010) followed as AR developed.

The maturity of technology AR applications use, such as smartphones and tablets, led the Horizon Report's expert panel to state that AR has strong potential to provide both powerful contextual and in situ learning experiences. Early in 2011, they predicted that AR would be a game-changer for education and would be used for visual and highly interactive forms of learning by overlaying data onto the real world, simulating dynamic processes, augmenting books by applying active and interactive technology, and aligning with situated learning. A significant benefit of AR is that it transfers learning from one context to another while breaking the boundaries between formal and informal learning and contributing to the evolution of a learning ecology that transcends educational institutions. The considerable potential of AR for just-in-time learning and exploration and for annotating existing spaces by overlaying extended information (e.g., in museums), as well as its positioning ability and ubiquitous services was already identified in 2012.

Further evolutions of AR in 2013 toward mobile and gaming applications, the introduction of elements of learning analytics (e.g., dashboards) and the integration of hybrid and collaborative learning in 2014 led the experts to see AR is a key emerging technology for visualization technologies in education. AR also has the ability to blend formal and informal learning and to support the Internet of Things. New advancements in VR technology in 2014 brought fresh perspectives, so big players in the digital industry, like Facebook and Microsoft, invested in the development of today's leading HMDs for AR and VR applications like the Microsoft Hololens and Oculus Rift. Today, AR has the potential to impact content mediation, content delivery and transformation of online education significantly and to support the growing focus on measuring learning and collaborative learning in blended learning environments. Moreover, AR, with its contribution to the design, creation and definition of prospective Virtual Learning Environments (VLEs) will redesign learning spaces to enable learners to have authentic learning experiences.

---

[5] The Horizon Report is available through the EDUCAUSE library, accessible via https://library.educause.edu/

The number of studies on the use of AR in education has grown steadily since 2005 (Chen et al., 2017; Johnson et al., 2012). Researchers have presented a wide range of positive impacts and affordances and have discussed the effectiveness of using AR in education in terms of increased content understanding (e.g., contextual visualization, learning of spatial structures, language associations), better learning performance (e.g., retention success, long-term memory retention), and improved soft skills (e.g., collaboration skills, enjoyment, motivation, positive attitudes, engagement) (Bacca, 2014; Chen et al. 2017; Radu 2012, 2014; Santos 2014). In their study, Dunleavy, Dede, and Mitchell (2009, p. 20) stated that AR's most significant advantage is its "unique ability to create immersive hybrid learning environments that combine digital and physical objects, thereby facilitating the development of processing skills such as critical thinking, problem-solving, and communicating through interdependent collaborative exercises." Furthermore, Sotiriou and Bogner (2008) found that AR helps learners to acquire better investigation skills. However, Dey et al. (2016) contended that there is a need for studies that are conducted as pilot studies and that explore quantitatively the effect of AR on learning outcomes.

Even though AR can be accessed with various technologies, such as tablet PCs and HMD, the delivery technology for AR in education that is usually preferred is mobile devices (Akçayır & Akçayır, 2017), since today's learners are familiar with this technology and use it in their daily lives. Mobile devices provide many advantages that support AR applications, as they are easy to use, cost-effective (Furio et al., 2013), and portable (Chiang et al., 2014); they provide a high level of social interactivity and independent operability (Hwang et al., 2012); and they are useful for outdoor activities (Chiang et al., 2014), thereby contributing to users' collaboration skills (Bressler & Bodzin, 2013; Yu et al., 2009) and facilitating meaningful learning (Bronack, 2011).

AR has also demonstrated some negative impacts, such as attention tunneling, usability difficulties, ineffective classroom integration, difficulty responding to learner differences (Radu, 2012, 2014), and increasing lecture time (Munoz-Cristobal et al., 2015). Although teachers recognized the benefits of using AR in classrooms, they complained about having little control over the content in the system so need for adaption to the needs of their students was extensive (Wu et al., 2013). In addition, technical issues caused by the devices that provide AR applications can lower the motivation to learn (Wu et al., 2013), as can the handling of bulky AR technologies like HMDs (Yu et al., 2009). In summary, even though AR offers new possibilities for the education sector, it also comes with challenges.

What all levels of education have in common is that teaching takes place in various learning environments, that it is based on commonly accepted learning theories, and that learning success is assessed by applying metrics. Since this thesis focuses on AR applied in teaching and training, the following sub-sections outline relevant aspects of AR and consolidate the scope of the research field into the educational sector.

2.2.1 AR in Formal Learning Environments

The twenty-first century has been called “the digital age,” as digitization affects nearly every industry, including the education sector. However, the educational system still uses traditional approaches to teaching and learning, with its roots in the eighteenth century. Even though technology like the internet has created the opportunity to learn about anything at any time, there is increasing interest in learning that is self-directed and less structured, aspect-oriented and less formally organized. According to many education experts, a blend of formal and informal methods of teaching and learning in diverse environments supports a learner’s creativity, curiosity, and motivation and fosters experimentation (Johnson et al., 2015). With the evolution of VR and AR, a virtual learning environment (VLE) has arisen as a new generation of learning environments (Dillenbourg et al., 2002; Van Raaij et al., 2008).

In their literature review, Bacca et al. (2014) analyzed target groups in which AR studies were carried out in schools. Their results showed that most studies were conducted at primary and lower secondary education levels, followed by the bachelor’s (or equivalent) level, the upper secondary education level, and a few in informal learning environments and short-cycle tertiary education. Both Akçayır and Akçayır (2017) and Chen et al. (2017) confirmed these findings in their reviews, and Chen et al. (2017) highlighted that the research methods applied most often in case studies are mixed methods (40%), followed by quantitative research methods (33%) and qualitative research methods (7%). The tool implanted most frequently for data collection were tests (47%), tests with pre-tests and post-tests (29%), interviews (31%), questionnaires (29%), video observations (18%) and surveys (16%), although some studies applied several tools at once. According to their results, image-based AR was preferred over location-based AR, and smartphones and tablets were used most often.

The literature presents no evidence regarding the extent to which AR is involved in the formal learning process with respect to a review of learning outcomes on which a qualification, certificate or similar can be issued. Only a few studies have been carried out in the VET sector, although VET institutions are promising research partners for demonstrating the possibilities of AR learning for improving and acquiring professional competencies, particularly in terms of cost reductions when creating learning experiences that require expensive learning materials and that place in special training environments are no longer necessary (Bacca et al., 2014).

2.2.2 AR in Informal Learning Environments

Learning takes place in various learning environments, including face-to-face (e.g., traditional classroom), online, and hybrid forms. However, the focus of this thesis project is on informal learning environments, so some boundaries between informal and formal learning environments must be established. According to the European policy document model published by the European Commission (2001), formal learning is typically provided by an education or training institution, structured in terms of having learning objectives, learning times and learning support, and leads to a learning certification. Formal learning is intentional from a learner's perspective, while informal learning results from daily life activities related to work, family or leisure. Informal learning is not structured and typically does not lead to a certification; it is less intentional than formal learning and often even unintentional. The overlapping space between formal and informal learning is described as non-formal learning, that is, learning that is not typically provided by an education or training institution and does not lead to a certification but is structured in terms of having learning objectives, learning times and learning support. It is intentional from a learner's perspective (EC, 2001).

Informal learning environments are cited in the literature as those out-of-class learning locations that are not necessarily designed to host formal learning. A key source is "Learning Science in Informal environments – People, Places, and Pursuits" (National Research Council, 2009). According to that study's definition, informal learning environments are places that are shaped by everyday experiences, such as activities around a visit to a science center, zoo, aquarium, botanical garden, planetarium, museum, exhibition, theme park, or historical site. Informal learning occurs equally at home, in schools and other education environments, and in workplaces.

Definitions of formal, non-formal and informal learning are contested in some research circles, since their interrelationships are complex (Colley et al., 2003; McGivney, 2002). In particular, the efforts of the European Commission in the last decade to validate, recognize and credit informally acquired skills and competencies has strengthened the controversy around this topic. Although this discussion is mentioned for the sake of completeness, it does not necessarily contribute to the objectives of this thesis. AR can shape its own augmented and virtual learning environment to create personal learning experiences in both formal and informal learning environments and to support learning on demand anywhere and anytime.

2.2.3 Learning Theories

Since AR refers to technologies that blend real-world environments and context-based digital information, its positive effect on learning can be theoretically grounded in a cognitive (Bujak et al., 2013) and a constructivist (Dunleavy & Dede, 2014) perspective on learning. AR also addresses other theories related to constructivism, such as experiential learning theory, animate vision theory and situated learning theory (Santos, 2014).

Learning theories can be presented from behavioral, cognitive and constructivist perspectives (Illeris, 2009; Parhizkar et al., 2012). Behaviorism focuses on controlling and modifying students' behavior in their learning process, and learning is a result of the association formed between stimuli and responses (Illeris, 2009). From a psychological perspective, behavior is a basic element involved in all learning activities (Skinner, 1974). In AR-based learning environments, the user constantly interacts with the environment and immediately receives feedback from it, which determines how to proceed, but always in connection to existing knowledge. AR can support behavioral learning by, for example, demonstrating a particular behavior in a specific situation, showing how to complete a single task, or explaining how to behave in an appropriate manner.

Cognitivism considers internal cognitive structures to explain how information is received, organized, stored and retrieved. The mind is understood as an information-processor (Anderson, 1983; Hutchins, 1995; Illeris, 2009; Wenger, 1987). From a pedagogical perspective, processing and transmitting information takes place when communication, explanation, recombination, contrast, inference and problem-solving are involved in knowledge acquisition. An applicable theory for AR that is based in cognitivism is the cognitive theory of multimedia learning (CTML) (Mayer, 2009). Mayer presents twelve principles that are grouped according three categories: reducing extraneous processing, managing essential processing and fostering generative processing. Based on AR's nature, well designed applications incorporate the multimedia principle (combining words and pictures), the spatial and temporal contiguity principle (presenting words and pictures simultaneously), the modality principle (preferring graphics and narrations over animation and on-screen text) and the signaling principle (adding cues that highlight the organization of essential material) (Chiang et al., 2014; Santos et al., 2016).

Constructivism is a learning paradigm in investigating processes that supports learners in building their own mental structures when they interact with an environment (Illeris, 2009; Richard, 2015). It equates learning with creating meaning from experience, as learning is more meaningful to students when they interact with a problem or concept. Students are actively engaged in problem-solving and motivated through meaningful contexts to gain new knowledge that is based on their own experiences. Such learning tasks can be implemented by using high-order thinking skills to transfer knowledge to new situations as is done, for example, in simulated worlds, role-plays, debating, cooperative learning, and self-directed task-based learning (Illeris, 2009; Papert, 1980; Piaget, 1954). Learning theories that are interdependent with and follow the constructivist paradigm include situated learning, experiential learning, game-based learning and simulation, and collaborative and mobile learning (Dunleavy & Dede, 2014; Herrington et al., 2009).

2.2.4 Metrics for Learning Success

In traditional classroom environments, the assessment used to determine learning success often follows formative and summative approaches. Formative assessment is conducted by a teacher every day and is an ongoing process, based on, for example, observations and informal tests. In contrast, summative assessments are given at certain times, are planned, and follow predefined rules (e.g., exams) (Garrison & Ehringhaus, 2007). Since most AR experiments in education have lasted less than a month (Garzón & Acevedo, 2019), the evaluation of learning success in such AR-related usability studies is short-term-oriented.

Two of the most frequently investigated aspects of the application of IT artifacts are perceived usefulness and perceived ease of use, as explained by Davies et al. (1989) in their technology acceptance model (TAM; Figure 3). TAM is often consulted in IT research, as it is a well-established and still valid model that specifies a central theory from the discipline of business informatics. Central questions derived from this model are given in terms of a system's usability and aspects of the system that affect a user's attitude and behavior. What makes a system user-friendly and which aspects of the system require attention should be considered in early development stages to raise the users' acceptance of the technology or system. However, perceived usefulness, perceived ease of use, perceived learning and perceived motivation are parameters that determine a user's perception but do not necessarily evaluate learning success based on independent indicators like the number of correct answers or fulfilled tasks, increased retention, and time to completion.

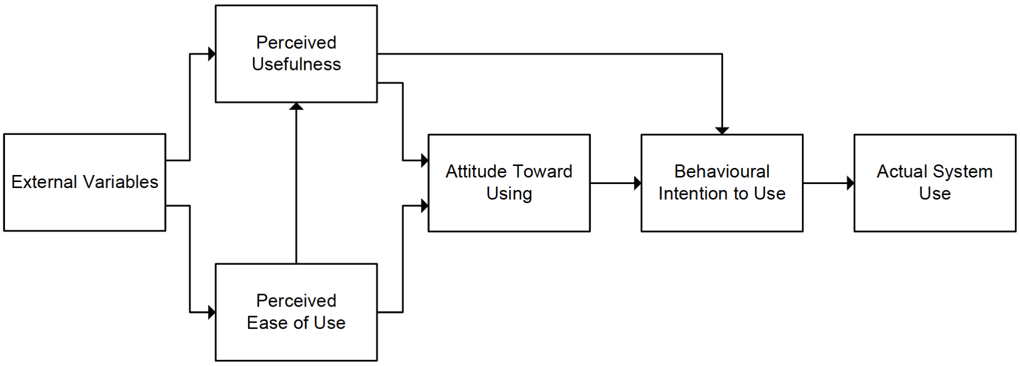


Figure 2.3: Technology Acceptance Model (Davies et al. 1989, p. 985)

By extension, Venkatesh and Davis (2000, p. 187) defined perceived usefulness as "the extent to which a person believes that using the system will enhance his or her job performance" and perceived ease of use as "the extent to which a person believes that using the system will be free of effort." However, a system's usefulness differs from its ease of use. Usefulness in the context of AR learning applications refer to the extent to which a user learns something by using the application, while ease of use is based on a user's individual abilities and inclinations. For example, an experienced AR user will probably experience the ease of an AR application's use differently than an inexperienced user will.

To identify and extract measures of learning success from the literature, Tullis and Albert (2013) categorized metrics along the dimensions of "performance" and "self-reported," where performance metrics are objective methods and self-reported metrics are largely subjective. Performance metrics are collected either through observation methodologies or by automated data collection and do not consider participants' opinions, while self-reported metrics focus on the reliability of a user's opinion.

As for data sampling and evaluation, Lim et al. (2019) grouped types of evaluations found in their literature review into within-subject evaluations, between-subjects evaluations and a combination of both. Within-subject evaluations are repeated measurements on experimental participants that are evaluated on more than one tested item, while between-subjects evaluation compares a single evaluation's results between participants (Tullis & Albert, 2013). Within-subject evaluation does not necessarily require a large sample size, but it entails the risk of participant carryover effects and experience biases. Between-subjects evaluations are based on a clean data collection, so while they reduce risks and side effects, more effort for data collection is required.

Metrics like effectiveness, efficiency, satisfaction, and learnability are some of the most frequently applied metrics in evaluating learning from AR (Tullis & Albert, 2013). Lim et al. (2019) provided an overview of metrics used in studies that have investigated mobile-based AR learning applications and categorized the metrics according to performance metrics vs. self-reported metrics, and within-subject evaluation vs. between-subject evaluation. They summarized eighteen categories of metrics, accompanied by the multiple interchangeable terminologies from the literature in each category. Their report on performance metrics covered accuracy for effectiveness and for efficiency. Table 2.1 shows these categories and their attendant metrics.

Table 2.1: Metrics for mobile AR learning applications (Lim et al., 2019)

Within- subject evaluation			Between- subject evaluation		
Usability/Experience	Learnability	Content	**Efficiency**	Fun / Amusement	Emotion
Motivation	Engagement	Adaption	Usefulness	Cognitive Load	Security
Satisfaction	**Effectiveness**	Behavior	Preference	Interface Design	Other

The results from a literature review that is grounded in learning theories, as provided in paper P.3 of this thesis, show that measures undertaken to investigate the influence of AR in learning predominantly apply pre- and post-test designs, relate to cognitivist theory, and compare memorized information before and after the intervention. Metrics applied in studies include the number of trials performed, time-related aspects of the lesson (e.g., study time, response time), and the number of attempts in a given timeframe. Metrics that can be considered new and are identified in Lim et al.'s (2019) review are escapism, facilitating conditions, bundled identification, pragmatic quality, stimulation, novelty, price value, and social influence (Lim et al., 2019).

2.3 Challenges and future directions for AR in educational settings

Besides the advantages that AR holds for the educational sector, some challenges reported in the literature remain to be overcome. One such challenge has to do with the use of performance metrics (Lim et al., 2019). Even though AR facilitates data collection for a continuous evaluation of its application in a particular setting, new models and methodologies remain to be proposed for using beneficial performance metrics that eliminate the limitations of objective measures. This challenge suggests that future research be conducted as pilot studies with larger sample sizes to explore quantitatively the effect of AR on learning outcomes (Dey et al., 2016). Future trends for AR in teaching and training will focus on measurement and evaluation of learning in personalized student learning experiences and will consider the acquisition of skills, competencies, creativity, and critical thinking (Becker et al., 2017).

The literature has suggested several directions for future research. In particular, AR-related learning outcomes like decreasing the cognitive load or enhancing spatial abilities require further investigation (Bacca et al., 2014), as demonstrated in, for example, Santos et al. (2014) by superimposing virtual text, images and videos to reduce cognitive load in the limited working memory. The use of mobile AR in educational games (Furió et al., 2013) promises to make valuable contributions to researchers who aim to explore AR's features, advantages and drawbacks (Bacca et al., 2014). However, a comprehensive explication of the educational effects and implications of AR is still missing (Radu, 2012). Aspects of interest for future research, as discussed in paper 3 of this thesis, include extensive subject matters, lengthening the research timeframe, considering teachers' requirements, incorporation of interactive strategies to enhance first-hand experiences and interactions, considering differences in cognitive processing, psychological immersion between AR and reality settings, individual interactions, sense of identity, adaptive applications, AR classroom design and evaluation, and the teacher's role in AR educational settings. In particular, future research should address the affordances and characteristics of AR in educational settings that differentiate this technology from others (Bacca et al., 2014), including AR's strengths for offering an inclusive experience for people with disabilities.

From the technology side, Azuma et al. (2001) reported on optical challenges for AR HMDs that remain, for example, in terms of resolution of the displays, distortion, safety, eye-offset (position of the video camera that differ from the position of eyes in a natural environment) that leads to difficulties in hand-eye coordination, limited field of view, delay between real and virtual views, and size and weight. Location-based AR also often lacks in tracking, such as static errors in the tracking system that lead to mechanical misalignments or incorrect viewing parameters, and dynamic errors like delays and motion lags (Cheng & Tsai, 2013; Chiang et al., 2014). Although technology will continue to evolve, and it is expected that the imperfections of today's devices will soon be remedied, future research should consider these challenges, especially for location-based AR.

Many studies have investigated the application of AR technologies in a specific context. According to Bacca et al. (2014) just few studies have considered the factors of accessibility and usability. The most frequently reported challenge for AR lies in its usability (Akçayır & Akçayır, 2017; Chang et al., 2014; Cheng & Tsai, 2013; Munoz-Cristobal et al., 2015). Usability issues lower the technology acceptance and affect educational effectiveness (Chang et al., 2014). Weak usability also leads to longer activity times, as reported in a case study from Gavish et al. (2015), who compared the training times between an AR-assisted group and a control group that used traditional tools. Chiang et al. (2014) emphasized that at least some user guidance or perhaps instant hints could be provided to AR users to overcome basic usability issues. Further research needs to be undertaken that focuses on usability for AR applications, especially in education, and that prepares guidelines to support the design of AR-based educational settings (Bacca et al., 2014).

Bujak et al. (2013) emphasized that learners skills must be considered to understand the conceptualization and the arrangement of AR learning experiences, especially the user experience and knowledge construction processes in AR applications (Lin et al., 2013) for, for example, the design of multisensory experiences and their impact on learning outcomes (Ho et al., 2011). In addition, to develop new methods for the design of interactive 3D learning environments (Chang et al., 2014), the underlying concepts should be further explored and evaluated. Bacca et al. (2014) emphasized that the conceptualization and construction of tools for teachers to create content requires their involvement in the design of the AR application. Moreover, the investigation of aspects like cost efficiency, as provided by, for example, impact studies of AR introduced with large groups or in laboratories to reduce training costs, should be considered.

3 Research Design

This chapter explains the theoretical framework and the research design applied in the Ph.D. project. The research question introduced in section 1.3 is linked to the research process, and the work processes and the research methods used therein are presented, along with the processes for the design, creation, evaluation and analysis of the IT artifacts developed for the underlying research purpose. Finally, the ontological and epistemological positions that underlie this thesis are outlined.

In discussing design science versus design theory, Walls et al. (1992, as cited in Walls et al, 1992, p. 48) contended “design practice creates ‘things that serve human needs’, while design science should create the theoretical foundations for design practice.”.

3.1 Design Science Research

Design science research (DSR) is a commonly accepted research approach for use in information systems (IS) research (Hevner & Chatterjee, 2010; Kuechler & Vaishnavi, 2008). To address unresolved challenges in IS, DSR provides effective and efficient processes and solutions that can deliver new and innovative IT artifacts (Iivari, 2015). However, there are two strategies to the approach for developing a DSR-based research process: one that focuses on the construction of IT artifacts that can be general solutions and possibly adapted and applied to construct specific solutions, and one that focuses on a particular problem in a particular domain and tries to solve it by creating a non-generic IT artifact that may be used to derive a generally valid solution concept later on. In addition, there are two fundamentally different views on DSR, which can be broken down into followers of design theory and those who focus on a pragmatic design of IT artifacts (Gregor & Hevner, 2013; Iivari, 2015).

The Ph.D. project documented in this thesis follows Walls et al. (1992) approach for theory-guided design and the understanding of DSR as “research with design as a method of investigation” (Iivari, 2015, p. 108), so it aims to produce new, innovative meta-artifacts and follows the epistemology of utility. Since this thesis guides the development and implementation of AR applications in informal learning environments (e.g., at museums and workplaces) and the development of a design framework to support AR development as a main result, the overall research design follows Hevner et al.’s (2004) DSR guidelines. The research process has five sub-processes, all of which implement the DSR guidelines.

Hevner et al. (2004) postulated that their seven guidelines for design science research could be used to create a construct, model, method or instantiation as an artifact (guideline 1). The thesis considers both a construct for the design, development and evaluation process and the development of several AR learning applications to demonstrate its effectiveness in informal learning environments. The second guideline addresses “research relevance” and targets the central objective of DSR, which is to develop a technology-based solution to an important problem. Chapter 1 of the thesis prepares the basis for satisfying guideline 2. Guideline 3 requires that the utility, quality and efficacy of a design artifact be demonstrated in an appropriate way using valid research methods. Guideline 4 says that, to present the contribution of the work to the research community, effective DSR-based projects should provide clear and verifiable contributions to the areas of the design artifact, its foundations and methodologies. To satisfy the requirements for research rigor expressed in guideline 5, DSR requires the application of rigorous methods in both the construction and evaluation of the designed artifact. Since the search for a good artifact design follows an iterative process in which the design task often involves the creation, utilization and assessment of heuristic search strategies, design should be constructed as a search process, as guideline 6 suggests. Finally, in guideline 7, DSR demands the communication of research in a way that enables practitioners to take advantage of the benefits offered by the artifact and researchers to build a cumulative knowledge base for extension and evaluation of the artifact. Table 3.1 summarizes the guidelines.

Table 3.1: Design Science Research Guidelines (Hevner et al., 2004, p. 83)

Guideline		Description
1	Design as Artifact	DSR must produce a viable artifact in the form of a construct, a model, a method, or an instantiation.
2	Problem Relevance	The objective of DSR is to develop technology-based solutions to important business problems.
3	Design Evaluation	The utility, quality, and efficacy of a design artifact must be rigorously demonstrated via well-executed evaluation methods.
4	Research Contribution	Effective DSR must provide clear and verifiable contributions in the areas of the design artifact, design foundations, and/or design methodologies.
5	Research Rigor	DSR relies on the application of rigorous methods in the construction and evaluation of the design artifact.
6	Design as Search Process	The search for an effective artifact requires using available means to reach desired ends while satisfying laws in the problem environment.
7	Communication of Research	DSR must be presented effectively to both technology-oriented and management-oriented audiences.

In addition to these guidelines, Hevner et al. (2004) provided a set of methods with which to satisfy the requirements for design evaluation as stipulated in their guideline 3 for DSR implementation (Table 3.2). Since evaluation is a crucial component of the research process, this source is applied to all sub-projects of this thesis.

Table 3.2: Design evaluation methods (Hevner et al., 2004, p. 83)

Evaluation		**Description**
1	Observational	Case study: Study the artifact in depth in a business environment. Field study: Monitor the use of the artifact in multiple projects.
2	Analytical	Static analysis: Examine the structure of the artifact for static qualities (e.g., complexity). Architecture analysis: Study the artifact's fit into technical IS architecture. Optimization 1: Demonstrate inherent optimal properties of the artifact; or Optimization 2: Provide optimality bounds on the artifact's behavior. Dynamic analysis: Study the artifact in use for dynamic qualities (e.g., performance)
3	Experimental	Controlled experiment: Study the artifact in a controlled environment for quality (e.g., usability).
4	Testing	Functional testing: Execute the artifact's interfaces to discover failures and identify defects. Structural testing: Perform coverage testing of some metrics in the artifact implementation (e.g., execution paths)
5	Descriptive	Informed argument: Use information from knowledge base to build a convincing argument for the artifact's utility (e.g., relevant research). Scenarios: Construct detailed scenarios around the artifact to demonstrate its utility.

3.2 Research Process

The research process that guides the thesis project and sub-projects are oriented toward Peffers et al.'s (2008) model, as shown in Figure 3.1. Notably, Gregor and Hevner (2013) accepted Peffers et al.'s research process as a useful synthesized general model that is compatible with their underlying ontological perspective in DSR. Based on the six research steps — (1) problem identification and motivation, (2) definition of the objectives of a solution, (3) design and development, (4) demonstration, (5) evaluation, and (6) communication — that describe the overall research process, Figure 3.1 illustrates the objectives for each step and describes their implementation in the thesis project and in each sub-project.

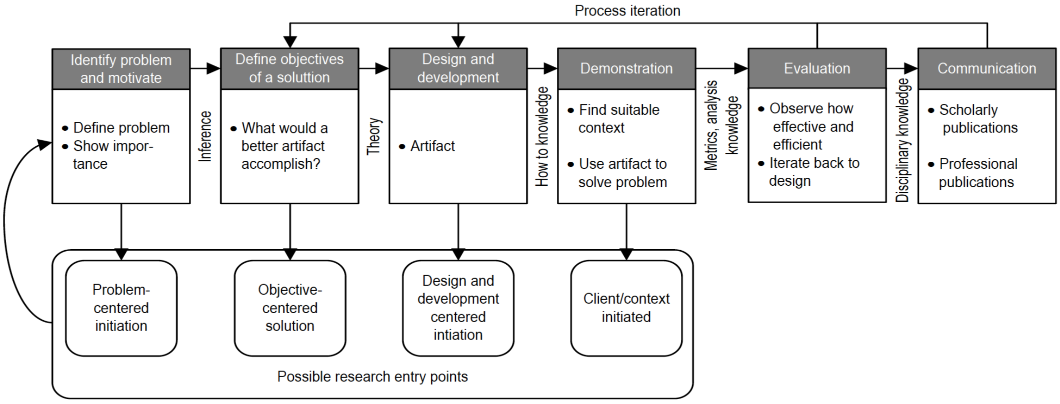


Figure 3.1: DSR process model (Peffers et al., 2008, p. 54)

Identify the problem and motivate: The initial step in the process meta-model covers the activities that define the research problem to address in the research project and justify the value of a solution. To proceed in that way, it is helpful to conceptualize and subdivide the problem into sub-tasks to understand the complexity (or lack of complexity) of a holistic solution to the problem so researchers and the audience can follow the reasoning for finding a solution and to obtain higher acceptance for the results.

Define the objectives of a solution: The derivation of the goals for a solution emerges from the problem definition. The goals may be quantitative, such as describing how a solution could solve a problem better than already existing solutions can. Qualitative goal descriptions can show how a new artifact supports solutions to a problem that has not been addressed. Such descriptions require a comprehensive knowledge of the state of the problem and the existing solutions and their effectiveness. Goals should be derived rationally from the problem specification.

Design and development: An artifact that solves the problem may be a well-defined construct, a model, a method, or an instance (Hevner et al., 2004). In this step, the desired functions are determined and the artifact's architecture is determined before the actual artifact is created. Comprehensive theoretical knowledge for finding solutions is required.

Demonstration: Experiments, simulations, case studies, evidence, and other appropriate activity may be used to demonstrate the artifact's effectiveness. This step requires that researchers have extensive knowledge about how the artifact solves the problem.

Evaluation: To be able to compare the objectives of a solution that the artifact achieves with the intended goals and to observe and measure these objectives in the application, knowledge of the relevant metrics and analysis techniques is required. Depending on the problem addressed and the artifact created, the evaluation can use items that describe the functionality of the artifact, analogous to the solution targets from Activity 2 (define of objectives of a solution), such as objective quantitative performance indicators like compliance costs, user satisfaction, customer feedback, and simulation results. Based on the findings, whether to return to Step 3, to continue to try to improve the effectiveness of the artifact, or to make suggestions for improvement for subsequent projects can be decided.

Communication: Profound knowledge of the appropriate way to communicate is required to communicate the problem, the solution obtained, the importance, utility, and novelty of the artifact, the accuracy of the design, its effectiveness for audiences like researchers and practicing professionals, and the associated knowledge gained. In particular, the structure of this process can be used to develop scientific research publications based on an empirical research process of problem definition, literature search, hypothesis development, data acquisition, analysis, results, discussion and conclusion.

3.3 Research Methods

The research method is a constitutive element in a research project that relates to a specific method of data collection and analysis (Creswell, 2003), and provides guidelines for its implementation. The literature distinguishes among qualitative, quantitative, and mixed methods, the last of which refers to the sequential or concurrent application of quantitative and qualitative research methods (Venkatesh et al., 2013). Validity and reliability are two key aspects of research acceptance. While validity refers to the legitimacy of the findings, reliability is connected with the quality of measurement and is a precondition for convincing quantitative research (Straub et al., 2004).

According to Gregor and Hevner (2013), research based on DSR should explain the specific DSR approach that is adopted for the research design, with reference to the referring authorities (e.g., Gregor & Hevner, 2013; Hevner et al., 2004; Peffers et al., 2008). The clear reasons for applying the chosen methods must be stated in the method section and to satisfy its role in the evaluation step of the DSR process.

3.3.1 Literature review

A *literature review* is part of every research project so the researcher can understand the current state of knowledge in related research field. The literature review plays a central role in IS research (Levy & Ellis, 2006; Webster & Watson, 2002). The search process to find appropriate literature is systematized to provide an effective review (e.g., Levy & Ellis, 2006; vom Brocke et al., 2009; Webster & Watson, 2002). According to Gough et al. (2017), the steps in systematic reviews consist of the four key activities:

1. Propose a research question.
2. Ascertain and qualify relevant research.
3. Critically evaluate research articles using a systematic and comprehensible process.
4. Run a conclusive analysis and make a final claim.

Vom Brocke et al. (2009) noted that the search process should involve databases but also techniques like keyword searches, backward and forward searches, and an ongoing evaluation of sources. Vom Brocke et al. (2009) emphasized documenting the literature search process and applying a concept matrix for the literature analysis based on Webster and Watson (2002). They also provided support for these steps in their guidelines.

This thesis applies vom Brocke et al.'s (2009) suggestions in papers P.1 and P.2. The systematic literature review in paper P.3 follows Gough et al.'s (2017) stages, and papers P.4 and P.5 also use the findings of literature reviews conducted by other authors.

3.3.2 Laboratory and Field Experiments

In research, an *experiment* is defined as a procedure performed to test a hypothesis as part of the scientific method, that is, “a study in which an intervention is deliberately introduced to observe its effects” (Shadish et al., 2002, p. 12). Experiments typically include an independent variable that is controlled to test its effect on dependent variables. In a *true experiment*, subjects of treatments are randomly assigned to two (or more) groups and then exposed to a different treatment under completely identical conditions (Ross & Morrison, 1996). The literature has offered a definitions of many kinds of experiments, roughly subdivided into *natural experiments* (which contain uncontrolled variables), *controlled experiments* (which contain an experimental group, a control group, and an independent variable), and *field experiments* (natural or controlled experiments conducted in real-world settings). A *randomized experiment* contains units that are assigned to receive a treatment or alternative condition based on a randomized process (e.g., rolling a die), in contrast to *quasi-experiments* in which the randomization does not take place (Shadish et al., 2002). Experimental studies should ensure internal validity of conditions and subject selection and external validity and avoid testing of trivial or inappropriate outcome measures, inappropriate analyses, and insufficient theoretical bas or rationale (Ross & Morrison, 1996).

Another experimental design is a framed field experiment, in which natural subjects (e.g., visitors) conducted natural tasks (e.g., engaging with exhibits) in a natural place (e.g., museum). The underlying research method for papers P.1 and P.2 is that of a crossover study. In crossover experimental designs, participants receive a series of treatments over time (Johnson, 2010; Mills et al., 2009). More specifically, a crossover experiment consists of groups of participants and multiple periods of treatment. Each group receives all treatments but at different times so the effects of different treatments on the same participant can be compared. Each participant serves as his or her own control, eliminating the potential for bias caused by between-subject variability; crossover designs are more efficient than parallel group designs, as they require fewer participants to reach a given level of statistical power; and crossover experiments are attractive for participants, as all participants receive all treatments (instead of some participants’ being part of the control group only). However, special attention has to be paid to minimizing carryover and order effects in crossover designs (Shadish, Cook, & Campbell, 2002).

Laboratory experiments are carried out in a largely controllable environment to keep the variables controllable and to prevent external influences on the research. The laboratory experiment applied in paper P.4 follows a static group design that comprises an experimental group and a control group. This design is used to investigate the differences in the task performance and learning performance of two groups: one that is supported by an AR tool and one that uses traditional tools (i.e., a catalogue). The dependent variable “time for task completion” is applied as a measure of task performance and the dependent variable “number of correctly identified flowers" from the questionnaire fielded after the treatment is an indicator of learning performance.

A *field experiment*, also referred to as a field study and fieldwork, provides in-depth knowledge of a social setting, group, or event (Burgess, 1984). Field experiments conducted in field settings like organizations are high in both internal and external validity (Bhattacherjee, 2012). A field experiment should provide detailed insights into the study's context, as well as the applied principle methodology, the research processes and research methods, a clear presentation of the results and their evidence, the limitations, and identified areas for further research (Burgess, 1984). The field experiment is the underlying research method for paper P.5.

3.3.3 Survey research

The *survey* is a method of gathering information directly from the target audience. Surveys can be applied in quantitative, qualitative and mixed method research. *Survey research* is a purely quantitative method and is described along three characteristics (Pinsonneault & Kraemer, 1993): it produces a quantitative representation of a study population, the information retrieval process follows a structured design and implements predefined questions that the target group are asked, and since just a fragment of the investigated population is involved in the information-collection process, the process is aligned in a way that makes it possible to generalize the results to the entire population (e.g., sample size for statistical analysis). In contrast, a qualitative survey, which is rarely specified in the literature, studies diversity in a population, not its distribution (Jansen, 2010). Mixed method research often uses surveys that contain a quantitative part (e.g., providing categories for answers based on a scale) and asks open-ended questions to receive freely expressed feedback. In such cases, how the data is interpreted and analyzed determines whether the research is qualitative or quantitative (Jansen, 2010). *Qualitative surveys* focus on diversity in a population and their analyses code the survey data in objects, dimensions and categories. In contrast, a *quantitative survey* focuses on frequency distribution and is more unit- and variable-oriented, implementing cluster and homogeneity analysis, correlation, factor analysis, index construction and scaling (Jansen, 2010). The survey research method is applied in papers P.1, P.2, P.4 and P.5 of this thesis.

3.4 Ontological and Epistemological Position

The previous section shows exemplarily the methodological pluralism in IS research, which includes multidisciplinary and multinational IS research (Becker & Niehaves, 2007), illustrating the importance of presenting the philosophical assumptions that frame a particular research project, especially to explain the foundation for the research results and its explanatory power. Burrell and Morgan (2017, p. 1) explained in 1979[6] “that it is convenient to conceptualize social science in terms of four sets of assumptions related to ontology, epistemology, human nature and methodology,” the first three of which have direct implications for the methodology. In the philosophical context, *ontology* relates to the nature of the social world and how it may be investigated. A basic ontological question asks about the investigated reality and whether it is external to the individual or the product of an individual’s cognition (Burrell & Morgan, 2017). In their epistemological framework, Becker and Niehaves (2007 p. 202) provided some distinctions for ontological *realism* (“a world exists independently of human cognition”), ontological *idealism* (“the world is a construct depending on human consciousness”) and *kantianism* (entities exist that are independent as well as dependent on a human mind).

Epistemology relates to nature of knowledge and its justification, how the world is understood and how this knowledge is communicated to others. Epistemological assumptions ask whether knowledge can be acquired or personally experienced, in that way to determine the position for one (Burrell & Morgan, 2017), and Becker and Niehaves (2007) distinguish between *epistemological realism* (the possibility of an independent reality as an objective cognition) and *constructivism* (the subject determines the relationship of cognition and the object of cognition).

The assumption set concerning *human nature* is conceptually separated from epistemological and ontological issues and is associated particularly with the relationship between human beings and their environment (Burrell & Morgan, 2017). Since Burrell and Morgan developed their concept in 1979, when how a society would change with digitization and artificial intelligence (AI) was hardly conceivable, it is understandable that the future research and development of philosophical assumptions are no longer limited to seeing people in the center. Becker and Niehaves (2007) renounced this assumption, but today it is imaginable that, in the future, next-to-human and artificial intelligences will enrich our society, so the philosophical assumption will probably be refined.

How humans investigate and obtain knowledge about the social world relies on the *methodological* aspect of epistemology (Becker & Niehaves, 2007; Burrell & Morgan, 2017). Becker and Niehaves (2007) discerned three directions for how cognition can be obtained: by *induction* (generalization from individual cases to universal laws), by *deduction* (statement derivation from other statements by logical conclusion), and by *hermeneutics* (gaining new knowledge in a circular way during the knowledge construction process and based on existing knowledge).

---

[6] The original source from Burrell and Morgan was published in 1979. In this thesis, a digitalized version published in 2017 is referenced.

IS research paradigms often involve more than one epistemological aspect, such as *positivism* and *interpretivism* (Becker & Niehaves, 2007). Positivists believe that reality is separated from the individual who observes it, while interpretivists believe that reality and the individual who observes it cannot be separated. This *pluralism* is a theoretical doctrine in the theory of science that argues for diversity of theories and methods (Brühl, 2017). Pluralism is based on three assumptions concerning the complexity of social reality: that the human perspective prevents one from fully describing, understanding or explaining the reality (1) by means of a unifying theory, (2) through a single method, or (3) by the social reality itself that is shaped by actors who act in socio-cultural systems based on complex and diverse norms and values that must also be considered. Pluralism is favored for social science, since one-sided theory formation and one-sided use of methods can produce one-sided perspectives on social science phenomena (Brühl, 2017). Accordingly, Mingers (2001, p.3) supported the pluralism approach for IS research to use "the idea of different paradigms to emphasize the desirability of combining together methods that have distinctively different assumptions, but does not wish to remain wedded to the particular paradigm boundaries that exist at the moment."

It is challenging and even contradictory to define an ontological and epistemological position in line with the philosophical assumptions and in the context of an AR that is enriched and extended with digital information, since ontology refers to what exists in the social world and the related assumptions about the form and nature of social reality. It seems obvious at first that AR, as a piece of technology, can be treated as part of the social world, but assigning AR as an additional, digitally enhanced level of our social reality makes clear that AR can create a new, digital, enriched environment that can host the rise of a new, independent social world. In that way, AR research could follow all three distinct ontological positions: realism, idealism and kantianism (Snape & Spencer, 2003). Even though AR is inspired by virtual nature, rather than real nature, its effects on humans and its social impact and perceptions are real.

AR and VR training applications allow people to learn skills like how to operate a laboratory device without ever having seen one. Questions arise about how real our emotions are when we stand in an AR/VR environment that simulates our being on the roof of the Empire State Building. Of course, such an experience has at least some effect on the real world, as therapists have used it to treat people who suffer from acrophobia.

In light of this reasoning, this thesis project follows the pluralist approach proposed by Brühl (2017) and Mingers (2001), rejecting the traditional view that the paradigms are not combinable and that they contradict each other. Mingers (2001) recommended using a pluralistic approach to make research results more meaningful and reliable by combining paradigms. Vaishnavi and Kuechler (2007) argued for a change in research paradigms and philosophical assumptions in the course of a design science project. Thus, the introduction of novel artifacts is changing the state of the world. However, the different states do not remain the same (as, for example, the different realities of the interpretive researcher). At the same time, the belief in a single, stable physical reality is not lost but is substantiated by the diversity of world states. As the phases of DSR proceed iteratively, the researcher's philosophical perspective changes. Hence, a reality will be created by constructive intervention, so the behavior of the system will be recorded and compared with the predictions (theory). In that way, the researcher's position changes to that of a positivist observer (Vaishnavi and Kuechler, 2007).

For this dissertation's various design science studies, the paradigms differ based on Mingers' (2001) and Vaishnavi and Kuechler's (2007) positions. The experiments carried out in papers P.1 and P.2 follow a positivistic position, while the literature analysis carried out in paper P.3 is interpretive in nature, and in papers P.4 and P.5 and the underlying IS design theory, elements of interpretativism and positivism are combined for the process of building and evaluation.

4 Research Results

The following sections summarize the main research results of this thesis: the core design elements of the developed design framework. The chapter has two parts: It begins with a presentation of the AR applications from the research papers that make Part B of this thesis to provide some insights into the AR apps and the application of the framework in a practical way and at an early stage. The second part contain the sections related to the development and implementation of the design framework, beginning with an introduction to explain how the framework is structured. Since the design framework comprises several layers, the layers and elements are presented accordingly. The evaluation of the AR prototypes in various settings is also reported.

All results of this dissertation have been published in academic outlets and are presented in full length in Part B of this thesis. Figure 4.1 illustrates how the research objectives relate to the phases of the design framework's development process, as introduced in Section 3.2, and based on their research objectives, maps them to the publications in Part B.

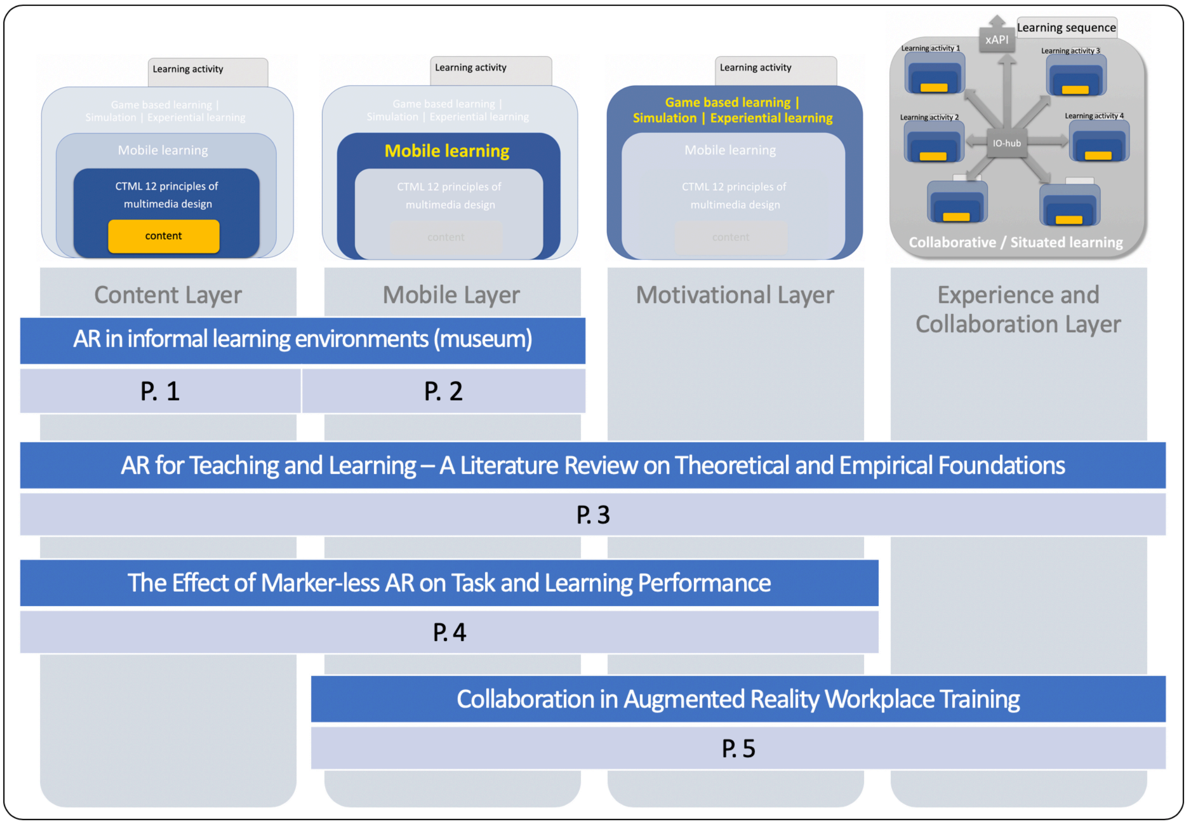


Figure 4.1: Overview research objectives and related publications (P.1 to P.5)

4.1 AR applications

The AR applications designed and developed in the context of this thesis primarily fulfill the purpose of applying the design framework in the app design and evaluate an app's impact in a specific learning setting. This approach also supports the evaluation of the framework, in particular with regard to receiving feedback from the application development about the framework's design elements. The illustration of the apps in this section, before details of the framework are presented, is intended to support the reader by clarifying the learning context and activities addressed with the apps and making the application of the design elements mapped on an app's features in section 4.2 understandable.

Artifact 1: "Museum app"

The research covered by P.1 and P.2 was conducted in the form of a framed field experiment during a mathematics exhibition at the Liechtenstein National Museum in spring 2013, where natural subjects (i.e., visitors) conducted natural tasks (i.e., engaging with exhibits) in a natural place (i.e., the museum). The only artificial component in this setup was that participants were aware they were taking part in an experiment and that their behavior was being recorded and analyzed. The experiment was driven by the hypothesis that visitors learn better from AR-augmented museum exhibits than they do from exhibits that are accompanied only by traditional information displays (e.g., information boards, posters, leaflets). The primary purpose of the study was to determine whether AR is an effective educational technology.

Figure 4.2: The "museum app" in action: trigger image in exhibition (left) and application (right)

Aurasma Studio (Version 2.0) was used to design the augmentations of twelve selected exhibits with videos (including audio) in which the curator explained the mathematical exhibits and animations of the mathematical phenomenon described in the exhibit. The content preparation for the "museum app" followed the cognitive theory of multimedia learning (CTML) principle for multimedia design. In addition, physical objects and virtual content were aligned in space (spatial contiguity principle) and time (temporal contiguity principle). The augmented exhibits were highlighted by trigger images (signaling principle). Visitors used the Aurasma mobile app running on iPads (4th generation) to discover and unlock augmentations by pointing their tablets' cameras at the exhibits and trigger images. All iPads were equipped with headphones to allow the visitor to listen to the audio without disturbing other visitors. Figure 4.2 shows the "museum app" in action.

Artifact 2: "Explore app"

The AR "explore app" in P.4 investigates the application and effectiveness of marker-less AR in supporting the execution of a task in a mundane setting and learning about the underlying domain. The app followed the idea of implementing object recognition and its application in both school and professional education (VET). The applied design principles follow Billinghurst et al. (2015) (i.e., real physical objects/virtual elements displayed and linked with an interaction metaphor) (2015) and the design framework, as presented in section 4.2 of this thesis.

The "explore app" supports the task of learning names related to physical objects used in a particular professional domain: the florist industry, as presented in P.4. More specifically, the app combines machine learning techniques for image recognition with machine translation to identify objects that are in the focus of the mobile phone camera in real time and superimposes information like the object's name onto the object in various languages. Trained in that way, the app can implement any theme from any domain. Figure 4.3 shows screenshots of the application in its explore mode and quiz mode.

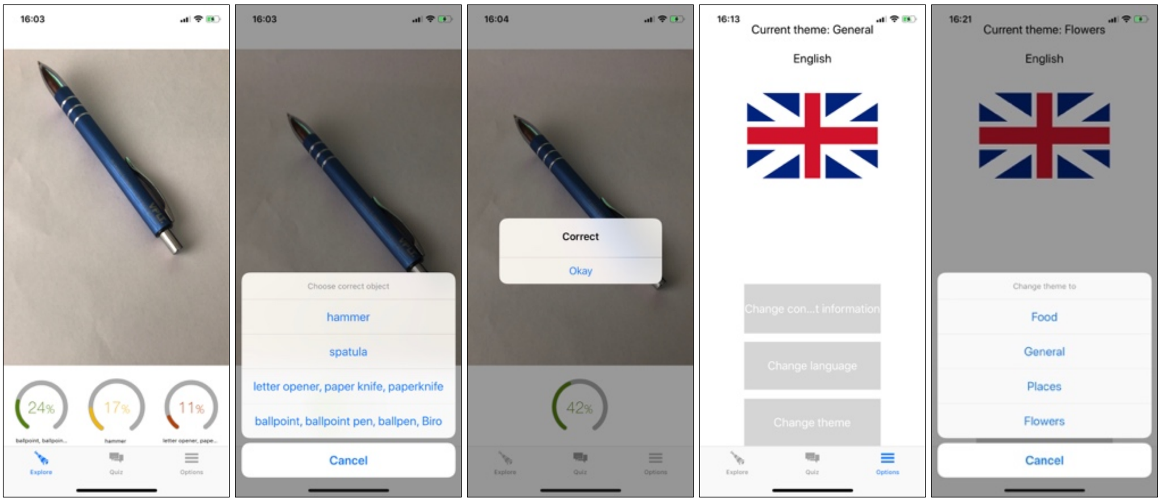
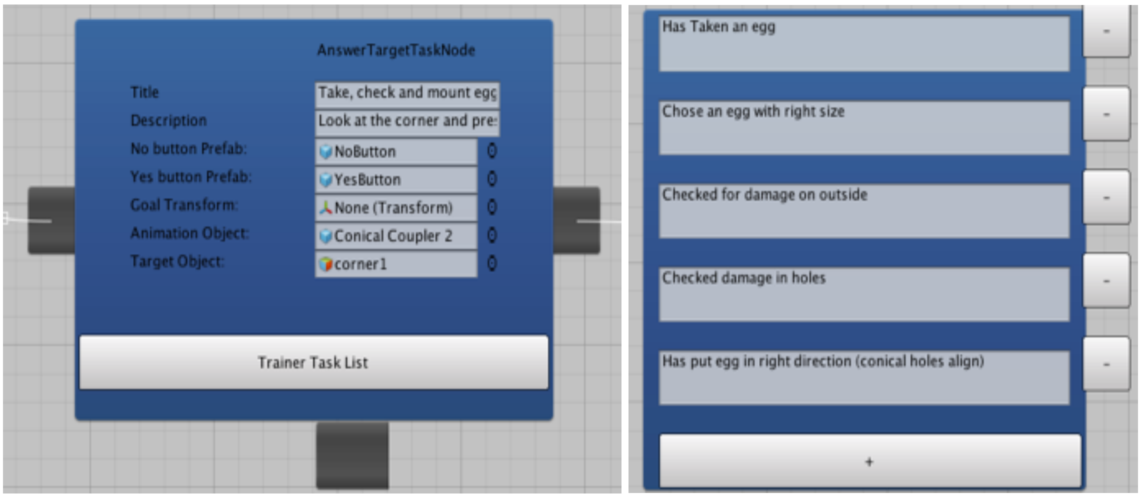


Figure 4.3: The "explore app" in explore mode, quiz mode and selection of language and theme

The multilingual composed training app, which can be used in any environment, implements an exploration mode and a quiz mode. In both modes the user focuses on a particular object using the device's camera (e.g., smartphone, tablet, any head-mounted device). in the exploration mode, the app shows the most likely label for the object for object identification. The quiz mode, which is implemented to support learning at any time and at any place, presents a selection of labels for an identified object, and the user chooses the one that is correct. The app gives feedback for correct and incorrect answers. The app was designed to work with various image sets from a variety of selectable domains.

Artifact 3: "Truss app"

The learning activity enriched with AR and implemented in the "truss app," as documented in P.5, was developed with the elements from the proposed design framework in mind. The setup of the training session followed training instructions from the event technology industry for connecting a truss and covered the identification of the items and tools used. A node editor (Figure 4.4) was implemented for the app development to follow the requirements of a modularized, process-oriented training sequence.



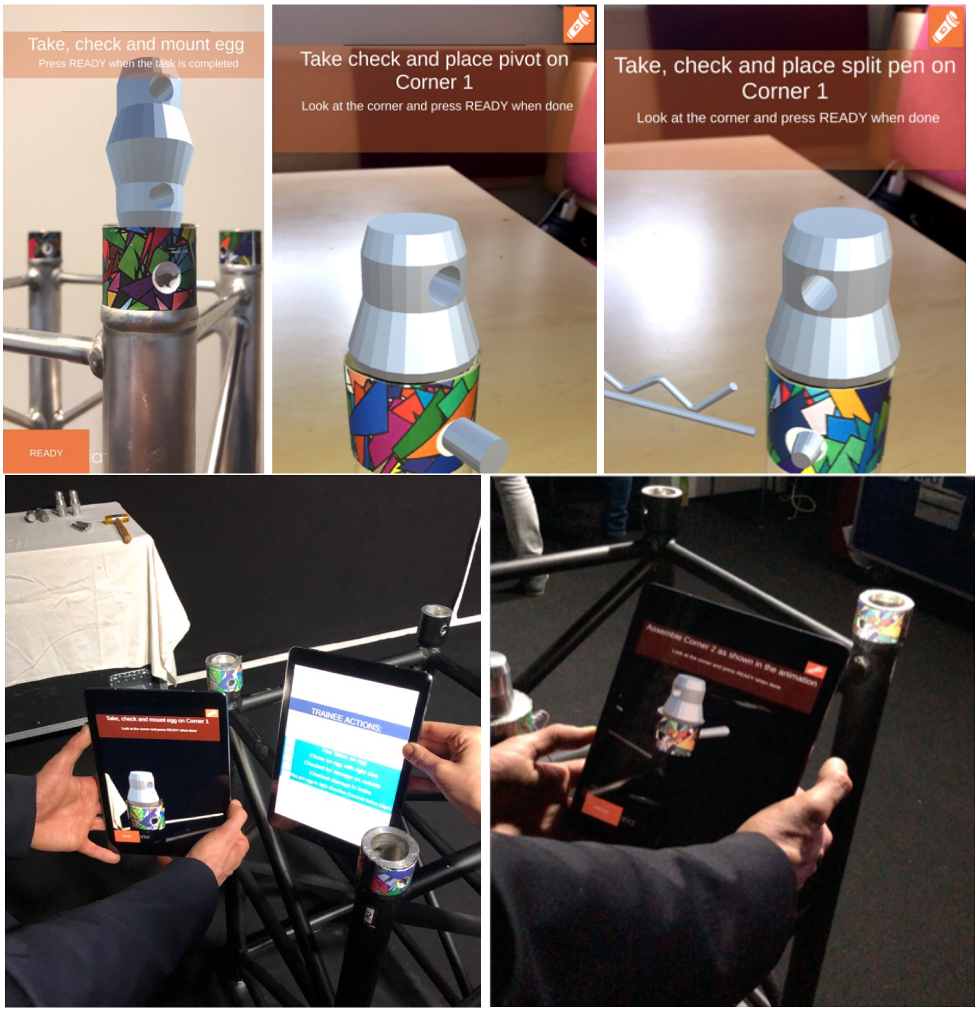
AnswerTargetTaskNode

Title: Take, check and mount eg
Description: Look at the corner and pre
No button Prefab: NoButton
Yes button Prefab: YesButton
Goal Transform: None (Transform)
Animation Object: Conical Coupler 2
Target Object: corner1

Has Taken an egg
Chose an egg with right size
Checked for damage on outside
Checked damage in holes
Has put egg in right direction (conical holes align)

Figure 4.4: Node editor to define training steps for the truss app

The key activity was to prepare a truss element for connection with another truss element. The training app had a trainer mode and a trainee mode. Both modes were connected via a multiplayer server environment to interact in a virtual room. The "truss app" running in trainee mode asked the user to point the camera to one of the four corners of the truss. The first corner was marked with a trigger image to start a 3D animation that showed step-by-step the requested activities; the user had to confirm completion for each step. The procedure for preparing the second corner was identical, but the app illustrated all steps combined in an 3D animation. For both corners, the animation was superimposed on the particular corner of the truss. For the third and fourth corners, the user was requested to prepare everything "on his/her own," so the user had to recall the steps, activities, and required components and tools from memory. At all steps, the app logged the learner's and the trainer's activity. Figure 4.5 shows screenshots from the "truss app" in trainee mode and from the experiment in both modes.


Take, check and mount egg
Press READY when the task is completed

Take check and place pivot on Corner 1
Look at the corner and press READY when done

Take, check and place split pen on Corner 1
Look at the corner and press READY when done

Take, check and mount egg on Corner 1

TRAINEE ACTIONS:

Assemble Corner 2 as shown in the animation

Figure 4.5: Screenshots from the truss app in trainee-mode, pictures below show both modes during the field experiment

The training session was guided by a trainer who used the AR training app in trainer mode. The app in trainer mode fulfill two core tasks: supporting the trainer in, first, leading the trainee's activity and in assisting with (to a low degree) and discussing the activities with the trainee, and second, in observing and evaluating the trainee's performance according to a prepared checklist.

4.2 Design framework

The design framework connects theoretical and empirical foundations with the form and function of concrete AR applications (apps). Therefore, a systematic literature review (P.3 of this thesis) was performed to identify and group the learning theories that have been used in empirical studies on AR for teaching and learning. In a second step, the main system features of the apps in each theory group were identified, grouped to design elements, and arranged in a logical and hierarchical structure (i.e., the layers in the final framework). As a result, design elements were identified that can be traced back to both abstract learning theories and concrete system features.

The first layer addresses the preparation of the learning content used in an AR application (app) and contains elements derived from Mayer's CTML. The second layer introduces mobile aspects of the learning design to break the boundaries of a single location and to enable, for example, the integration of location awareness. Since motivation and engagement are fundamental components of a learning process (Martin, 2012, as cited in Christenson et al., 2012), and AR has the potential to support these aspects of learning, the design framework incorporates elements in the third layer that include motivational aspects of AR learning design. Finally, a primary added value of implementing AR learning is that learning can be an isolated individual task or a collaborative activity. Therefore, the framework includes in the fourth layer the design of a single learning activity and multiple activities composed in a learning sequence. This layer also contains elements derived from situated learning theory (i.e. learning at specific places) and collaborative learning. The layers are described in more detail in the following sub-sections of this thesis.

4.2.1 Design of the Content Layer

Since AR is a supplement to the real world, it superimposes computer-generated additional information onto real-world places and objects, thus allowing users to experience virtual and real layers in the same space and supplying information that users are unlikely to grasp otherwise. The content layer of the proposed design framework contains design elements that help developers and creators generate content that supports learning. Since there is a lack of research on theoretically grounded principles for creating AR content, we draw on work from the field of multimedia learning to identify design elements for the content layer. According to Mayer (2005, p. 31), "A fundamental hypothesis underlying research on multimedia learning is that multimedia instructional messages that are designed in light of how the human mind works are more likely to lead to meaningful learning than those that are not."

From a cognitive perspective, the CTML provides explanations for why AR may improve learning. Mayer posited that people learn better from words and pictures than they do from words alone (1997, 2009). The three basic assumptions of a CTML are:

- Dual-channel assumption: To process visual and auditory information, humans must have access to two separate channels (Paivio, 1990).
- Limited capacity: Each channel is limited in the amount of information that can be processed at one time (Sweller, Ayres, & Kalyuga, 2011).
- Active processing: Learning is an active process in which in-depth and relevant or selected information is organized in coherent mental relationships and linked to other knowledge (Wittrock, 1992).

Studies that have based AR application design on CTML use various subsets of the twelve design principles outlined in this theory to translate CTML's basic ideas into concrete features. To reduce learners' extraneous cognitive load, thereby enhancing their cognitive information-processing processes, researchers focus on four CTML principles (e.g. Parhizkar et al., 2012 and Santos et al., 2014). Designed and applied in the right way, AR incorporates a subset of these design principles:

- Preparing content by using words and pictures; playing spoken words instead of displaying written words (multimedia principle)
- Aligning physical objects and virtual content in space (spatial contiguity principle)
- Aligning physical objects and virtual content in time (temporal contiguity principle)
- Using cues to highlight the organization of essential material (signaling principle)

The multimedia principle states that people learn better from words and pictures than they do from words alone. AR implements this principle by overlaying printed text with virtual pictorial content (e.g., integrating videos into a textbook) or by augmenting physical objects with virtual text (e.g., displaying labels and measures when the learner focuses on a technical object). The spatial and temporal contiguity principles state that learning is enhanced when the space and/or time between disparate but related elements of information is minimized. AR can implement these two contiguity principles by overlaying physical objects with digital content in real time, thereby spatially and temporally aligning related physical and virtual information. The modality principle states that learning can be enhanced by presenting textual information in an auditory format rather than a visual format when it accompanies related visual content. AR can implement the modality principle by playing spoken text instead of displaying printed text when it recognizes a trigger event. Finally, the signaling principle states that people learn better when cues highlight the organization of essential information in a learning environment. AR can implement signaling by directing and guiding people through learning environments using geographic location information and visual triggers.

It is commonly accepted that three types of human memory can be distinguished: sensory memory, short-term memory, and long-term memory. (For an overview of models of human memory, see, e.g., Craik and Lockhart, 1972). External stimuli enter the human memory system through the sensory stores, which are characterized by their pre-attentive, modality-specific, and transient nature. If a subject pays attention to the information that is entering the sensory storage, that information can be transferred to short-term storage (STS), also known as working memory. Compared to sensory storage, STS has a much more restricted capacity but also a slower rate of information loss. Through repeated rehearsal, information can be transferred from STS to long-term storage (LTS), which has no known capacity limits. Compared to the STS, where verbal information is coded phonemically, information in LTS is stored largely semantically and maintained through repetition, organization, and integration with prior knowledge. CTML is largely based on the multi-store model of human memory (Figure 3.1).

By representing information in efficient formats, multimedia technologies can overcome the capacity limitations of our working memories, thereby enabling short- and long-term learning.

Following the design elements proposed in the content layer design for AR development, the application of AR in the "museum app" was evaluated, as presented in P.1 and P.2 of this thesis. Table 4.1 contains examples of the design elements, forms of implementation, and corresponding metrics applied.

Table 4.1: Design elements for the content layer

Layer	Design elements	Implementation	Corresponding metrics
Content layer	Multimedia principle	Include narration in content representation (museum app, P.1, P.2)	Number of correct answers
	Spatial contiguity principle	Place digital information next to object (all apps, P.1, P.2, P.4, P.5)	+ time to single task completion
	Temporal contiguity principle	Align narration and visualization (all apps, P.1, P.2, P.4, P.5)	+ time to mission completion
	Signaling principle	Use AR trigger image to highlight object in the real world (all apps, P.1, P.2, P.4, P.5)	+ Number of attempts to access content
		Use geographic location information	+ log user paths
		Highlight cues to structure content sequentially (truss app, P.5)	+ sequence of tasks fulfilled

As already described in chapter 4.1, the research covered in P.1 and P.2 was conducted as a framed field experiment during a mathematics exhibition at the Liechtenstein National Museum in spring 2013, where natural subjects (i.e., visitors) conducted natural tasks (i.e., engaging with exhibits) in a natural place (i.e., the museum). The only artificial component in this setup was that participants were aware that they were taking part in an experiment and that their behavior was being recorded and analyzed. The experiment was driven by the hypothesis that visitors learn better from augmented museum exhibits than they do from exhibits that are accompanied only by traditional information displays (e.g., information boards, posters, leaflets). The analysis of the results from the treatment group and the control group is based on a pretest and post-test questionnaire consisting of single-choice questions and designed to measure knowledge retention. The field experiment produced empirical evidence that provides strong support for the hypothesis, as visitors performed significantly better on post-test questions related to augmented exhibits than they did on post-test questions related to non-augmented exhibits. They also showed significantly greater gains in scores in terms of comparing post-test and pretest scores. The analysis of the effect size for both tests indicated that AR has a medium effect on learning performance. Concluding from the evaluations in P.1 an P.2, there is first promising quantitative evidence that AR can improve students' learning performance, especially in informal learning environments.

The AR application design for the “explore app,” as presented in P.4, used the app’s core function to implement the spatial and temporal contiguity principle. The app was designed to support a user’s task to identify an object. The AR app design for “the truss app,” demonstrated in P.5, considered all four design elements from the content layer and used a trigger image to highlight the availability of AR content (signaling principle) that was mounted directly on the target object (spatial and temporal contiguity principle) and to present a 3D animation about how to perform the task (multimedia principle).

4.2.2 Design of the Mobile Layer

Today’s mobile AR devices, such as smartphones, tablets, and HMDs, enable ubiquitous computing, so AR applications leverage these devices’ built-in cameras, GPS sensors, and Internet access to overlay real-world environments with dynamic, context-based, and interactive digital content. Since mobile AR is still an emergent technology and field of study, most studies have been qualitative (i.e., they have used methods like observations and interviews) and have focused on eliciting the affordances and constraints of AR for teaching and learning. Only a few quantitative studies have tried to measure the effect of AR on learning performance considering mobile aspects derived from learning theories.

The design elements for the mobile layer follow a constructivist learning theory approach like that proposed in Carlson and Gagnon’s (2016) conceptual model and in the three-phase learning model from Parhizkar et al. (2012). To summarize the requirements for mobile AR, the composition of the mobile layer of the design framework should contain aspects for the design of mobile learning, along with parameters for evaluation of learning success and performance.

Introducing mobility in learning design breaks the boundaries of a single location and enables the integration of location awareness, as described in Tseng et al. (2001) and Wu et al. (2010). Such design elements include maps and features that indicate objects of interest nearby or invitations for students to move in class or to visit a specific place, as Furió et al. (2015) and Kamarainen et al. (2013), respectively, implemented in their studies. Herrington et al. (2009) formulated eleven design principles for mobile learning that constitute the foundation for the mobile layer of the design framework (p. 134):

1. Real-world relevance: Use mobile learning in authentic contexts
2. Mobile contexts: Use mobile learning in contexts where learners are mobile
3. Explore: Provide time for exploration of mobile technologies
4. Blended: Blend mobile and non-mobile technologies
5. Whenever: Use mobile learning spontaneously
6. Wherever: Use mobile learning in non-traditional learning spaces
7. Whoever: Use mobile learning both individually and collaboratively
8. Affordances: Exploit the affordances of mobile technologies
9. Personalize: Employ the learners’ own mobile devices
10. Mediation: Use mobile learning to mediate knowledge construction.
11. Produce: Use mobile learning to produce and consume

Table 4.2: Design elements for the mobile layer

Layer	Design elements	Implementation	Corresponding metrics
Mobile layer	1 Real world relevance	The "museum app" in a museum (P.1, P.2), the "explore app" anywhere (P.4), the "truss app" in workplaces (P.5)	-
	2 Mobile contexts	The "museum app" for visiting an exhibition (P.1, P.2), the "explore app" and the "truss app" to fulfill given tasks (P.4, P.5)	-
	3 Explore	Visitors had 90 minutes to visit the math exhibition ("museum app", P.1, P.2); participants' time to complete a task is measured ("explore app" P.4, "truss app" P.5)	+ time to mission completion
	4 Blended	Interaction with AR and non-augmented exhibits (all apps, P.1, P.2, P.4, P.5)	+ Number of attempts to access content
	5 Whenever	Museum visitors were asked to participate in the experiment (P.1, P.2), independent learning with "explore app" (P.4)	+ log user paths
	6 Wherever	The "museum app" at the museum (P.1, P.2), the "explore app" anywhere (P.4), the "truss app" at a workplace (P5); all non-traditional, informal learning spaces	-
	7 Whomsoever	All apps: individual support (P.1, P.2, P.4, P.5), the "truss app": in collaboration (P.5)	-
	8 Affordances	The "museum app" and the "truss app" on iPads (P.1, P.2, P.5), the "explore app" on iPhone 8+ (P.4), (requirements for computing capacity, machine learning)	-
	9 Personalise	Visitors in museum (P.1, P.2) and participants (P.4, P.5) could download and use the AR app on their own devices.	+ Number of app downloads
	10 Mediation	The "museum app" used videos, animations to explain mathematical phenomenon (P.1, P.2), explore mode and quiz mode (P.4), animation and process steps (P.5)	-
	11 Produse	The "explore app2 in P.4 produces pictures from identified objects; the "truss app" in trainer mode in P.5 produces data to evaluate learning success;	-

The studies in P.1, P.2, P.4, and P.5 demonstrate how these eleven design principles can be embedded into AR applications for teaching and learning. Table 4.2 provides a summary of the design elements, concrete implementations, and corresponding metrics applied.

The AR application design illustrated for the "explore app" in paper P.4 and the "truss app" in paper P.5 implemented design elements like real-world relevance (P4: identifying flowers from the nature, P5: realistic workplace activity), mobile context (both are mobile AR apps), explore (P4: explore various flower species, P5: explore the features of elements for truss connection), whenever and wherever (no time or place restrictions for using both apps), whoever (not restricted to a specific user group), personalize (P4: language selection, P5: session code), and mediation (P4: providing various modes, P5: incorporate and expand various stages).

Whether an implementation in mobile AR application environments is successful can be determined by a simple check for the corresponding design elements. In a few cases, the implementation of a design principle can be measured directly in terms of an added value when measuring learning success. Examples are given in Table 4.2.

4.2.3 Design of the Motivational Layer

Motivation and engagement are fundamental components of a learning process (Martin in Christenson et al., 2012), and AR promises to increase both. Therefore, the guiding questions of the motivational layer concern what increases student motivation and engagement and how AR can support it. Martin (2012, as cited in Christenson et al., 2012) added that a learning setting that supports human nature and inspires a learner raises a learner's curiosity, thus enhancing motivation and engagement.

According to Martin's (2012) model of motivation, self-efficacy, mastery orientation, and valuing support adaptive motivation. In contrast, anxiety, failure avoidance, and uncertain control lead to maladaptive motivation. Since students' motivation influences and is influenced by their engagement in a particular learning situation, aspects of learning like persistence, planning, and task management support adaptive engagement, while disengagement and self-handicapping lead to maladaptive engagement (Martin, 2012, as cited in Christenson et al., 2012). Consequently, to generate a stimulating learning situation, the design framework should incorporate elements that address adaptive motivation and engagement and prevent maladaptive motivation and engagement.

It is a central aim of applying AR in learning environments to turn simple learning into a motivational learning experience. Similarly, the central objective in game-based learning theory follows the equation that having more fun increases the motivation that leads to learning (Prensky, 2001). Prensky's six key structural elements for games and simulations are rules, goals and objectives, outcomes and feedback, conflict/competition/challenge/opposition, interaction, and representation or story. Table 4.3 combines Martin's model of motivation with elements derived from learning theories.

Table 4.3: Design elements for game-based learning and simulation and experiential learning theory

Motivational aspects		**Game based learning and simulation:**	**Experiential learning**
Adaptive motivation	Self-efficacy	Accomplishing missions	Experience
	Mastery orientation	Goals, storytelling	Reflection
	Valuing	Competition, leaderboards, badges, points, rewards, …	Reflection
Maladaptive motivation	Anxiety	Instructional content, collaboration	Observational tasks, substitute experiences
	Failure avoidance	Variation, sandbox mode	Substitute experiences
	Uncertain control	Rules, interactivity	Observational tasks
Adaptive engagement	Persistence	Rules	Reflection
	Planning	Outcomes and feedback	Planning active
	Task management	Storytelling, rules,	Process orientation
Maladaptive motivation	Disengagement	Win states, storytelling, missions, interaction	Reflection
	Self-handicapping	Variation, sandbox mode	Observational tasks, substitute experiences

The design of the motivational layer is complex and covers aspects of constructivist learning theory, particularly game-based learning and simulation, and experiential learning. Therefore, interaction, navigation, and communication in and collaboration between learning activities often follow a predefined process order (Carlson and Gagnon, 2016; Ibáñez et al., 2012; Squire and Jan, 2007). For example, Prensky (2001, p. 6) provided a comprehensive list of features (e.g., motivating aims that affect enjoyment, intense and passionate involvement, structure by rules, goals, interactivity and variability, feedback and gratification, competition and emotion), supported by design elements like storytelling, accomplishing missions, and implementing variation using mini-games between learning steps. Other elements are interaction, navigation, drama and presentation, storytelling, three-dimensional interaction, human controller interface (HCI), programming, pattern analysis, and visual content analysis (Hirumi et al., 2010; Kiili, 2005; Prensky, 2001; van Eck, 2006). Table 4.4 contains the design elements for the motivational layer from literature.

Table 4.4: Design elements for the motivational layer

Layer	Design elements	Implementation	Corresponding metrics
Motivational layer	Rules	Follow the instructions first, then do the task activity or step (in the "truss app")	-
	Goals and objectives	"Explore app" in quiz-mode: give correct answer (P.4); "truss app" for workplace training: prepare a truss for connecting with another (P.5)	+ Fail attempts + Checklist items from trainer
	Outcomes and feedback	"Explore app" in quiz-mode: give the correct answer (P.4); "truss app" for workplace training provides a dashboard that presents learning results (P.5)	+ Number of correctly identified objects + Number of attempts to access content + Time for task completion
	Conflict / competition / challenge / opposition	"Truss app" for workplace training provides dashboard that presents learning results in competition with others (P.5)	+ Number of attempts to access content + Time for task completion
	Interaction	"Explore app": identify flowers to pick from a meadow (P.4); "truss app": illustrates task in animation to complete in the real world (P.5)	+ Number of correct flowers + Time for task completion
	Representation or story	Introductory story / mission (P.4, P.5)	-
	Observation	"Truss app" demonstrates how to prepare a corner of the truss for connection (P.5)	+ Number of attempts to access content
	Reflection	"Truss app" for workplace training provides a dashboard that presents trainer's feedback according to a checklist (P.5)	+ Number of tasks performed correctly
	Experimentation	"Explore app" quiz-mode: give correct answer (P.4); "truss app" for workplace training: prepare a truss for connecting with another (P.5)	+ Fail attempts + Checklist items from trainer
	Connection to the real world	"Explore app": identify flowers from a meadow (P.4); "truss app" demonstrates task in an animation to complete in real world (P.5)	+ Number of attempts to access content

Experiential learning is constructed in a process-oriented way, rather than through single tasks. Following Kolb (2014), such a process starts with a concrete experience, which is deepened through an observation and reflection step, followed by abstract conceptualization to lead to further active experimentation. Studies from a literature search implemented, for example, storytelling as a first instruction, followed by observations in real-world environments (e.g., botanical garden, museum, in nature) either singly or in groups. Findings were shared later with their colleagues in class in a reflection and discussion session to experiment with their own and others' findings.

The laboratory experiment described in P.4 used the "explore app" and followed the advice from Dunleavy and Dede (2009; i.e., decreasing cognitive load by creating a simplified experience structure), Diegmann et al. (2015; i.e., causality between benefits of AR), Chen et al. (2017; i.e., AR classroom design and evaluation research, design and implementation of AR learning resources), and Palmarini (2018; i.e., use of marker-less AR). The experiment also examined the usability of AR, its effectiveness, and its potential for use in teaching and learning. The evaluation covered measures for perceived usefulness, perceived learning, and students' motivation, as well as objective measures in terms of time to task completion and number of mistakes made in a recall and retention post-test. In addition, the SUS was implemented to evaluate the usability of the applied AR system. The item "time for task completion" was used as a measure for task performance and "number of correctly identified flowers" from a questionnaire after the treatment was used as an indicator of learning performance. According to the regression results, participants who use the AR app did not perform significantly better in terms of correctly identifying flowers than participants who used a paper catalogue did. With regard to the time needed to complete the task, participants in the AR group performed significantly worse than participants in the paper catalogue group did. Hence, we found no empirical evidence that the AR app increased participants' objective task performance in terms of task accuracy and speed to completion. Consistent with this finding, participants in the AR group evaluated the perceived usefulness of their tool, the AR app, significantly worse than participants in the non-AR group did. However, in terms of objective learning performance measured by the number of questions answered correctly in the post-test questionnaire, participants in the AR group performed significantly better than those in non-AR group did.

The design principles discussed in this chapter were implemented in the design and development of the "truss app," a collaborative AR training application for the field study covered in P.5. The app followed a predefined storyline and considered the training requirements derived from an existing competence-requirements catalogue, which was developed in collaboration with training experts from industry. The focus of the training app was that learners can first train in a safe environment on their own and at their own pace, supported by a trainer who collaborates in the training situation. Situated in the context of the event technology industry, the learning objective was to be able to prepare the four corners of a truss (technically described as a SD square heavy steel truss element) for connection with another truss element. The application was tested in a realistic setup at an international fair with experts from the industry.

The quantitative evaluation in P.5 focused on a participant's overall performance and addressed the number of correctly and incorrectly fulfilled tasks, the completion time (i.e., for corners 3 and 4 of the truss), and the time taken for the whole training scenario (corners 1-4). There was a significant difference in the scores for time to completion of corner 3 and corner 4. Since participants were faster at corner 4 than they were at corner 3, they trained this skill with the app to improve performance on this task. A weak positive correlation based on a Pearson correlation analysis suggests that participants who needed less time to complete corner 3 were also faster at corner 4. A questionnaire that was used to get participants' feedback based on closed and open questions addressed the participants' impression of the system's usefulness (perceived usefulness), perceived learning, and motivation. The statistical analysis of the quantitative answers to the questionnaire shows a high rating for the system's perceived usefulness (N=57, min=1, max=5, mean=4), an above-average rating for perceived learning (N=6, min=1, max=5, mean=3.7), and a high rating for the participants' motivation/engagement in the training (N=61, min=1, max=5, mean=4.2). As for how participants perceived the AR app's usability, the calculated SUS for each participant shows a mean of 72.8 (minimum 25, maximum 100), which is comparable with the SUS values of good products (Bangor et al., 2008).

4.2.4 Design of the Experience and Collaboration Layer

There are differences between implementing AR learning for a single isolated task and combining two or more learning tasks in a learning sequence and embedding collaboration between users. While single-task applications can easily save the data produced by the app on a user's device, applications that cover more than one task must often also exchange data with external systems (e.g., multiplayer environments), especially for communication with other users.

Communication is central to success in collaborative and situated learning environments (Ibáñez et al., 2012; Santos et al., 2014). For example, Ibáñez et al. (2012) used a developer environment for multiplayer games for their AR app development to support communication between the app and the training system and between users. They implemented their own interface and communication module to migrate objects and users between virtual environments and between these environments and the real world. However, only a few studies have incorporated any standards in their AR app designs for data exchange and communication, especially for internal and/or external communication and capturing learner's behavior, activities, and results. Consequently, a comprehensive and persuasive design framework should incorporate design elements that cover internal and external communication in AR learning applications for both users' and objects' communication based on a well-established standard.

In addition, situated learning incorporates learning at specific places (e.g., in a library, at home, at a botanical garden, in nature, in particular areas in a town). Therefore, the design elements influence and are derived from elements in the environment itself—its atmosphere, impression, environmental and real-world experiences—and integrate discovered objects. Herrington and Oliver (1995) provided the critical characteristics of situated learning for instructional design based on three categories, as shown in Table 4.5. In particular, situated learning entail placing several learning activities into a learning sequence. Therefore, in addition to the design of a single learning activity, the framework also includes in the fourth layer the design of multiple activities composed in a learning sequence. This layer contains elements derived from situated learning theory (i.e., learning at specific places and collaborative learning).

Table 4.5: Design elements from situated learning theory

Designing for …	**Constitutive elements of situated learning in interactive multimedia**
… the role of the interactive multimedia program	Provide authentic context that reflects how the knowledge will be used in real-life;
	Provide authentic activities;
	Provide access to expert performances and the modelling of processes;
	Provide multiple roles and perspectives;
… the role of the student	Support collaborative construction of knowledge;
	Promote reflection to enable abstractions to be formed;
	Promote articulation to enable tacit knowledge to be made explicit;
… the implementation of the program	Provide coaching and scaffolding at critical times;
	Provide for integrated assessment of learning within the tasks.

Future trends for AR in teaching and learning focus on measuring and evaluating learning in personalized student learning experiences, which requires collecting information about the learner’s behavior, activities, and results and exchanging this information with external systems. Both the internal and external process management and the communication between learning activities and with learning management systems can then be supported by the implementation of an input-output hub (IO-hub). Such an approach addresses an important issue that must be considered if AR is to be used successfully in learning (Martgetis et al., 2013; Parhizkar et al., 2012; Santos et al., 2013, Veliz Reyes, 2015) in both online (Lundblad et al., 2012) and network-independent versions (Ternier et al., 2012). However, communication between AR apps and external entities allows data to be collected in the form of activity statements that describe the learner’s behavior during the learning sequence. Thus, it addresses current challenges and future trends in AR teaching and learning. Table 4.6 contains the design elements for the experience and collaboration layer.

Table 4.6: Design elements for the experience and collaboration layer

Layer	Design elements	Implementation	Corresponding metrics
Experience and collaboration layer	Authentic context and activities	To embed learners in the environment; introductory story / mission (P.4, P.5)	-
	Expert performance	“Truss app” for workplace training (P.5) provides a trainer mode; trainer guides the learning process according checklists	-
	Multiple roles and perspectives	“Explore app” provides explore-mode and quiz-mode (P.4); “truss app” for workplace training provides a trainer mode and a trainee mode; trainer guides the learning process supported by different checklists (P.5)	-
	Collaboration	“Truss app” for workplace training provides a trainer mode and a trainee mode; trainer guides the learning process supported by different checklists (P.5)	+ Number of attempts to access content + Time for task completion
	Reflection	Post-session-activity to reflect learning experience; “explore app” in quiz-mode gives feedback for correct/incorrect answers (P.4); “truss app” for workplace training provides dashboard and trainer feedback according to a checklist (P.5)	+ Fail attempts, checklist items from trainer + Results in comparison to other learners
	Articulation	-	-
	Coaching	“Explore app” identifies flowers to collect and in quiz-mode gives feedback for correct/incorrect answers (P.4); “truss app” demonstrates task in animation in the real world and checklist function guides trainer to give instant feedback during the task activity (P.5)	+ Number of correct flowers + Time for task completion
	Integrated assessment	“Explore app” in quiz-mode (P.4) gives feedback for correct/incorrect answers; “truss app” for workplace training (P.5) provides checklists for trainer feedback	+ Fail attempts, checklist items from trainer + Results in comparison to other learners

The AR application design illustrated for the “explore app” in P.4 and the “truss app” in P.5 implemented design elements like authentic context and activities (P.4: explore the real world, P5: realistic workplace activity), expert performance (P.5: collaborative session with trainer), multiple roles and perspectives (P4: explore mode and quiz mode, P5: trainer mode and trainee mode), collaboration (trainer and trainee work in the same session), reflection (P.4: quiz mode, P5: trainer feedback according to a checklist), coaching (P5: trainer interactivity during the whole session), and assessment (P4: quiz mode, P5: checklist and data collection to measure time to completion).

The evaluation of the “truss app” in P.5 uses qualitative and quantitative analyses. The first part of the analysis targeted the measure for time for task completion. The paired t-test to compare the time to completion for corners 3 and 4 showed a significant difference in the scores for time to completion (corner 3: M=39.03, SD=20.95; corner 4: M=23.37, SD=15.44), with t(66)=5.707, p<0.001), so the participants were faster at completing corner 4 than they were at corner 3. A weak positive correlation (r=0.268, n=67, p=0.029) was identified based on a Pearson correlation analysis, which suggested that participants who needed less time to complete corner 3 were also faster at corner 4.

The quantitative evaluation of the questionnaires focused on perceived usefulness and perceived learning and on whether participants learned more, faster, and with higher motivation with the AR support. The results from a Pearson correlation analysis provide some answers. First, participants who valued the app as helpful assessed its usability on a higher level than those who did not see it as helpful, and participants who valued the system’s usability also perceived the app as helpful. Second, participants who found the app helpful in fulfilling the task were also more motivated than those who did not find it helpful. Third, participants who found the app helpful in fulfilling the task also perceived that they had learned more and vice versa. Fourth, participants who agreed that the app was useful also gave higher approval to the effectiveness of AR in training than did those who did not find the app as useful. Finally, participants who were more motivated felt more confident about AR’s effectiveness than did those who were less motivated.

Learning is a complex process and today’s concept of learning goes far beyond a simple acquisition of knowledge and skills (Illeris, 2009, p. 1). Emotional and social dimensions need to be included for future learning, as “all learning implies the integration of two very different processes, namely an external interaction process between the learner and his or her social, cultural, or material environment, and an internal psychological process of elaboration and acquisition” (Illeris, 2009, p. 8). The modular conceptual framework for AR application design is intended to support this complexity of learning.

Connecting two or more learning activities in a learning sequence allows a trainer to treat learning as a process. In most of the studies analyzed for this thesis, the definition and management of such processes was handled in the app itself. Future trends for AR in teaching and learning will focus on measuring and evaluating learning in personalized student learning experiences, which requires collecting information about the learner’s behavior, activities, and results and exchanging this information with external systems. Both the internal and external process management and the communication between learning activities and with learning management systems can then be supported by the implementation of an IO-hub. Such an approach was not implemented in any of the analyzed studies, but it addresses an important issue if AR is to be used successfully in learning (Lundblad et al.,2012; Margetis et al., 2013; Parhizkar et al., 2012, Santos et al., 2013; Veliz Reyes, 2015). However, communication between AR apps and external entities allows data to be collected in the form of activity statements that describe the learner's behavior during the learning sequence, so it addresses current challenges and future trends in AR teaching and learning.

Established standards support the integration of the application into various technology environments so data can be exchanged between systems. According to the Experience Application Programming Interface (xAPI) specification, which is designed to support the collection of formal and informal distributed learning activities (Kevan & Ryan, 2016), a data set contains single activity statements and is stored in a learning record store. Thus, series and various types of experiences, including data from, for example, wearables, mobile applications, and workplace environments, as well as geo data, can be collected to be analyzed later (Silvers, 2017).

From a theory perspective, the xAPI specification is influenced by the socio-cultural framework of Activity Theory (Silvers, 2017) and is in close alignment with constructivist learning theory. The xAPI standard in a learning sequence should be applied early in the design process. Thus, constructivist-aligned strategies are implemented from design through evaluation of the learning activity (Kevan & Ryan, 2016). The data acquisition and data analysis then follow the main aspects of constructivist learning theory.

As a central result of P.3, the composition of the design framework integrates the layers presented in this chapter to support the development process of AR learning applications based on learning theories for content creation and to integrate mobile support and motivational aspects, thus supporting the application of teaching-learning sequences (Méheut & Psillos, 2004) and the implementation of expanding-seeding and contracting-soloing learning sequences, as explored in organizational knowledge management systems (Bingham & Davis, 2012). Figure 4.6 shows a graphic representation of the proposed design framework.

The core of a learning activity in this model is imparting information and knowledge, represented as learning content. In the content layer, the learning content should be prepared according the content layer design presented in section 4.2. The integration of mobile aspects, as presented in the mobile layer design in section 4.3, should be considered right after content creation. At this point, whether a learning application is immovable or mobile should be determined because this determination affects further design elements at the motivation level. The motivational layer considers aspects of game-based learning, simulation-based learning, and experiential learning, especially in terms of interaction and navigation and communication within and collaboration between learning activities. Considering these aspects also lays the basis from which to support Kolb's elements of experiential learning, as presented in section 4.4. Therefore, this layer includes a communication interface with which to collect and exchange information about users' learning experiences with the learning activity.

Finally, one or more learning activities can be assembled into a learning sequence that includes the design elements proposed in section 4.5, which covers learning sequence design. A learning sequence should include elements of coaching, collaboration, and reflection, as well as the application of multiple practices, learning skills, and technology.

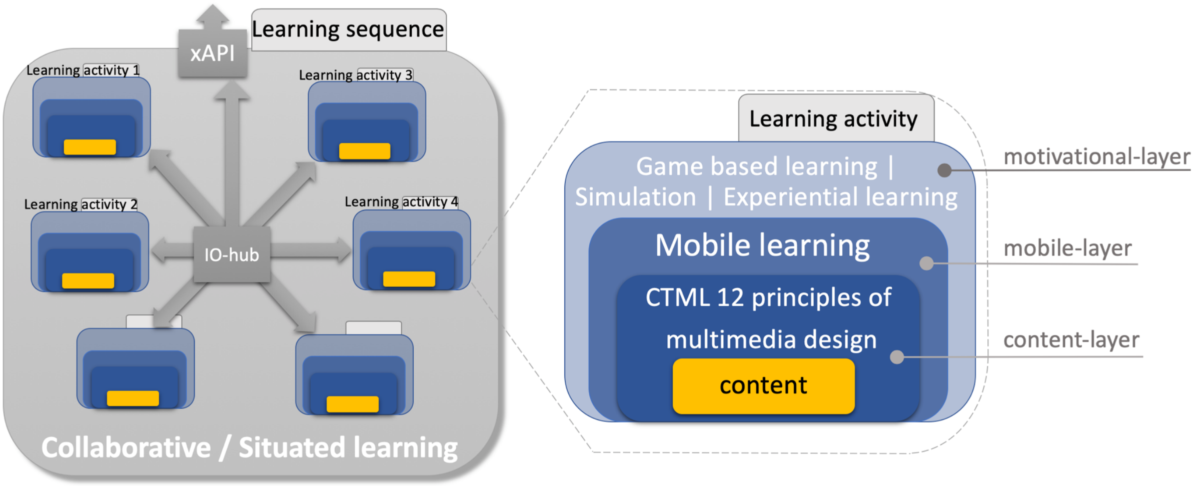


Figure 4.6: Design framework

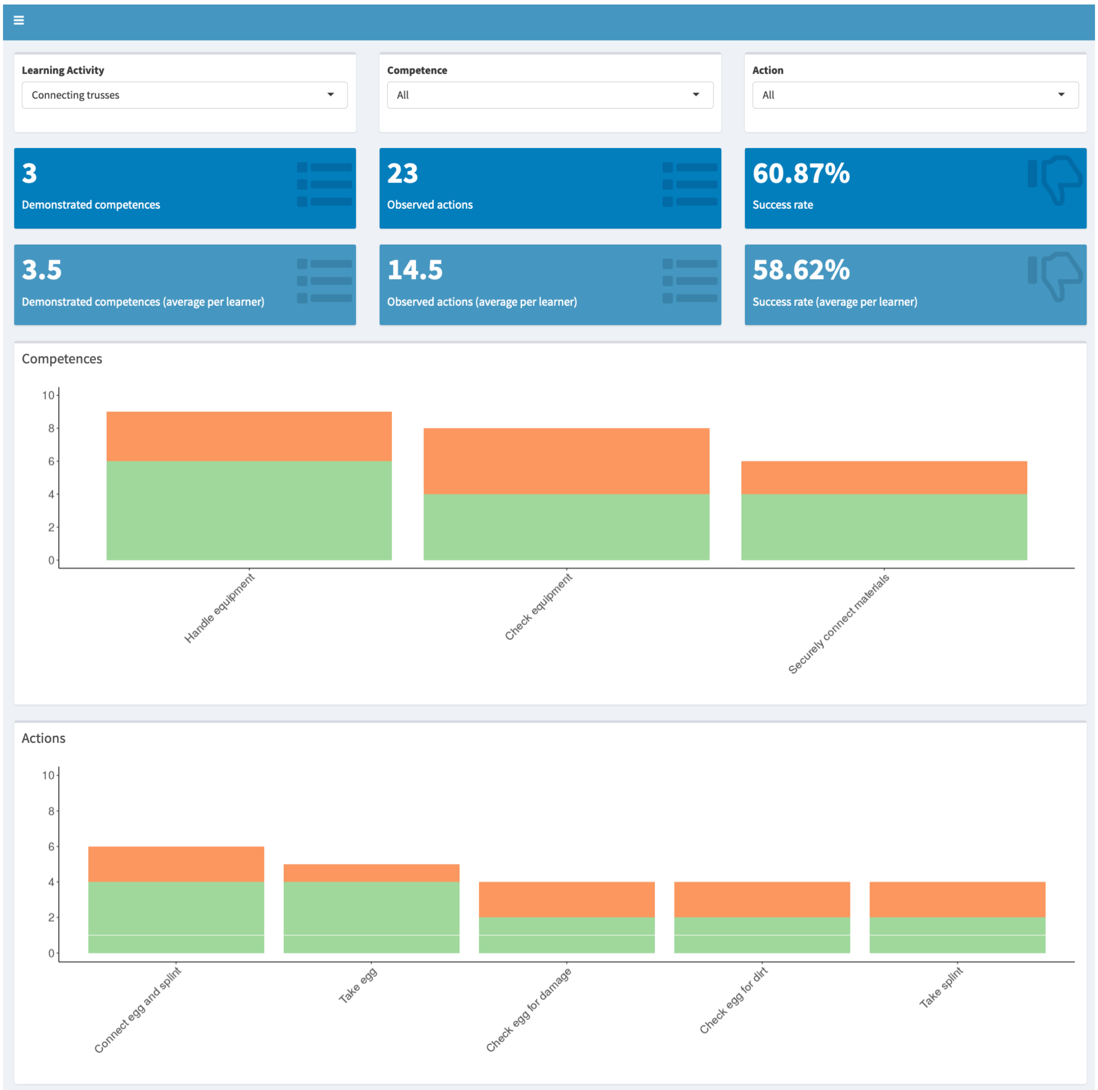
Implementation of the design framework for the development of AR learning applications has several major benefits. First, the design framework can be used as a general guide with which to structure the application-development process systematically according the proposed level design. Second, the framework considers a variety of design elements derived from learning theory, so it follows pedagogical and didactical aspects of learning. Third, the modularized approach to structuring learning is based on single activities that can be combined into learning sequences. Thus, an organized learning database can be used and existing applications can be integrated easily into new learning modules.

As an example, the learning activity enriched with AR for the “truss app” in P.5 was developed using the elements from the proposed design framework. In all training steps, the data from both users’ (trainee and trainer) experience were logged by the app, and the xAPI was used to send the data to a learning record store (LRS), supporting the evaluation of learning success. The focus was on “number of correctly and incorrectly fulfilled tasks”, “completion time for corners 3 and 4”, and “time taken for the whole training session”. The timestamps of the records were used to calculate the respective times. Table 4.7 shows the complete list of the xAPI-statements implemented in the AR app.

Table 4.7: xAPI-statements designed to log users' experience

Actor	**Verb**	**Verb URL**	**Object**	**Description**
Trainer Trainee	initialized	http://adlnet.gov/expapi/verbs/initialized	Application	Start learning scenario
Trainee	launched	http://adlnet.gov/expapi/verbs/launched	Learning step	Record the start of each step
Trainee	viewed	http://id.tincanapi.com/verb/viewed	Content	Record that they've viewed the content
Trainer	evaluated	http://www.tincanapi.co.uk/verbs/evaluated	Trainee attempted statement	The beginning of evaluating
Trainer	rejected	http://activitystrea.ms/schema/1.0/reject	Checklist item	Record when they click through
Trainer	accepted	http://activitystrea.ms/schema/1.0/accept	Checklist item	Record when they click through
Trainee	attempted	http://adlnet.gov/expapi/verbs/attempted	Learning step	Record that they've marked a node complete for testing
Trainee	failed	http://adlnet.gov/expapi/verbs/failed	Learning step	Restart
Trainee	passed	http://adlnet.gov/expapi/verbs/passed	Learning step	Record pass
Trainee	skipped	http://id.tincanapi.com/verb/skipped	Learning step	Not all checklist items ticked but continuing anyway
Trainee	completed	http://activitystrea.ms/schema/1.0/complete	AR app	
Trainee Trainer	exited	http://adlnet.gov/expapi/verbs/exited	AR app	
Trainee Trainer	abandoned	https://w3id.org/xapi/adl/abandoned	AR app	

The training activity ends by displaying a dashboard to provide the user with feedback about the individual training performance. Figure 4.7 shows a prototype of the dashboard. The analysis of the participant's training performance in terms of time to task completion showed a significant difference between the scores for time to completion of corner 3 ($M=39.03$, $SD=20.95$) and that for corner 4 ($M=23.37$, $SD=15.44$), with $t(66)=5.707$, $p<0.001$), so participants were faster at corner 4 than they were at corner 3. Furthermore, based on a Pearson correlation analysis, a weak positive correlation ($r=0.268$, $n=67$, $p=0.029$) suggests that participants who needed less time to complete corner 3 were also faster at corner 4.



Learning Activity: Connecting trusses

Competence: All

Action: All

3 Demonstrated competences

23 Observed actions

60.87% Success rate

3.5 Demonstrated competences (average per learner)

14.5 Observed actions (average per learner)

58.62% Success rate (average per learner)

Figure 4.7: Screenshot from the dashboard the "truss app" presents in trainee-mode at the end of the training

5 Closure

This chapter outlines the contributions of this thesis to the body of knowledge regarding theory-based AR application design for learning applications. The thesis includes quantitative field research and provides theory-based findings for AR introduced in informal learning environments. It also presents some practical implications derived from the framework design and the AR learning apps as IT artifacts that support learning in a particular domain. Next, the conclusions and limitations are discussed, along with an outlook for future AR research.

5.1 Contributions to Research

The overall objective of this research is to contribute knowledge on theory-based AR app design for learning applications and for AR introduced in informal learning environments (see chapter 1.3). This objective was summarized in a central research question, and the research gaps were identified from the literature and addressed in the studies covered by this thesis.

Research Question:	How can mobile AR applications be designed to support learning in informal learning environments?
Research Gap 1:	Few confirmatory studies show the effectiveness of AR in teaching and learning.
Research Gap 2:	Many studies lack theoretical grounding in learning theories and empirical evaluation of prototypes (qualitative and quantitative).
Research Gap 3:	Most studies focus on prototypes, their development, and their application instead of following a well-founded design theory.

The research covered by P.1 and P.2 documents the results from a large-scale field experiment to test the effect of AR on learning performance. The experiment was driven by the hypothesis that visitors learn better from augmented museum exhibits than they do from exhibits that are accompanied by written text and photographs (e.g., boards, posters, leaflets, quizzes, books, screens). The theoretical foundation for this hypothesis is based upon the CTML. P.1 and P.2 contributes to answering the research question in terms of content presentation in AR learning applications in informal learning environments, and it prepared the basis for the content layer design. With its finding that museum visitors learned significantly more from augmented exhibits than they did from non-augmented exhibits, the field experiment addresses Research Gap 1. In addition, since the study was based on CTML, and the AR prototypes were evaluated in qualitative and quantitative ways, both studies contribute to filling Research Gap 2. Finally, the study followed the DSR guidelines as presented in section 3.1 of this thesis, thus addressing Research Gap 3. The AR application, designed in multiple development iterations as prototypes (artifacts), was introduced in a math exhibition to encourage museum visitors to interact with the exhibits, and to determine to what extent AR can enrich the learning from such exhibitions. The utility, quality, and efficacy of the AR app was evaluated during the prototype-development process, by its application in the field experiment, and in the analysis of the research results. P.1 documents the primary results in an article published in *Computers & Education*, and the results related to long-term learning effects from P.2 were presented at the Hawaii International Conference on System Sciences (HICSS2018).

A central contribution to answering the underlying research question is provided by the literature review in P.3, which was undertaken to build a body of knowledge on AR-based instructional design and its effectiveness. Based on a combination of the theoretical foundations of and the empirical evidence for using AR for teaching and learning, the design framework for effective AR-supported teaching and learning was proposed. P.3 laid the basis for the first draft of the framework presented in chapter 4 of this thesis. The framework itself provides a guideline for developing AR learning applications and descriptions of the design elements, covering their reasoning and impact for learning and containing their corresponding metrics for measuring learning success. Although the framework does not claim to be exhaustive, it addresses major aspects of the design of AR learning applications, including design elements for collaboration and communication between learning activities and external systems. Thus, it gives a comprehensive answer to the RQ and provides the basis for closing the identified research gaps (RG1, RG2, RG3). The design framework was used in P.4 and P.5. The results from P.3 were presented at the European Conference of Information Systems (ECIS2018).

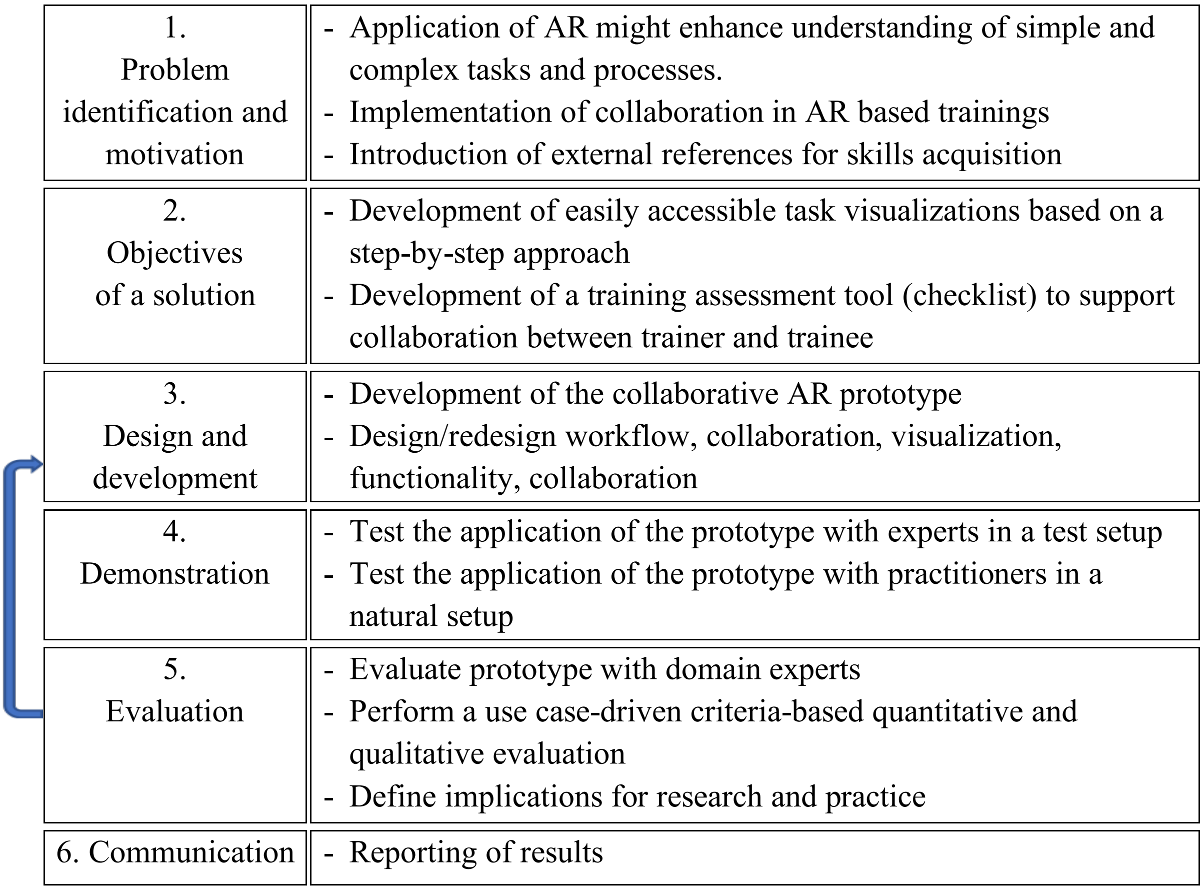
P.4 investigates the application and effectiveness of marker-less AR in supporting the execution of a specific task in an everyday setting and learning about the underlying domain. The laboratory experiment also addresses two additional research questions (RQ1: How can marker-less AR be implemented in a real-world environment? RQ2: How does marker-less AR affect task and learning performance?). These research questions contribute to AR design research by addressing marker-less AR as an alternative to marker-based AR. The AR app, the design of which followed the guidelines provided in the design framework from research P.3, supports the task of learning names related to physical objects used in a particular professional domain. In so doing, it combines machine learning techniques for image recognition and machine translation to identify objects that are in the focus of the device's camera in real time to superimpose information onto the object. The multilingual training app, which can be used in any environment, also implements an exploration mode and a quiz mode. In the exploration mode, the app shows the most likely label for an object, along with a confidence value. The quiz mode, which can support learning anytime and anywhere, presents a choice of labels for an identified object, and the user chooses the one he or she thinks is correct. The app gives feedback for correct and incorrect answers. The app was designed to work with image sets from various domains, which are selectable.

In the final study, P.5, the design framework was implemented for the design and development of a collaborative AR training application, as discussed in section 4.6, and to answer the research question. The qualitative analysis of the participants' performance data, presented in section 4.5 of this thesis, provides some input for Research Gap1. The findings from the evaluation of the training activity, especially findings from qualitative feedback given by domain experts, contribute to addressing Research Gap 2. The feedback was mapped on aspects of valuation, sorted by frequency, and assigned to the corresponding layer of the design framework. Table 5.1 provides the categorized qualitative feedback.

Table 5.1: Categorized qualitative feedback from research in P.5

Layer	Aspects	Quotation
Content	Language	Imparts without language barrier what needs to be done. Language and nationalities independent.
	Descriptive	The visual representation is very helpful because it is easy to understand what needs to be done.
	Self-explaining	Trainer may need to correct only minor issues and it is self-explanatory.
	Under-standing	The visual, very clear presentation / instruction makes it easy for everyone to understand how to proceed.
	Complexity	Learning about complex tasks. Ability to combine a series of steps.
	Clarity	Intuitive operation and clearly defined activities.
	Interactivity	Step by step instruction on the object.
	Multimedia	Higher memorability through multisensory learning.
Mobile	Independence	Learning …, time and place independent.
		Can be used for several people on a construction site.
		For trainees and interns as an exercise in the storage.
Motivation / Engagement	Simple	simple handling; simple to learn;
	Quick	You quickly learn how to handle the traverse.
	Entertaining	It's quick and entertaining. Hands-on approach.
	Costs	No need to travel. Cost efficient.
	Safety	Training with no danger.
	Pace	You can train multiple students on their own tempo.
	Fun	Have fun, enjoy the work.
Situated Learning, Collaboration	Collaborative	You learn together and make no mistakes.
	Complex	Learning about complex tasks. Ability to combine a series of steps.
	Realistic	Realistic, simple and descriptive training.

P.5 integrated DSR, systems development, and action research. Using design science and systems development methods, the AR prototype development used the design framework as a proof-of-concept and enabled the collection of empirical data from the field. Subsequently, in the spirit of action research, the prototype was used to intervene into a real-world training setting, and the usefulness, usability, and learning support of the prototype were evaluated through quantitative (survey) and qualitative (participant feedback and observation) methods. To contribute to the RG3, the study applied the DSR process model as illustrated in Figure 5.1.



1. Problem identification and motivation	- Application of AR might enhance understanding of simple and complex tasks and processes. - Implementation of collaboration in AR based trainings - Introduction of external references for skills acquisition
2. Objectives of a solution	- Development of easily accessible task visualizations based on a step-by-step approach - Development of a training assessment tool (checklist) to support collaboration between trainer and trainee
3. Design and development	- Development of the collaborative AR prototype - Design/redesign workflow, collaboration, visualization, functionality, collaboration
4. Demonstration	- Test the application of the prototype with experts in a test setup - Test the application of the prototype with practitioners in a natural setup
5. Evaluation	- Evaluate prototype with domain experts - Perform a use case-driven criteria-based quantitative and qualitative evaluation - Define implications for research and practice
6. Communication	- Reporting of results

Figure 5.1: Applied DSR process

5.2 Practical Implications

The research covered by this thesis provides concrete guidance to practitioners, especially in terms of the proposed design framework for the development of AR learning applications. The multi-layer design acts as a guideline with its comprehensive presentation of the resulting design elements and their related measures. Since the multi-layer design is modularized, it also enables the integration of subcategories of specialized learning theories and, following the structure of the findings presented, determines new metrics for measuring learning success in a particular learning activity.

The integration of the proposed design principles into the organization and development of learning applications (e.g., for workplace training), contributes to the digital transformation of using and sharing organizational knowledge for both school, higher education, and professional education. As it addresses King’s (2009) understanding of knowledge management, the design framework can be applied in education for the organization, motivation, and controlling of people and to support educational processes and systems in organizations (e.g., companies and universities). Moreover, the framework’s elements for designing a training application enables developers to specify delivery points for submitting information during a learning activity and sequence. In this way, learning assessment becomes an integral part of the learning application’s design.

The apps developed for this thesis project provide some practical implications, mostly in terms of reusing the prototypes, even partially, since the app designs followed the guidelines from the design framework and so feature a modularized approach. The node editor for AR app development, as demonstrated for the “truss app” (P.5), has significant potential since either the node editor itself or the “truss app” can be reused, adapted to, or extended to additional training sequences. The definition of the xAPI-statements designed to log a user’s experience can be reused in any other project and adds to the existing body of knowledge in this field.

A large majority of participants in the field experiment conducted for P.1 and P.2 reported that the “museum app” was a valuable add-on for the exhibition, that the AR experience did not overload them, and that they wished to see more AR in museums in the future. These results indicate that AR is not only an effective tool for learning in museums but also a technology that museum visitors perceive as valuable and desirable. Technically, AR can be seen as an expansion of traditional audio guides. With regard to content, AR can enlarge a visitor’s experience in any number of exhibitions. Thus, the research in P.1 and P.2 can motivate museums and curators of exhibitions from various domains to use AR as a valuable augmentation of a museum visit, especially in terms of information, interaction, communication, and connection with other exhibitions and museums.

The “explore app” (P.4) has several themes that can be extended and applied in various domains. This first prototype supports the task of learning about objects used in a particular professional domain. Aspects of image recognition and analysis were implemented to connected the application to a database of images related to specific topics. The app can be trained using predefined images from objects. The study also investigated how marker-less AR affects task and learning performance in an everyday setting using the simulation of a florist’s job.

With the “truss app” (P.5), a process-oriented set of instructions was successfully implemented during the AR app development that was aligned with the requirements derived from a standardized training curriculum from industry, and features for collaboration in the given setup were introduced. The demonstration of the collaborative “truss app” in a workplace training session provided insights into the AR design framework. Since trainers and trainees benefitted from the motivational, collaborative and realistic training setting and appreciated what was derived from the design. For example, the interactive, intuitive and safe application, independency, efficiency, and process-orientation were identified as added value of AR in workplace training.

5.3 Limitations and outlook

The results from research P.1 and P.2 suggest a number of possible directions for future research. First, while the study was based on CTML and situated learning theory (SLT) as part of the constructivist learning theory, the experimental design was not set up to “prove” that these theories explain the observed effects. Therefore, future studies could, for example, compare the effect of AR experiences that are designed in accordance with the principles of CTML with AR experiences that intentionally violate these principles. Likewise, to investigate the role of SLT further, future studies could compare the effect of AR in terms of social contexts (e.g., formal versus informal learning environments, individual versus group learning) or compare AR applications that offer different levels of immersion (e.g., two- versus three-dimensional experiences, passive versus interactivity experiences).

The “museum app” and the “truss app” were installed on tablet computers. This technology is omnipresent today, but it is not without drawbacks. Users complained that the tablets are heavy to carry around and hold when pointing them at exhibits. As a result, some users’ hands sometimes started shaking, which caused the camera to lose focus and the app to stop the AR experience. In addition, in using the “truss app,” the users found it cumbersome to fulfill a task activity when they needed to work with both hands and handle the tablet at the same time. In such cases, the trainer was held the device. However, future research should investigate the consequences of such usability issues on the effect of AR and test other kinds of AR hardware (e.g., lightweight, head-mounted displays).

External validity of the field experiments was ensured through the realistic setting of each experiment, but many field experiments must also consider threats to internal validity. For example, the participants in the museum experiment were not under the researchers’ control while they performed their activities during their 90-minute museum visit, which could be a confounding factor with influence on the results. Especially in self-directed learning settings like that used in this particular experiment, participants’ interactions with the exhibits differ terms of in time and number of trials, and some participants paid more attention to either augmented or non-augmented exhibits issued in the pretest and post-test. Future research should investigate, whether this effect could be understood as a positive effect derived from the use of technology, rather than a threat.

Another confounding factor emerged because complete equivalence of AR and non-AR materials could not be ensured, even though the AR materials were another representation of the exhibit's information.

The study in P.3 is also not free of limitations. Since the literature review followed a strict process for identifying relevant research articles, some articles could have been missed that would have fit the search profile. To compensate for this possibility, the literature search was also based on recent literature reviews. Another limitation of the study and the proposed design framework is that the study includes only design elements derived from learning theories and does not incorporate additional design theories that are not directly related to learning. Nevertheless, real world annotation, contextual visualization and vision-haptic visualization are the main strengths of AR, which are all supported by learning theories (Santos et al., 2014). Future research could demonstrate the integration of additional design elements. Since the framework follows a modular structure, such aspects of design can easily be integrated as an extension of the framework.

Despite the articles revealed by the literature search, research related to AR implemented in workplace training is still in its infancy, so there is significant potential for future development, especially in terms of distinct learning (i.e., at a student's own pace) and supported employment. Since the thesis arrives at the beginning of the development and evaluation of a number of workplace-based AR training applications, and today's AR technologies are still limited in their functionality and usability, future studies should address the use of smartphones, tablets, and HMDs in workplace environments.

Other limitations for AR as applied in the "explore app" presented in P.4 and the "truss app" in P.5 that are still present today have to do with limited tracking techniques, interaction techniques, user interfaces, and AR displays, especially for head mounted displays (HMD). Although the development of AR hardware became more sophisticated during the past decade, some major technical issues have yet to be overcome, such as low sensitivity trigger to recognition. The experiment in P.4 revealed that the recognition capability is sometimes lacking because of optical influences, which is still a common issue for AR applications. Hence, future technological development should focus on recognition algorithms and the preparation of large and validated datasets that can support the implementation of marker-less AR in education and in other real-life situations. The application of object detection instead of image recognition inside AR applications provides potential for new findings about how full three-dimensional support for such AR apps assists in learning and improves understanding. First results from such research have already confirmed that object detection facilitates the recognition of a series of objects in a single viewpoint.

The limitations for the field study documented in P.5 address usability, technology, and pedagogical and motivational aspects of learning. Observation and the participants' feedback indicated that the task performance was slowed by the AR application, mostly because of usability issues. Especially in the first task of the training, participants had to get used to the system (i.e., pointing the camera at a good angle to the trigger image to start the AR visualization).

The training situations were not recorded to identify any differences in the individual training sessions, but the use of different trainers in the field study could have influenced the results. Although the staff was instructed to follow the structured process for the field study, personal aspects of the trainer and his or her relationship with the trainee (e.g., sympathy, level of details explained) could have influenced participants' motivation and behavior in the training session. Since communication is a key aspect of collaboration, future research could investigate how active collaboration can be explored and how collaborative AR applications can be designed and implemented to support communication processes and their measurement.

Many participants noted that the task was too simple and referred to more complex tasks that would be interesting to use in investigating the application of collaborative AR at workplaces. Thus, the task simplicity could have had an influence on a participant's motivation and aspects of perception like perceived learning and perceived usefulness in the study. The app was prepared in a way that makes it possible to map more complex tasks, which further research will implement.

Part II

Research Papers

Papers of the Dissertation

P.1

Sommerauer, P., & Müller, O. (2014).
Augmented reality in informal learning environments: A field experiment in a mathematics exhibition.
Computers & Education, *79*, 59-68.

P.2

Sommerauer, P., & Müller, O. (2018, January).
Augmented Reality in Informal Learning Environments: Investigating Short-term and Long-term Effects.
In *Hawaii International Conference on System Sciences* (Vol. 51, pp. 1423-1430).
(nominated for best paper award)

P.3

Sommerauer, P., & Müller, O. (2018, November).
AUGMENTED REALITY FOR TEACHING AND LEARNING – A LITERATURE REVIEW ON THEORETICAL AND EMPIRICAL FOUNDATIONS.
Proceedings of the European Conference of Information Systems (ECIS 2018).

P.4

Sommerauer, P., Müller, O., Maxim L. & Østman N. (2019, March),
The Effect of Marker-less Augmented Reality on Task and Learning Performance.
In *14th International Conference on Wirtschaftinformatik (WI2019), February 24-27, 2019, Siegen, Germany,* pp.1696-1710.
(nominated for best paper award)

P.5

Sommerauer, P., Müller, O. & Maxim L. (n.d.),
Collaboration in Augmented Reality Supported Workplace Training.
(Manuscript submitted for publication).

Fact sheet P.1

Title	"Augmented reality in informal learning environments: A field experiment in a mathematics exhibition"
Authors	**Peter Sommerauer** and Oliver Müller
Publication type	Journal paper
Publication outlet	Computers & Education
Year	2014
Status	Published
Full citation	Sommerauer, P., & Müller, O. (2014). Augmented reality in informal learning environments: A field experiment in a mathematics exhibition. *Computers & Education*, *79*, 59-68.

6 Augmented reality in informal learning environments: A field experiment in a mathematics exhibition

Abstract

Recent advances in mobile technologies (esp., smartphones and tablets with built-in cameras, GPS and Internet access) made augmented reality (AR) applications available for the broad public. While many researchers have examined the affordances and constraints of AR for teaching and learning, quantitative evidence for its effectiveness is still scarce. To contribute to filling this research gap, we designed and conducted a pretest-posttest crossover field experiment with 101 participants at a mathematics exhibition to measure the effect of AR on acquiring and retaining mathematical knowledge in an informal learning environment. We hypothesized that visitors acquire more knowledge from augmented exhibits than from exhibits without AR. The theoretical rationale for our hypothesis is that AR allows for the efficient and effective implementation of a subset of the design principles defined in the cognitive theory of multimedia. The empirical results we obtained show that museum visitors performed better on knowledge acquisition and retention tests related to augmented exhibits than to non-augmented exhibits and that they perceived AR as a valuable and desirable add-on for museum exhibitions.

Keywords: Augmented Reality, Informal Learning, Mathematics, Field Experiment, Museum, Cognitive Theory of Multimedia Learning

6.1 Introduction

Augmented reality (AR) refers to technologies that dynamically blend real world environments and context-based digital information. More formally, AR has been defined as a system that fulfills three characteristics (Azuma, 1997): First, it combines the real and virtual world. Second, it allows real-time interaction. Third, it aligns real objects or places and digital information in 3D. In some professional contexts (e.g., military), AR technologies have been around for more than 50 years, but only the recent proliferation and consumerization of mobile technologies (e.g., smartphones, tablets) made affordable AR systems available for the broad public. Today's mobile AR applications leverage the built-in cameras, GPS sensors, and Internet access of mobile devices to overlay real-world environments with dynamic, context-based, and interactive digital content.

It has been asserted that education is one of the most promising application areas for AR (Wu, Lee, Chang, & Liang, 2013). The NMC Horizon Report 2012 identified AR as an emerging technology with high relevance for teaching, learning, and creative inquiry and predicted broad adoption by 2015 (NMC, 2012). Yet, in a recent literature review on AR teaching and learning Dunleavy and Dede (2014) stated that "[d]ue to the nascent and exploratory nature of AR, it is in many ways a solution looking for a problem" (p. 26) and that "relatively few research and development teams are actively exploring how mobile, context-aware AR could be used to enhance K- 20 teaching and learning" (p. 8). In fact, the majority of existing empirical research is of a qualitative nature (e.g., observations, interviews, focus groups) and concentrates on the elicitation of affordances and constraints of AR in education. Up to now, only few quantitative studies (e.g., experiments) exist that try to measure the effect of AR on learning outcomes.

In order to contribute to filing this research gap, we conducted a large-scale field experiment to test the effect of AR on learning performance. Due to its context-awareness and interactivity, many researchers see the biggest potentials in leveraging AR in informal learning environments (Dede, 2009; Greenfield, 2009), that is, voluntary and self-directed learning that takes place outside of the classroom (OECD, n.d.). We concur with this view and, therefore, conducted a field experiment at a mathematics exhibition, a typical example of an informal learning environment (Screven, 1993).

Our experiment was driven by the hypothesis that visitors learn better from augmented museum exhibits than from exhibits that are accompanied by traditional physical information displays only (e.g., boards, posters, leaflets, quizzes, books, screens). The theoretical foundation for this hypothesis is based upon the cognitive theory of multimedia learning (CTML). We argue that AR inherently implements a subset of the design principles formulated in the CTML, namely, the multimedia principle, the spatial contiguity principle, the temporal contiguity principle, the modality principle, and the signaling principle. The empirical results we obtained provide strong evidence for our hypothesis. Museum visitors learned significantly more from augmented exhibits than from non-augmented exhibits, perceived AR as a valuable add-on of the exhibition, and wish to see more AR technologies in museums in the future.

The remainder of this paper is structured as follows. We first present theoretical background on AR in education and related experimental studies that tried to quantify the effect of AR on learning outcomes. We then describe our experimental design in detail before we come to the statistical analysis of the results. In the discussion section we compare and contrast our findings with other studies and point out directions for future research. We conclude with a brief summary and outlook.

6.2 Theoretical Background

The cognitive theory of multimedia learning (CTML) provides potential explanations why AR may improve learning. In broad terms, CTML posits that people learn better from words and pictures than from words alone (Mayer, 1997, 2009). CTML is based on three assumptions. First, humans possess two channels for processing information, an auditory/verbal channel and a visual/pictorial channel (Paivio, 1990). Second, each channel can process only a limited amount of information at one time (Sweller, Ayres, & Kalyuga, 2011). Third, learning is an active process consisting of selecting relevant incoming information, organizing selected information into coherent mental representations, and integrating mental representations with existing knowledge (Wittrock, 1992). Based upon these theoretical assumptions, CTML postulates principles for the design of effective multimedia instructions (Mayer, 2009). We argue that AR, designed and applied in the right way, inherently incorporates a subset of these design principles, namely, the (1) multimedia principle, (2) the spatial contiguity principle, (3) the temporal contiguity principle, (4) the modality principle, and (5) the signaling principle.

The multimedia principle states that people learn better from words and pictures than words alone. AR can implement this principle by overlaying printed texts with virtual pictorial content (e.g., integrating videos into a textbook) or, vice versa, by augmenting physical objects with virtual texts (e.g., displaying labels and measures when focusing on a technical object). The spatial and temporal contiguity principles state that learning is enhanced when the space and/or time between disparate but related elements of information is minimized. AR can implement the contiguity principles by superimposing virtual content onto physical objects in real-time and thereby spatially and temporally aligning related physical and virtual information. The modality principle states that learning can be enhanced by presenting textual information in an auditory format, rather than a visual format, when accompanying related visual content. AR can implement the modality principle by playing spoken text, instead of displaying printed text, when recognizing a trigger event. Finally, the signaling principle states that people learn better when cues highlight the organization of essential information in a learning environment. AR can implement signaling by directing and guiding people through learning environments using geographic location information and visual triggers.

6.3 Related Work

Empirical studies have examined the use AR-based technologies for teaching and learning in natural science, medicine, engineering, languages, history, arts, and other subjects and in various learning environments, for example, kindergartens, schools, universities, laboratories, museums, parks, and zoos (Dunleavy & Dede, 2014; Wu et al., 2013). Given that mobile AR is still an emergent technology and field of study, it is not surprising that the majority of these studies is of a qualitative nature (using methods such as observations or interviews) and concentrates on the elicitation of affordances and constraints of AR for teaching and learning. Up to now, only few quantitative studies exist that try to rigorously measure the effect of AR on learning performance. In the following, we will briefly review extant experimental studies of AR for teaching and learning. As our field experiment focused on teaching general mathematical knowledge, we focused our review on studies that looked at teaching classical K-20 learning contents and excluded studies that looked at specialized professional trainings (e.g., maintenance, repair, medical training). We also excluded studies that lacked the rigorousness of true experimental designs (e.g., control groups, sufficient sample sizes, statistical hypothesis testing). Table 6.1 shows an overview of the studies we were able to identify.

About half of the studies we found examined the effect of AR on learning spatial abilities; a finding that is not surprising as 3D is one of the key affordances of AR. In one of the first large-scale experiments Dünser et al. (2006) investigated the efficacy of AR for training spatial abilities using 215 high school students as participants. Applying a pretest-posttest control group design, the researchers compared an AR-based training application running on a head-mounted display with a CAD application running on a traditional computer with screen, keyboard, and mouse. A between groups comparison could not find clear evidence for the advantageousness of AR as a spatial ability learning tool. Martin-Gutierrez et al. (2010) also studied the effect of AR on learning spatial abilities using a textbook enhanced by a desktop AR system and found more promising results. In a pretest-posttest classroom experiment with 49 university students the AR group showed a significant gain in spatial abilities, whereas the control group using a traditional textbook did not show significant improvements. Finally, in a quasi-experimental study, Fonseca et al. (2014) used a mobile AR application as an educational tool in an architecture and building engineering course with 57 university students. Comparing students' final grades related to practical skills and spatial abilities with the grades of students of the same course in the previous year (control group without AR), they found a significant statistical difference indicating that the application of AR technology in the course helped to improve students' performance.

A second group of studies investigated the effect of AR on the acquisition of theoretical natural science knowledge. For example, Liu et al. (2009) conducted an experiment to measure the effect of a mobile AR application on the acquisition of ecological knowledge during a field trip to a nature park with 72 elementary school students. The researchers used a pretest-posttest design with a control group and found that the AR group significantly outperformed the control group in terms of learning improvement. Echeverria et al. (2012) compared an AR game running on tablet computers with touch screens and additional head-up displays with a multi-mice computer game running on standard PCs. In a pretest-posttest design they measured the acquisition of physics knowledge for both groups. The evaluation showed that both technologies had a significant effect on learning performance, but there was no statistical significant difference between groups. Finally, Ibanez (2014) conducted a classroom experiment with 64 high school students to test whether a mobile AR application for smartphones is more effective in supporting the acquisition of physics knowledge than a similar web-based application. The experiment indicated that students in the AR group perceived higher levels of flow experience during the lecture and also gained significantly more knowledge.

Table 6.1: Overview of Experimental Studies on AR for Teaching and Learning

Study	Domain	Setting	Participants	AR Treatment	Control Group Treatment	Dependent Variables	Positive effect of AR
Dünser, Steinbügl, Kaufmann and Glück (2006)	Engineering	Classroom	215 high school students	AR via head-mounted displays	PC with CAD software	Spatial abilities	No
Liu, Tan and Chu (2009)	Ecology	Field trip	72 elementary school students	AR app on a PDA	Paper-based materials	Knowledge acquisition	Yes
Martín-Gutiérrez et al. (2010)	Engineering	Classroom	49 university students	AR book	Paper-based materials	Spatial abilities	Yes
Echeverría et al. (2012)	Physics	Classroom	45 secondary school students	AR game	Multi-player computer game	Knowledge acquisition	No
Fonseca,Martí, Redondo, Navarro, and Sánchez (2014)	Engineering	Classroom	57 university students	AR smartpho ne app	Paper-based materials	Academic performance (practical skills and spatial abilities)	Yes
Ibáñez, Di Serio, Villarán and Delgado Kloos	Physics	Classroom	64 high school students	AR smartpho ne app	Web-based learning applicatio	Knowledge acquisition; Flow experience	Yes
Chang et al. (2014)	Arts	Museum	135 college students	AR smartphone app	Audio guide; No guide	Painting appreciation; Engagement with paintings; Flow experience	Yes

Finally, we found one experimental study that examined the use of AR in the context of arts education. Chang et al. (2014) designed a AR museum guide and tested its effectiveness against an audio guide and no guide at all. 135 college students participated in the experiment and the

AR group showed significantly greater scores in a painting appreciation test than the two control groups. The researchers also investigated flow levels and amount of time spent focusing on paintings, but did not find clear differences between groups.

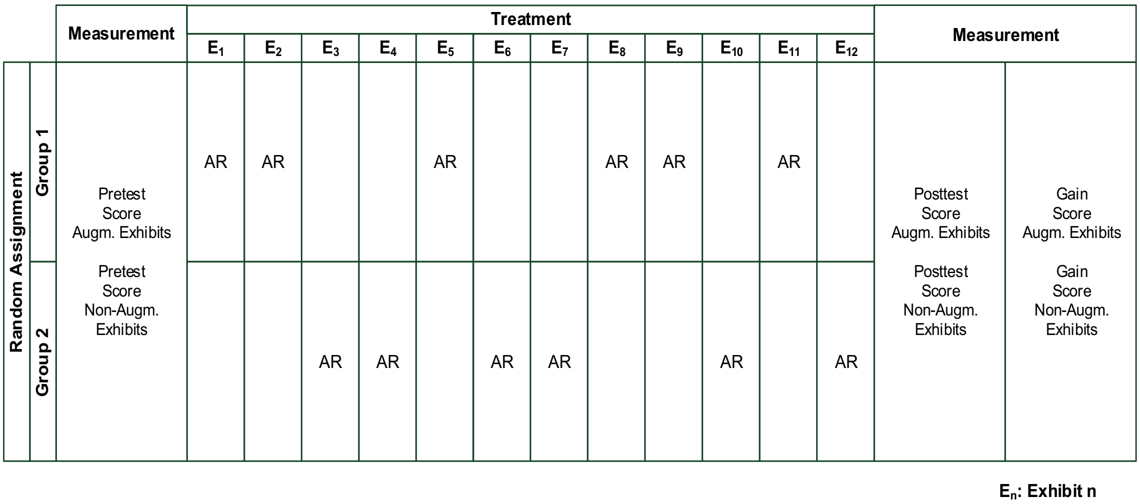
In sum, we can conclude that there is first promising quantitative evidence that AR has the potential to improve students' learning performance. Yet, the experimental results are not completely concordant. Two out of the seven reviewed studies did not find a significant difference between the AR group and the control group. Interestingly, both studies compared AR to other computer-based learning technologies, and not to paper-based learning materials. When looking at teaching and learning mathematics-related contents, which is in the focus of this paper, the picture is even more inconclusive. Three studies found positive evidence for the effectiveness of AR, while two studies did not. Finally, our brief review shows that the majority of studies (five out of seven) investigated the effect of AR on structured, organized, and intentional learning in the classroom (formal learning); only two studies were situated in informal learning environments.

6.4 Materials and Methods

6.4.1 Experimental Design

The objective of our study was to investigate whether AR is an effective educational technology in informal learning environments. Consequently, the hypothesis underlying our study, which was conducted in the form of a field experiment during a mathematics exhibition at the Liechtenstein national museum in spring 2013, was that museum visitors learn better from augmented exhibits than from non-augmented exhibits.

We chose to conduct a framed field experiment (Harrison & List, 2004), in which natural subjects (i.e., visitors) performed natural tasks (i.e., engaging with exhibits) in a natural place (i.e., museum). The only artificial component in the experimental setup was the fact that participants were aware that they are taking part in an experiment and that their behavior is recorded and analyzed. The field experiment was designed as a crossover study (Johnson, 2010; Mills et al., 2009), that is, participants received a series of different treatments over time (i.e., augmented and non-augmented exhibits) so that each participant could serve as its own control, thereby eliminating potential bias caused by between-subject variability. To rule out carryover and order effects, we designed experimental tasks that were logically and temporally independent of each other and let participants roam through the exhibition and complete tasks at their own order and pace.



		Measurement	Treatment												Measurement	
			E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8	E_9	E_{10}	E_{11}	E_{12}		
Random Assignment	Group 1	Pretest Score Augm. Exhibits	AR	AR			AR			AR	AR		AR		Posttest Score Augm. Exhibits	Gain Score Augm. Exhibits
	Group 2	Pretest Score Non-Augm. Exhibits			AR	AR		AR	AR			AR		AR	Posttest Score Non-Augm. Exhibits	Gain Score Non-Augm. Exhibits

E_n: Exhibit n

Figure 6.1: Overview of the randomized crossover field experiment

Figure 6.1 graphically summarizes the design of the experiment. Participants were randomly assigned to one of two groups. Participants in both groups were given 90 minutes to visit the mathematics exhibition individually and at their own pace. Before entering the exhibition, participants received a short hands-on training how to use the mobile AR app to discover and activate hidden virtual contents within the exhibition. In addition, all participants had 15 minutes to take a pretest with 16 questions regarding the mathematical exhibits they will later see. The same test, plus additional questions on demographics and user experience, was administered to all participants as a posttest after visiting the exhibition (participants were not told that the same questionnaire is used for the posttest).

The exhibition consisted of four separate rooms covering eight mathematical topics with a total of 275 exhibits. All objects of the exhibition were accompanied by traditional physical information displays (i.e., boards, posters, leaflets, quizzes, books, screens). For twelve exhibits, we created additional virtual augmentations, six accessible for participants in Group 1 and six accessible for participants in Group 2. All twelve augmented exhibits were tagged with markers.

6.4.2 Participants

We recruited 101 participants to take part in the field experiment. The sample included heterogeneous genders, age groups, and educational levels (Table 6.2). Participants were recruited via mailing lists and local media as well as at the entrance of the museum itself. Participants received free entry into the exhibition as a compensation for taking part in the experiment.

Table 6.2: Participants of the field experiment

Gender		Age				Education (highest degree achieved)		
Male	Female	14-20	21-40	41-60	61-79	Primary school	Secondary school	University
62 (61%)	39 (39%)	35 (34%)	27 (27%)	26 (26%)	13 (13%)	40 (40%)	34 (33%)	27 (27%)

6.4.3 Treatments

We used Aurasma Studio (Version 2.0) to design augmentations for twelve selected exhibits (Table 6.3). Nine objects were augmented with videos (incl. audio) in which the curator explained and demonstrated the mathematical exhibits, three objects were augmented with animations of the mathematical phenomenon described in the exhibit (Figure 6.2). The length of the augmentations varied between 60 and 252 seconds. Visitors used the Aurasma mobile app running on iPads (4th generation) to discover und unlock augmentations by pointing the tablet's camera at exhibits and trigger images. All tablets were equipped with headphones to allow listening to sound without disturbing other visitors. Manipulation of treatments was done by assigning each augmentation to only one of the two experimental groups. Thereby we ensured that for each exhibit half of the participants were able to access the augmented virtual content and the other half had to rely on the physical information displays only. We used the channel concept of Aurasma to implement the grouping of participants and treatments.

As outlined in the Theoretical Background section, we argue that AR enables the efficient and effective implementation of a subset of the design principles stated in the cognitive theory of multimedia learning. In the following, we explain how we incorporated these design principles into the experimental AR materials. We incorporated the *multimedia principle* into the AR materials by explaining the mathematical concepts of an exhibit through rich motion pictures, that is, animations and videos, instead of static graphics and texts. For example, while the physical information display for Exhibit 9 (Linear and exponential growth) illustrated exponential growth through a number series (2, 4, 8, 16, 32, 64, 128, …), the corresponding AR experience showed an animation of the wheat and chessboard problem using time-lapse and zooming features (Figure 6.2). The *spatial contiguity principle* was implemented by superimposing virtual information onto physical exhibits. This removes the need to visually search the environment of an exhibit for explanatory information. For example, in the AR experience of Exhibit 7 (The various nets of a cube's surface, Figure 6.2) the animation unfolded directly on top of the trigger image, while participants in the non-AR group for this exhibit had to spent cognitive resources to constantly switch their visual focus between a model of a cube and surrounding models of its eleven possible nets, and had to integrate these disparate information sources.

Table 6.3: Exhibits and AR experiences

Exhibit	Group	Exhibit and topic	AR Experience
1	1	Interactive model of a cycloid constructed of a three-lane marble track	*Video* in which the curator explains and illustrates that a cycloid has the properties of a tautochrone curve
2	1	Interactive model of a cycloid constructed of a three-lane marble track	*Video* in which the curator explains and illustrates that a cycloid has the properties of a brachistochrone curve
3	2	Interactive model of a hyperboloid constructed of strings	*Video* in which the curator explains why the cooling towers of nuclear power plants are constructed in the form of hyperboloids
4	2	Interactive model of a hyperboloid that is used for plugs in aircrafts; real aircraft plugs	*Video* in which the curator explains why a hyperboloid form guarantees full galvanic isolation of plugs
5	1	Interactive model of a double cone on a diverging monorail	*Video* in which the curator shows that a double cone on a diverging monorail seemingly rolls upwards
6	2	Explanation of the approximation of Pi in an annexed book and on exercise sheets	*Video* in which the curator explains how to approximate Pi by tying a rope around the earth's equator
7	2	Physical models of a cube and the various nets of its surface	*Animation* showing the unfolding of all different nets of a cube's surface (Figure 6.2)
8	1	Interactive installation illustrating the attributes of a plain mirror; additional descriptions on exercise sheets	*Video* in which the curator illustrates the correlation between distance and height of the objects in the mirror
9	1	Illustration of linear and exponential growth through an interactive paper folding experiment and a representation of a exponentially growing number series on the steps of the entrance hall's stairs	*Animation* illustrating the exponential growth through the wheat and chessboard problem (Figure 6.2)
10	2	The Monty Hall problem explained in book in the exhibition's reader's corner	*Animation* explaining the Monty Hall paradox
11	1	Fully functional exemplar of the Arithmometré mechanical calculator from Thomas de Colmar in a glass cabinet	*Video* in which the curator explains and demonstrates the functionalities of the Arithmometré calculator
12	2	Fully functional exemplar of the Heureka mechanical calculator in a glass cabinet	*Video* in which the curator explains and demonstrates the functionalities of the Heureka calculator

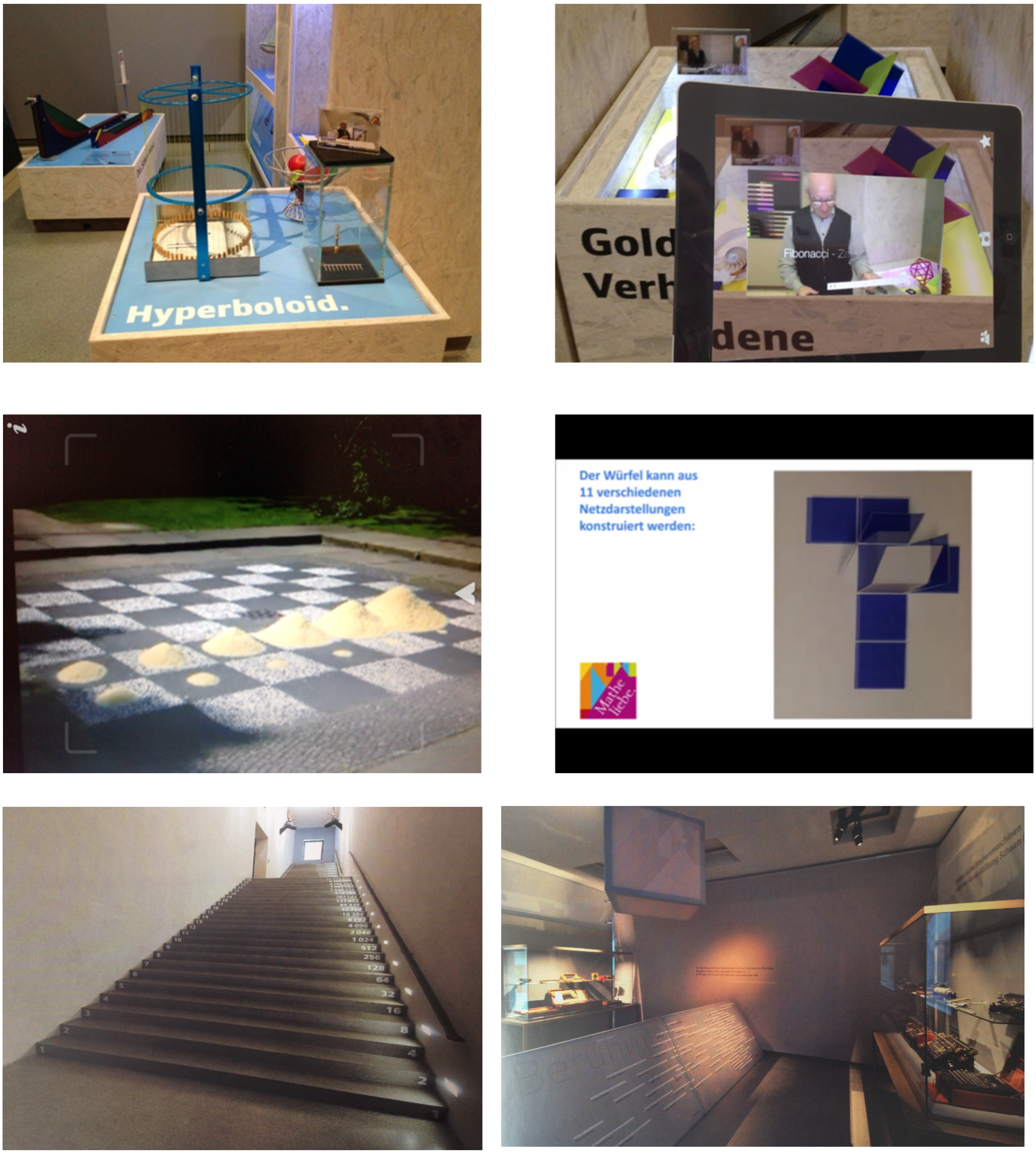


Figure 6.2: Interactive exhibits of hyperboloids (top left); AR experiences (top right: video in which the curator demonstrates an exhibit, middle: two animations illustrating mathematical problems); Illustration of exponential growth on the stairs of the entrance hall (bottom left); Historical calculators in glass cabinets (bottom right)

In a similar vein, we used spoken narration by the curator to provide information about an exhibit at the same time at which the visitor is focusing on the exhibit, thereby implementing the *temporal contiguity principle*. Visitors in the control group, in contrast, had to decide whether to first take a look at the exhibit and then read through the accompanying information, or vice versa, and then needed to integrate both types of information into one congruent mental model. This simultaneous visual and auditory information provisioning is also in line with the *modality principle* of CTML, which states that people learn better from animations with spoken narration than from animations with on-screen text. Finally, we implemented the *signaling principle* within and across AR experiences. Within individual AR experiences, we inserted headings for subsections in order to give structure to videos and animations. Across the whole exhibition, we chose to augment only selected exhibits with AR in order to organize the overall museum visit and highlight the most important objects of each part of the exhibition.

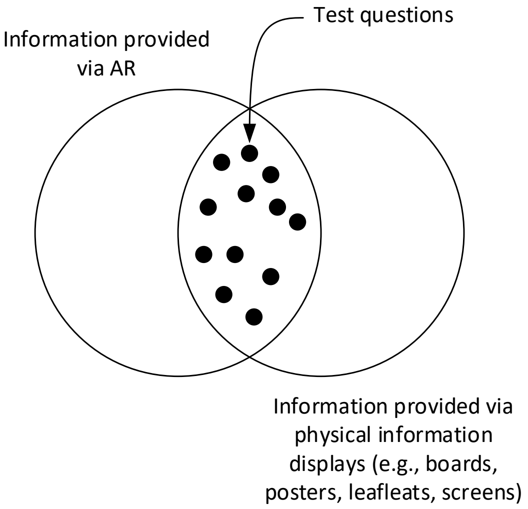


Figure 6.3: Alignment of information provided via AR, information provided via physical information displays and test questions

A key challenge when designing AR materials for experimental treatments is the issue of informational equivalence. According to Larkin and Simon (1987), two representations are informationally equivalent if all the information from one representation can also be inferred from the other representation, and vice versa. On the one hand, informational equivalence is clearly a desirable feature for controlled laboratory experiments on educational technologies as it ensures that differences in effects stem from the mode of representation and not from the content of a representation. On the other hand, we argue that when designing realistic AR experiences, it is difficult to achieve full informational equivalence without undermining the affordances of AR. For example, transcribing all spoken information of a two minutes AR experience would lead to long texts that no museum visitor would read, and, vice versa, transforming all information contained in the physical displays accompanying an exhibit in a science museum into AR would lead to overloaded AR experiences. Therefore, we designed AR materials that overlapped, rather than were equivalent, with physical information displays. Following the guidelines regarding informational equivalence in experimental studies given by Parsons and Cole (2005), our questionnaire was then designed in a way that it was "possible to answer [all] questions correctly with any of the representational forms used as treatments in [the] experimental study" (p. 330). This way, we ensured that both learning experiences were “educationally equivalent”, that is that they support the same learning objectives.[7] Figure 6.3 illustrates this approach graphically.

6.4.4 Measures

Following related experimental studies on the use of AR in education, we focused on knowledge retention as a measure of learning performance using a pretest-posttest measurement approach. This decision was driven by the guidelines outlined in Parsons and Cole (2005), who advocate the use of simple comprehension tests to compare different representations of information, as such tests focus on a representation’s ability to effectively and efficiently convey information. Knowledge application or problem-solving tests, in contrast, are intended to measure a deeper level of domain understanding in which information provided by a representation needs to be integrated with existing knowledge schema (e.g., a person’s general mathematical understanding or mental arithmetic skills).

All pretest and posttest questions were single-choice questions. In the selection and design of the test questions we paid special attention that all question could be answered through both the virtual augmentations of the exhibits and the physical information displays accompanying the exhibits. We created one test question for each of the twelve exhibits being part of the experiment. We selected questions that were adaptations of well-known mathematical problems, for example: *“What is the fastest descent between two points that are not above each other? A) Slope B) S-Curve C) Circular arc D) Cycloid”* or *“How tall a mirror do you need to see yourself? A) Half your height B) Two thirds of your height C) Equal to your height D) Twice your height”*. To establish content validity all questions were reviewed by the curator of the exhibition, who was a retired mathematics high school teacher.

We aggregated the answers to the individual questions to six test scores (Figure 6.1). The pretest score for augmented objects and the pretest score for non-augmented objects captured the level of previous knowledge regarding the mathematical exhibits. The posttest score for augmented objects and the posttest score for non-augmented objects captured the knowledge level after visiting the exhibition. The possible values of pretest and posttest scores ranged between 0 and 6. Knowledge acquisition and retention was measured by computing gain scores as the difference between a participant’s posttest and pretest scores. Analog to the pretest and posttest scores, we computed gain scores for augmented and non-augmented objects separately. Possible values of gain scores ranged between -6 and 6.

In addition to the above test questions we included four control questions into the pretest and posttest questionnaires to check for potential confounding factors. We added three control questions related to exhibits that were not augmented at all, neither for Group 1 nor for Group 2,

---

[7] We gratefully thank one of the anonymous reviewers for providing us with the notion of “educational equivalence”.

and that were not tagged in any way. The answers to these questions were used to check whether visitors were biased towards tagged exhibits, even if they were not able to access the corresponding augmentation (as it was only accessible for participants in the other group). We also added one control question related to an additional exhibit which's augmentation was accessible for both groups. This question was used to check for unintended group differences (e.g., due to inappropriate randomization). The posttest questionnaire also contained a number of simple user experience questions and standard demographics questions.

6.5 Results

6.5.1 Descriptive Statistics

Table 6.4 and Figure 6.4 give an overview of the test scores. All results are in line with expectations. The low scores on the pretest suggest that participants had only little prior knowledge about the topics covered in the exhibition. Even after the visit, participants answered only about half of the test question correctly.

Table 6.4: Descriptive statistics of test scores

	Pretest Scores					Posttest Scores					Gain Scores				
	M	Mdn	SD	Min	Max	M	Mdn	SD	Min	Max	M	Mdn	SD	Min	Max
Augmented Exhibits	1.75	2	1.11	0	5	3.64	4	1.31	0	6	1.89	2	1.50	-2	6
Non-Augmented Exhibits	1.81	2	1.16	0	5	2.59	3	1.28	0	6	0.78	1	1.46	-2	4

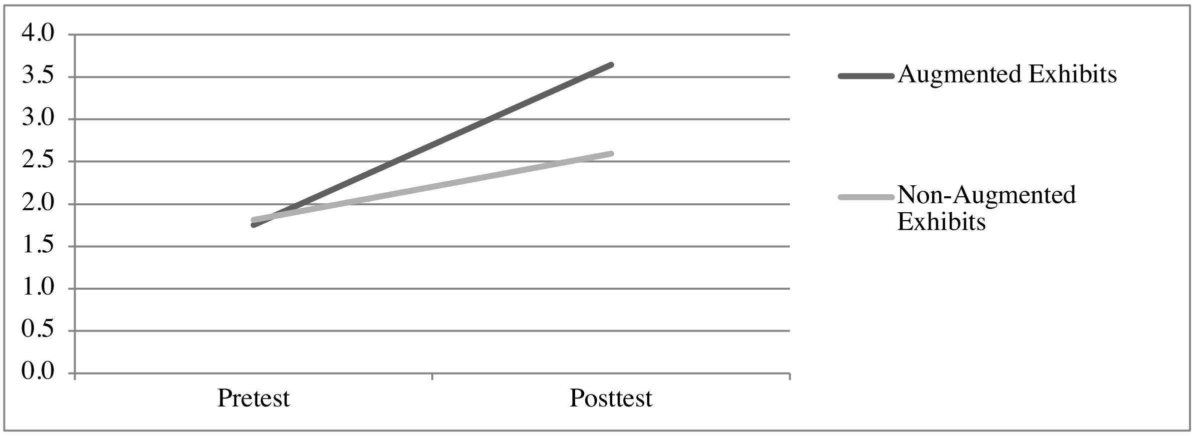


Figure 6.4: Comparison of pretest and posttest scores related to augmented and non-augmented exhibits

6.5.2 Hypothesis Testing

Usually, the statistical analysis of paired pretest-posttest data is done via paired t-tests or a repeated measures analysis of variance (Dimitrov & Rumrill, 2003). Yet, a Kolmogorov-Smirnov test indicated that the required assumption of normality for the dependent variables of the experiment was violated. Hence, we used the equivalent non-parametric Wilcoxon signed-rank test for statistical hypothesis testing. Specifically, we conducted Wilcoxon signed-rank tests on pretest scores, posttest scores, and gain scores for augmented and non-augmented exhibits (Table 6.5) and for the additional control questions.

Table 6.5: Results of Wilcoxon signed-rank tests

		N	Mean Rank	Sum of Ranks
Pretest Score Non-Augmented Exhibits – Pretest Score Augmented Exhibits	Positive Ranks	37[a]	33.58	1242.50
	Negative Ranks	35[b]	39.59	1385.50
	Ties	29[c]	-	-
a. PretestScoreNonAugmentedObjects < PretestScoreAugmentedObjects b. PretestScoreNonAugmentedObjects > PretestScoreAugmentedObjects c. PretestScoreNonAugmentedObjects = PretestScoreAugmentedObjects				
Posttest Score Non-Augmented Exhibits – Posttest Score Augmented Exhibits	Positive Ranks	66[a]	43.11	2845.50
	Negative Ranks	17[b]	37.68	640.50
	Ties	18[b]	-	-
a. PosttestScoreNonAugmentedObjects < PosttestScoreAugmentedObjects b. PosttestScoreNonAugmentedObjects > PosttestScoreAugmentedObjects c. PosttestScoreNonAugmentedObjects = PosttestScoreAugmentedObjects				
Gain Score Non-Augmented Exhibits – Gain Score Augmented Exhibits	Positive Ranks	62[a]	43.63	2705.00
	Negative Ranks	20[b]	34.90	698.00
	Ties	19[c]	-	-
a. GainScoreNonAugmentedObjects < GainScoreAugmentedObjects b. GainScoreNonAugmentedObjects > GainScoreAugmentedObjects c. GainScoreNonAugmentedObjects = GainScoreAugmentedObjects				

To rule out that differences in test scores were caused by different levels of difficulty of question sets related to augmented and non-augmented exhibits, we first performed a Wilcoxon signed-rank test on the pretest scores. The test showed no statistically significant differences in median scores between the two pretest question sets, $z = -0.409$, $p = 0.682$. From 101 participants, 37 participants performed better on questions related to augmented exhibits, 35 participants performed better on questions related to non-augmented exhibits, and 29 participants showed no difference in performance between questions related to augmented and non-augmented exhibits.

Next, we compared medians of posttest scores. Participants performed significantly better on posttest questions related to augmented exhibits (Mdn = 4) than on posttest questions related to non-augmented exhibits (Mdn = 3), $z = -5.069$, $p < 0.005$. From the 101 participants, 66 were better on questions related to augmented exhibits, whereas 17 were better on questions related to non-augmented exhibits; 18 participants showed no difference in performance.

To examine the magnitude of learning improvements, we proceeded with an analysis of gain scores. Participants learned significantly more from augmented exhibits (Mdn = 2) than from non-augmented exhibits (Mdn = 1), $z = -4.679$, $p < 0.005$. 62 participants gained more on questions related to augmented exhibits, 20 participants gained more on questions related to non-augmented exhibits, and 19 participants showed no difference.

We also computed the effect sizes for the Wilcoxon's signed-rank tests using the formula given in Rosenthal (1991, p. 19). The effect size for the difference in posttest scores was $r = 0.36$ and the effect size for the difference in gain scores was $r = 0.33$, which can be considered medium effects (Cohen, 1992).

Finally, we analyzed the control questions to rule out further potential confounding factors. A Wilcoxon signed-rank test[8] showed no significant differences in median gain scores per question for control questions and for questions related to exhibits with inaccessible augmentations. We interpreted this as an indicator that visitors were not biased toward exhibits with inaccessible augmentations, as compared to totally "naked" exhibits, and vice versa. Regarding the control question related to the one exhibit which's augmentation was accessible for both groups, a Mann-Whitney U[9] test found no significant between-subjects difference in median gain scores. This gives indication that there were no differences in the use of AR between the two groups.

6.5.3 Post-hoc Analysis

In addition to the hypothesis tests, we carried out tests to check whether there were any differences in the effect of AR on learning performance between subgroups of our sample. A Mann-Whitney U test[10] with gender as a grouping variable showed neither for augmented nor for non-augmented exhibits a statistically significant difference in median gain scores between males and females. We performed two Kruskal-Wallis tests[11] to inspect whether the effect of AR on learning performance was different across educational and age groups. For the category

---

[8] A visual inspection of the shapes of the distributions of difference scores showed that the scores were approximately symmetrical and, hence, that all required assumptions of the Wilcoxon signed-rank test were met.

[9] A visual inspection of the shape of the distribution of gain scores in each group showed that they were reasonably similar and, hence, that all required assumptions of the Mann-Whitney U test were met.

[10] A visual inspection of the shape of the distribution of gain scores in each group showed that they were reasonably similar and, hence, that all required assumptions of the Mann-Whitney U test were met.

[11] A visual inspection of the shape of the distribution of gain scores in each group showed that they were reasonably similar and, hence, that all required assumptions of the Kruskal-Wallis test were met.

education, the tests showed no significant differences. However, the scores were significantly different between the different age groups for augmented exhibits, $X^2(3) = 10.973$, $p = 0.012$. There were significant differences in gain scores for augmented exhibits between the age group 41-60 (Mdn = 3) and the age group 14-20 (Mdn = 2) ($p = .028$) and the age group 41-60 and the age group 61-79 (Mdn = 1) ($p = .035$), but not between any other combinations. For non-augmented exhibits, no statistically significant differences in gain scores across age groups were found.

6.5.4 Visitor Feedback

Besides measuring learning performance, we also asked participants whether they perceived the augmented exhibits as a positive experience. An overwhelming majority of participants reported that the mobile AR app was a valuable add-on for the exhibition, that the AR experience did not overload them, and that they wish to see more AR in museums in the future (Table 6.6). These results indicate that AR is not only an effective tool for learning in museums, but also a technology that museum visitors perceive as valuable and desirable.

Table 6.6: Visitor feedback on the AR experience

Do you think that AR is a valuable add-on for museum exhibitions?			
Yes, absolutely	Yes, partly	Not really	Not at all
72 (71.3%)	26 (25.7%)	2 (2.0%)	1 (1.0%)
Do you think that the enhancement of exhibitions through AR is “too much”?			
Yes, absolutely	Yes, partly	Not really	Not at all
4 (4.0%)	13 (12.9%)	32 (31.7%)	50 (49.5%)
Do you wish to see more AR in museums in the future?			
Yes, absolutely	Yes, partly	Not really	Not at all
58 (57.4%)	34 (33.7%)	6 (5.9%)	1 (1.0%)

6.6 Discussion

Our field experiment was driven by the hypothesis that museum visitors learn more from augmented exhibits than from non-augmented exhibits. We grounded this hypothesis in the cognitive theory of multimedia learning. The conducted field experiment produced empirical evidence that provides strong support for our hypothesis. Visitors performed significantly better on posttest questions related to augmented exhibits than on posttest questions related to non-augmented exhibits. Also, they showed significantly greater gain scores when comparing posttest and pretest question scores. The analysis of the effect size for both tests indicated that AR has a medium effect on learning performance.

This study contributes to the still emerging body of quantitative empirical evidence on the effect of AR on learning performance, especially learning mathematics-related contents in informal environments. Experimental results on the application of AR in this field are still inconclusive. For example, in contrast to the findings of Dünser et al. (2006) and Echeverria et al. (2012), who could not find a significant advantage of AR learning materials over other materials, we were able to obtain positive evidences for the efficacy of AR. However, it has to be noted that both studies compared AR to other computer-based treatments, and not to physical learning materials. Interestingly, Dünser et al. (2006) and Echeverria et al. (2012) discovered significant gender differences; in both studies male subjects profited from AR as compared to non-AR technologies and outperformed females using AR. We could not replicate these gender differences in our study. Our results are consistent with the results of other studies (Fonseca et al., 2014; Ibáñez et al., 2014; Martín-Gutiérrez et al., 2010), which found that AR can have a significant positive effect on knowledge acquisition performance. In particular, we could replicate and transfer the findings of a recent study of Ibáñez et al. (2014), who found that students using AR performed better on retention tests of physics knowledge than students using a web-base learning tool. The authors explained their results by arguing that AR technologies, as compared to traditional computer technologies, require a lower cogntive effort from users. This rationale is in line with our theoretical argument that AR allows for the efficient and effective implementation of CTML design principles, which, in turn, are partly based on cognitive load theory. When looking at the use of AR in informal learning environemts, our study extends the findings of Liu et al. (2009) and Chang et al. (2014). Both studies found empirical evidences for the efficacy of AR in field settings, but in non-matehmtical contexts. We demonstrated the value of AR for teaching formal contents (mathematics) in informal environments (museums). All extant AR studies in the mathematics context have been conducted in formal classroom situations. Our study, in contrast, investigated natural subjects (i.e., visitors) conducting natural tasks (i.e., engaging with exhibits) in a natural place (i.e., museum). Learning was not an organized and intentional process, but voluntary and self-directed. Taken together, the findings of our study and the above discussed studies suggest that AR has the potential to be an effective learning tool for mathematics-related and other contents in formal and informal learning environments.

The realistic field setting of our experiment added to its external validity. Yet, field experiments come with a number of threats to internal validity. For example, we were not able to control the actions of the experimental subjects during their 90 minutes museum visits. Hence, we cannot rule out that visitors paid more attention to augmented exhibits or to exhibits that were covered in the pretest. Especially the first case is a potential confounding factor that may have influenced our results. Yet, in self-directed learning settings, like the one used in this study, increasing voluntarily time spent on a task could also be understood as a positive side effect of a technology, and not as a threat. A second potential confounding factor stems from the fact that we were not able to ensure full informational equivalence of AR and non-AR materials, as the AR experiences we have designed were not artificial, but used in the museum on a daily basis.

Our study points out a number of possible directions for future research. First, although we have provided theoretical arguments for the proposition that the implementation of the principles of CTML makes AR an effective educational technology, our experimental design was not set up to "prove" that this theory really explains the causes for the observed effects. To do this, future studies should compare the effect of AR experiences that are designed in accordance to the principles of CTML with AR experiences that intentionally violate these principles.

Second, the post-hoc analysis of our experimental results showed that the effect of AR on learning performance differed significantly between age groups. In our experiment, the age group 41-60 profited the most from the use of AR. This is somewhat surprising, as one would usually expect that AR is especially effective with younger people. At the moment, we can only speculate about potential explanations. Our assumption, that builds upon the observations we made and the feedback we got during and after the experiments, is that this age group perceived the AR technology as something new and exciting and, at the same time, was not alienated by it. Yet, further research is needed to replicate, if possible, this result and find theoretically and empirically grounded explanations.

Third, we used a mobile AR app in combination with tablet computers and headphones for the experiment. This technology is omnipresent today, however, not without drawbacks. Some users complained that the tablets are heavy to carry around and hold when pointing at exhibits. As a result, users sometimes started shaking which, in turn, caused the camera to lose the focus and the app to stop the AR experience. Future research should investigate the consequences of such usability issues on the effect of AR and test different AR hardware (e.g., lightweight head-mounted displays).

Finally, our study is not without limitations. In particular, we solely focused on short-term knowledge acquisition and retention. First, it would be interesting to examine whether AR also has a positive effect on long-term knowledge retention. Second, we suggest that future studies should try to replicate our results for higher-order learning tasks, especially knowledge application (problem solving). The studies conducted by Martín-Gutiérrez et al. (2010) regarding the effect of AR on spatial abilities and Fonseca et al. (2014) regarding the effect of AR on general academic performance have already provided first promising results in this respect.

6.7 Conclusion

Recent advances in mobile technologies – mobile cameras, GPS and Internet access – made AR available for everybody owning a smartphone. Consequently, many educators and developers started exploring the potential of AR for teaching and learning in various subjects and contexts. Yet, so far only few studies exist that tried to quantify the effect of AR on learning outcomes. To the best of our knowledge, the here presented study is the first *field* experiment on the effect of AR in learning mathematical contents. The empirical evidence we gathered provides strong support for the proposition that AR has the potential to be an effective tool for learning formal contents (mathematics) in informal learning environments (museums). Museum visitors learned significantly more from augmented exhibits than from non-augmented exhibits, perceived AR as a valuable add-on of the exhibition, and wish to see more AR experiences in museums in the future. Due to this combination of measurable utility and perceived user acceptance we think that AR bears the potential to replace traditional audio guides in museums in the near future; especially when considering the advent of next generation AR devices such as Google Glass.

Fact sheet P.2

Title	"Augmented Reality in Informal Learning Environments: Investigating Short-term and Long-term Effects"
Authors	**Peter Sommerauer** and Oliver Müller
Publication type	Conference
Publication outlet	Hawaii International Conference on System Sciences (nominated for best paper award)
Year	2018
Status	Published
Full citation	Sommerauer, P., & Müller, O. (2018). Augmented Reality in Informal Learning Environments: Investigating Short-term and Long-term Effects. In *Hawaii International Conference on System Sciences* (Vol. 51, pp. 1423-1430).

7 Augmented Reality in Informal Learning Environments: Investigating Short-term and Long-term Effects

Abstract

While many researchers have qualitatively examined the affordances and constraints of AR in educational settings, only few studies exist that tried to quantify the effect of AR on learning performance. To contribute to filling this research gap, we conducted a pretest-posttest-posttest crossover field experiment with 24 participants at a mathematics exhibition to measure the effect of AR on acquiring and retaining mathematical knowledge in an informal learning environment, both short-term (i.e., directly after visiting the exhibition) and long-term (i.e., two months after the museum visit). Our empirical results show that museum visitors performed significantly better on knowledge acquisition and retention tests related to augmented exhibits than to non-augmented exhibits directly after visiting the exhibition (i.e., short-term), but this positive effect of AR vanished in the long run.

7.1 Introduction

Augmented reality (AR) dynamically blends real world environments and context-based digital information (Azuma, 1997). Recent advancements in mobile computing made AR systems affordable for the broad public. Such mobile AR applications use cameras, GPS sensors, and Internet access of mobile devices to overlay real-world environments with dynamic, context-based, and interactive digital content.

It has been argued that education is one of the most promising application areas for AR (Wu et al., 2013). The NMC Horizon Report 2016 identified AR as a technology to bring new opportunities for learning and to offer compelling applications for higher education; AR is especially expected to empower students in STEM (Science, Technology, Engineering, Mathematics) disciplines to engage in deep learning and prepare them for the future workplace (NMC, 2016). Nonetheless, in their literature review on AR for teaching and learning Dunleavy and Dede (2014) stated that “[d]ue to the nascent and exploratory nature of AR, it is in many ways a solution looking for a problem” (p. 26) and that “relatively few research and development teams are actively exploring how mobile, context-aware AR could be used to enhance K-20 teaching and learning” (p. 8). Up to date, most empirical research on AR for teaching and learning is of a qualitative nature and focuses on exploring the affordances and constraints of AR. So far, relatively few quantitative studies exist that tried to measure the effect of AR on learning outcomes (exceptions include, e.g., Fonsecca et al., 2014; Ibáñez et al., 2014; Sommerauer and Müller, 2014).

In order to address the current gap in the body of knowledge on AR for education, we conducted a field experiment in a mathematics exhibition to test the effect of AR on learning performance, both short-term and long-term. Our study was driven by the hypothesis that museum visitors learn better from museum exhibits enriched through AR than from exhibits that are accompanied by traditional physical information displays only (e.g., info boards, posters). The theoretical foundation for this hypothesis is based upon the cognitive theory of multimedia learning (CTML). More specifically, we posit that AR implements a subset of the design principles formulated in CTML, namely, the multimedia principle, the spatial and temporal contiguity principles, the modality principle, and the signaling principle.

The results of our experiment provide evidence for the short-term effectiveness of AR as a tool for supporting learning. Directly after the museum visit, participants were able to retain significantly more knowledge about augmented exhibits than about non-augmented exhibits. However, the advantage of AR over traditional learning materials disappeared when re-testing participants two months after the museum visit, pointing to the need for more research on the design of AR learning materials for supporting sustainable and deep learning experiences.

This paper is structured as follows: We first provide the theoretical background on the cognitive theory of multimedia learning and on the differences of short-term and long-term memory. We then outline our experimental design and present the analysis of our empirical results. After discussing our results, we conclude with a brief summary and directions for future research.

7.2 Theoretical background

7.2.1 Cognitive Theory of Multimedia Learning

The cognitive theory of multimedia learning (CTML) provides potential explanations why AR may improve learning. In broad terms, CTML posits that people learn better from words and pictures than from words alone (Mayer, 1997; Mayer 2009). CTML is based on three assumptions. First, humans possess two channels for processing information, an auditory/verbal channel and a visual/pictorial channel (Paivio, 1990). Second, each channel can process only a limited amount of information at one time (Sweller et al., 2011). Third, learning is an active process consisting of selecting relevant incoming information, organizing selected information into coherent mental representations, and integrating mental representations with existing knowledge (Wittrock, 1992). Based upon these theoretical assumptions, CTML postulates principles for the design of effective multimedia instructions (Mayer, 2009). We argue that AR, designed and applied in the right way, inherently incorporates a subset of these design principles, namely, the (1) multimedia principle, (2) the spatial contiguity principle, (3) the temporal contiguity principle, (4) the modality principle, and (5) the signaling principle.

The multimedia principle states that people learn better from words and pictures than words alone. AR can implement this principle by overlaying printed texts with virtual pictorial content (e.g., integrating videos into a textbook) or, vice versa, by augmenting physical objects with virtual texts (e.g., displaying labels and measures when focusing on a technical object). The spatial and temporal contiguity principles state that learning is enhanced when the space and/or time between disparate but related elements of information is minimized. AR can implement the contiguity principles by superimposing virtual content onto physical objects in real-time and thereby spatially and temporally aligning related physical and virtual information. The modality principle states that learning can be enhanced by presenting textual information in an auditory format, rather than a visual format, when accompanying related visual content. AR can implement the modality principle by playing spoken text, instead of displaying printed text, when recognizing a trigger event. Finally, the signaling principle states that people learn better when cues highlight the organization of essential information in a learning environment. AR can implement signaling by directing and guiding people through learning environments using geographic location information and visual triggers.

7.2.2 Short-term and long-term memory

It is commonly accepted that three different types of human memory can be distinguished, namely, sensory memory, short-term memory, and long-term memory (for an overview of models of human memory see, e.g., Craik and Lockhart, 1972). External stimuli enter the human memory system through the sensory stores, which are characterized by their preattentive, modality-specific, and transient nature. If a subject pays attention to the information entering the sensory storage, it can be transferred to the short-term storage (STS), also known as working memory. Compared to the sensory storage, the STS has a much more restricted capacity but also a slower rate of forgetting. Through repeated rehearsal information can be transferred from the STS to the long-term storage (LTS), which has no known capacity limits. Compared to the STS, in which verbal information is coded phonemically, it is assumed that information in the LTS is stored largely semantically and maintained through repetition, organization, and integration with prior knowledge. CTML is largely based on the multi-store model of human memory (Figure 7.1). By representing information in efficient formats multimedia technologies bear the potential to overcome the capacity limitations of our working memory and thereby enable more effective short- and long-term learning.

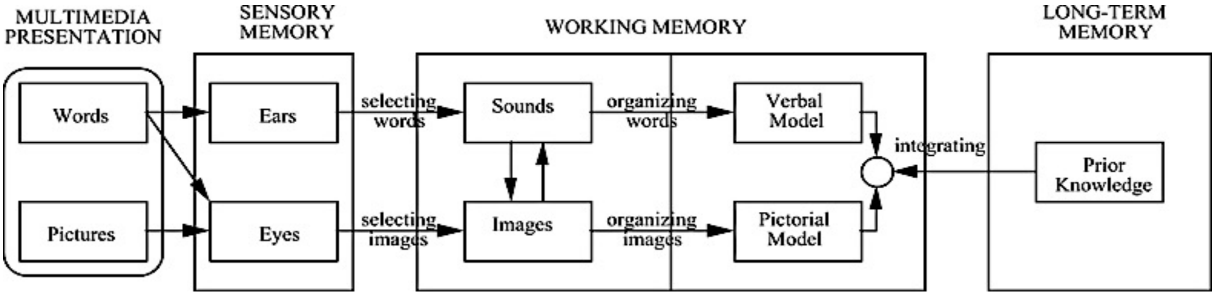


Figure 7.1: Overview of CTML (Mayer, 2009)

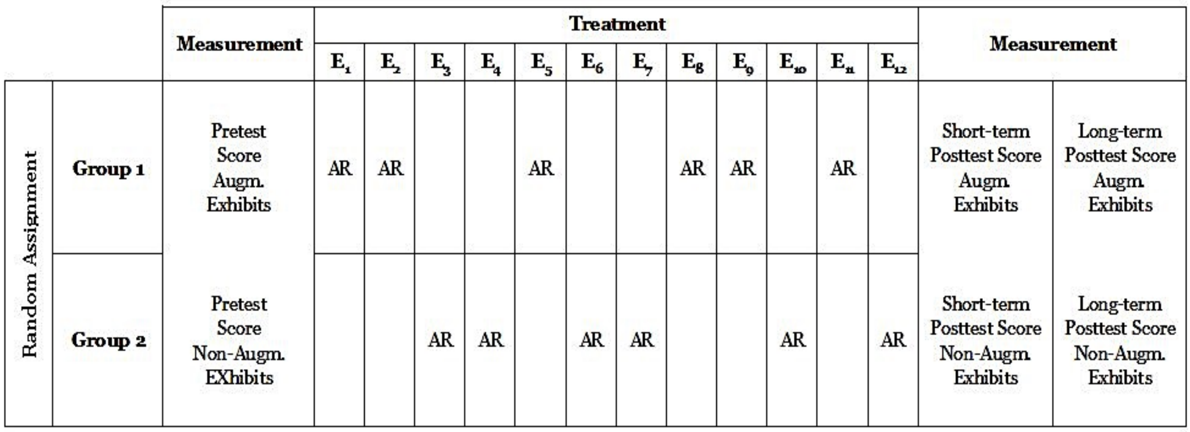
7.3 Experimental design

7.3.1 Setup and participants

The objective of our study was to investigate whether AR is an effective educational technology in informal learning environments, both in the short run and in the long run. Consequently, the hypothesis underlying our study was that museum visitors learn better from augmented exhibits than from non-augmented exhibits.

We chose to conduct a framed field experiment (Harrison and List, 2004), in which natural subjects (i.e., visitors) performed natural tasks (i.e., engaging with exhibits) in a natural place (i.e., museum). The only artificial component in the experimental setup was the fact that participants were aware that they are taking part in an experiment and that their behavior is recorded and analyzed. The field experiment was designed as a crossover study (Johnson, 2010; Mills et al., 2009), that is, participants received a series of different treatments over time (i.e., augmented and non-augmented exhibits) so that each participant could serve as its own control, thereby eliminating potential bias caused by between-subject variability. To rule out carryover and order effects, we designed experimental tasks that were logically and temporally independent of each other and let participants roam through the exhibition and complete tasks at their own order and pace.

Figure 7.2 graphically summarizes the design of the experiment. Participants were randomly assigned to one of two groups and had 15 minutes to take a pretest with 16 questions regarding the mathematical exhibits they were later to see. Participants in both groups were then given 90 minutes to visit the mathematics exhibition individually and at their own pace. Before entering the exhibition, all participants received a short hands-on training how to use the mobile AR app on their own devices in order to discover and activate hidden virtual contents within the exhibition. In addition, ten pre-configured iPads, two smartphones and various headsets were offered to those participants that had problems to get the application running on their own device. The same test, plus additional questions on demographics, was administered to all participants as a posttest directly after visiting the exhibition and in addition two months after the museum visit (participants were not told that the same questionnaire is used for the posttests).



		Measurement	Treatment												Measurement	
			E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8	E_9	E_{10}	E_{11}	E_{12}		
Random Assignment	**Group 1**	Pretest Score Augm. Exhibits	AR	AR			AR			AR	AR		AR		Short-term Posttest Score Augm. Exhibits	Long-term Posttest Score Augm. Exhibits
	Group 2	Pretest Score Non-Augm. EXhibits			AR	AR		AR	AR			AR		AR	Short-term Posttest Score Non-Augm. Exhibits	Long-term Posttest Score Non-Augm. Exhibits

E_n: Exhibit n

Figure 7.2: Overview of Experimental Design

The exhibition consisted of four separate rooms covering eight mathematical topics with a total of 275 exhibits. All objects of the exhibition were accompanied by traditional physical information displays (i.e., boards, posters, leaflets, quizzes, books, screens). For twelve exhibits, we created additional virtual augmentations, six accessible for participants in Group 1 and six accessible for participants in Group 2. All twelve augmented exhibits were tagged with markers.

We recruited a class of 26 pupils (K-20) and their mathematics teacher to take part in the experiment. The group consisted of two female and 24 male students and one male teacher. The students were between 15 and 18 years old and the teacher was 62 years old. While all 26 pupils participated in the first part of the experiment, only 23 could attend the long-term post-test session. Hence and due to the crossover design of the experiment (Jonson, 2010), in which all participants receive to all treatments, both experimental groups contained 23 participants.

7.3.2 Treatments

We used Aurasma Studio (Version 2.0) to design augmentations for twelve selected exhibits. Nine objects were augmented with videos (incl. audio) in which the curator explained and demonstrated the mathematical exhibits, three objects were augmented with animations of the mathematical phenomenon described in the exhibit (Figure 7.2). The length of the augmentations varied between 60 and 252 seconds. Visitors used the Aurasma mobile app running on their own mobile devices like smartphones or on iPads to discover und unlock augmentations by pointing the tablet's camera at exhibits and trigger images. All devices were equipped with headphones to allow listening to sound without disturbing other visitors. We manipulated the treatments by assigning each augmentation to only one of the two experimental groups. Thereby we ensured that for each exhibit half of the participants were able to access the augmented virtual content and the other half had to rely on the physical information displays only. We used the channel concept of Aurasma to implement the grouping of participants and treatments.

As outlined in the background section, we argue that AR enables the efficient and effective implementation of a subset of the design principles stated in the cognitive theory of multimedia learning. For example, we incorporated the multimedia principle into the AR materials by explaining the mathematical concepts of an exhibit through rich motion pictures, that is, animations and videos, instead of static graphics and texts. For instance, while the physical information display for Exhibit 9 (Linear and exponential growth) illustrated exponential growth through a number series (2, 4, 8, 16, 32, 64, 128, …), the corresponding AR experience showed an animation of the wheat and chessboard problem using time-lapse and zooming features. The spatial contiguity principle was implemented by superimposing digital information onto physical exhibits. This removes the need to visually search the environment of an exhibit for explanatory information.

For example, in the AR experience of Exhibit 7 (The various nets of a cube's surface) the animation unfolded directly on top of the trigger image, while participants in the non-AR group for this exhibit had to spent cognitive resources to constantly switch their visual focus between a model of a cube and surrounding models of its eleven possible nets, and had to integrate these disparate information sources. In a similar vein, we used spoken narration by the curator to provide information about an exhibit at the same time at which the visitor is focusing on the exhibit, thereby implementing the temporal contiguity principle. Visitors in the control group, in contrast, had to decide whether to first take a look at the exhibit and then read through the accompanying information, or vice versa, and then needed to integrate both types of information into one congruent mental model. This simultaneous visual and auditory information provisioning is also in line with the modality principle of CTML, which states that people learn better from animations with spoken narration than from animations with on-screen text. Finally, we implemented the signaling principle within and across AR experiences. Within individual AR 14 experiences, we inserted headings for subsections in order to give structure to videos and animations. Across the whole exhibition, we chose to augment only selected exhibits with AR in order to organize the overall museum visit and highlight the most important objects of each part of the exhibition. Table 7.1 provides an overview of the exhibits and AR experiences.

Table 7.1: Exhibits and AR experiences

Exhibit	**Group**	**Exhibit and topic**	**AR Experience**
1	1	Interactive model of a cycloid constructed of a three-lane marble track	*Video* in which the curator explains and illustrates that a cycloid has the properties of a tautochrone curve
2	1	Interactive model of a cycloid constructed of a three-lane marble track	*Video* in which the curator explains and illustrates that a cycloid has the properties of a brachistochrone curve
3	2	Interactive model of a hyper-boloid constructed of strings	*Video* in which the curator explains why the cooling towers of nuclear power plants are constructed in the form of hyperboloids
4	2	Interactive model of a hyper-boloid that is used for plugs in aircrafts; real aircraft plugs	*Video* in which the curator explains why a hyperboloid form guarantees full galvanic isolation of plugs
5	1	Interactive model of a double cone on a diverging monorail	*Video* in which the curator shows that a double cone on a diverging monorail seemingly rolls upwards
6	2	Explanation of the approximation of Pi in an annexed book and on exercise sheets	*Video* in which the curator explains how to approximate Pi by tying a rope around the earth's equator
7	2	Physical models of a cube and the various nets of its surface	*Animation* showing the unfolding of all different nets of a cube's surface
8	1	Interactive installation illustrating the attributes of a plain mirror; additional descriptions on exercise sheets	*Video* in which the curator illustrates the correlation between distance and height of the objects in the mirror
9	1	Illustration of linear/exponential growth through an interactive paper folding experiment and a representation of a	*Animation* illustrating the exponential growth through the wheat and chessboard problem
10	2	The Monty Hall problem explained in book in the exhibition's reader's corner	*Animation* explaining the Monty Hall paradox
11	1	Fully functional exemplar of the Arith-mometré mechanical calculator from Thomas de Colmar in a glass cabinet	*Video* in which the curator explains and demonstrates the functionalities of the Arithmometré calculator
12	2	Fully functional exemplar of the Heureka mechanical calculator in a glass cabinet	*Video* in which the curator explains and demonstrates the functionalities of the Heureka calculator

A key challenge when designing AR materials for experimental treatments is the issue of informational equivalence. According to Larkin and Simon (1987), two representations are informationally equivalent if all the information from one representation can also be inferred from the other representation, and vice versa. On the one hand, informational equivalence is clearly a desirable feature for controlled laboratory experiments on educational technologies as it ensures that differences in effects stem from the mode of representation and not from the content of a representation. On the other hand, we argue that when designing realistic AR experiences it is difficult to achieve full informational equivalence without undermining the affordances of AR. For example, transcribing all spoken information of a two minutes AR experience would lead to long texts that no museum visitor would read, and, vice versa, transforming all information contained in the physical displays accompanying an exhibit in a science museum into AR would lead to overloaded AR experiences. Therefore, we designed AR materials that overlapped, rather than were equivalent, with physical information displays. Following the guidelines regarding informational equivalence in experimental studies given by Parsons and Cole (2005), our questionnaire was then designed in a way that it was "possible to answer [all] questions correctly with any of the representational forms used as treatments in [the] experimental study" (p. 330). This way, we ensured that both learning experiences were "educationally equivalent", that is that they support the same learning objectives. Figure 7.3 illustrates this approach graphically.

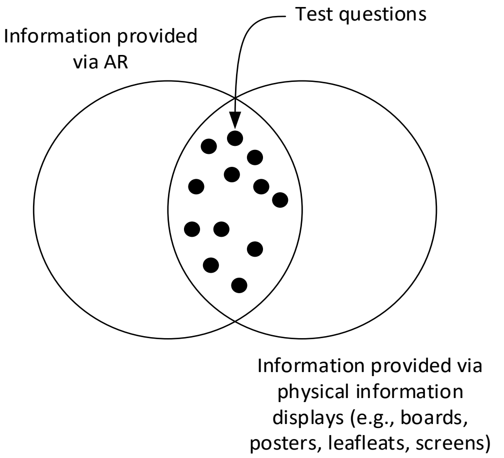


Figure 7.3: Alignment of information provided via AR, information provided via physical information displays and test questions

7.3.3 Measures

Following related experimental studies on the use of AR in education, we focused on knowledge retention as a measure of learning performance using a pretest-posttest measurement approach. All test questions were single-choice questions. In the selection and design of the test questions we paid special attention that all question could be answered through both the virtual augmentations of the exhibits and the physical information displays accompanying the exhibits. We created one test question for each of the twelve exhibits being part of the experiment. We selected questions that were adaptations of well-known mathematical problems, for example: "What is the fastest descent between two points that are not above each other? A) Slope B) S-Curve C) Circular arc D) Cycloid" or "How tall a mirror do you need to see yourself? A) Half your height B) Two thirds of your height C) Equal to your height D) Twice your height". To establish content validity the curator of the exhibition, who was a retired mathematics high school teacher, reviewed all questions.

We aggregated the answers to the individual questions to six test scores (Figure 7.2). The pretest score for augmented objects and the pretest score for non-augmented objects captured the level of previous knowledge regarding the mathematical exhibits. The short-term and long-term posttest scores for augmented objects and non-augmented objects captured the knowledge level after visiting the exhibition. The possible values of all scores ranged between 0 and 6.

7.4 Results

7.4.1 Descriptive statistics

Table 7.2 gives an overview of the test scores. The low scores on the pretest (on average 2 out of 6) suggest that participants had only little prior knowledge about the topics covered in the exhibition. Even after the visit, participants answered only about half of the test question correctly.

Table 7.2: Descriptive Statistics

Test	Measure	Non-Augmented Exhibits	Augmented Exhibits
Participants	N	23	23
Pretest Scores	Mean	2.05	2.00
	Median	2	2
	SD	1.27	1.03
	Min	1	0
	Max	5	4
Short-term Posttest Scores	Mean	2.59	3.42
	Median	3	4
	SD	1.34	0.65
	Min	1	0
	Max	5	5
Long-term Posttest Scores	Mean	3.04	2.88
	Median	3	4
	SD	1.11	1.68
	Min	1	0
	Max	5	5

7.4.2 Hypothesis tests

Usually, the statistical analysis of paired pretest-posttest data is done via paired t-tests or a repeated measures analysis of variance (Dimitrov and Rumrill, 2003). Yet, a Kolmogorov-Smirnov test indicated that the required assumption of normality for the dependent variables of the experiment was violated. Hence, we used the equivalent non-parametric Wilcoxon signed-rank test for statistical hypothesis testing. We conducted Wilcoxon signed-rank tests on various combinations of pretest scores, short-term posttest scores, and long-term posttest scores for augmented and non-augmented exhibits.

Figure 7.4 graphically summarizes the results of our statistical tests. First, to rule out that differences in test scores were caused by different levels of difficulty of question sets related to augmented and non-augmented exhibits we performed a Wilcoxon signed-rank test on the pretest scores of the two groups. The test showed no statistically significant differences in median scores between the two pretest-question sets (p = 0.8649).

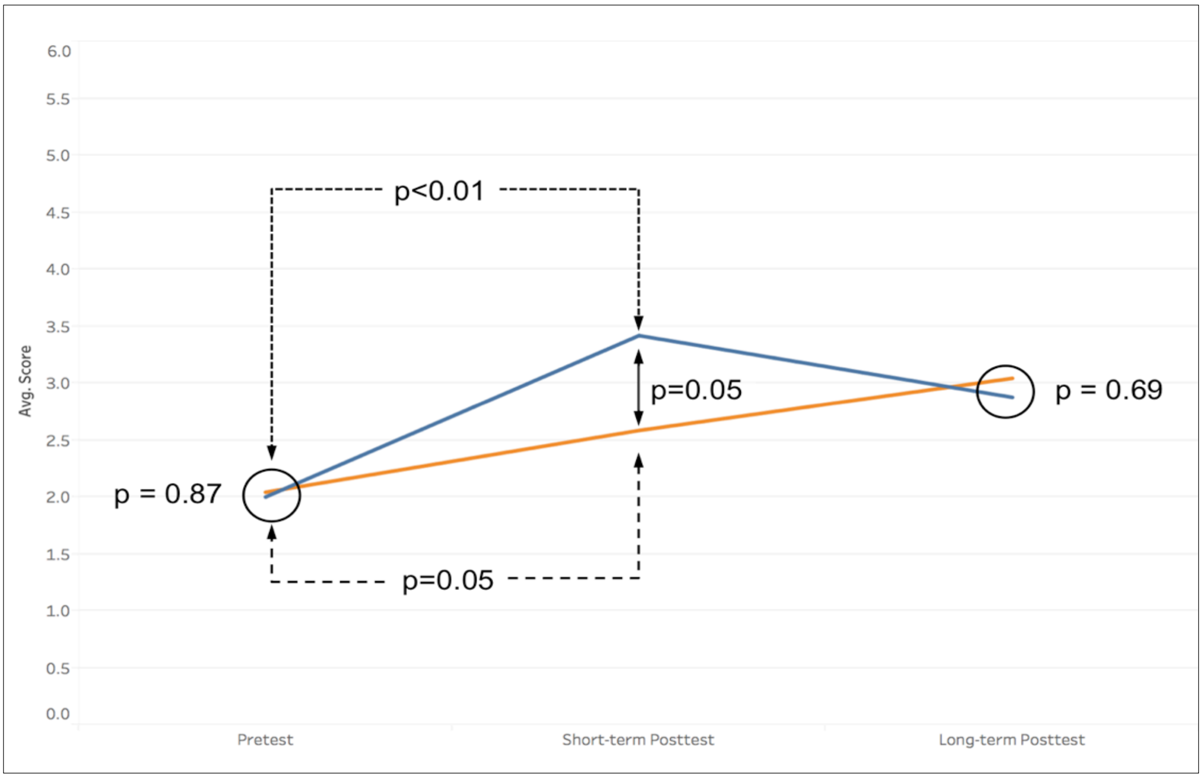


Figure 7.4: Results of Statistical Tests

Second, we compared medians of pretest scores and short-term posttest scores. For questions on both augmented and non-augmented exhibits participants showed significantly better performance at the short-term posttest than the pretest (AR: p = 0.0005, Non-AR: p = 0.0516). Hence, we can conclude that participants learned from both types of exhibits. However, the analysis also showed that participants performed significantly better for short-term posttest questions related to augmented exhibits (Mdn = 4) than for questions related to non-augmented exhibits (Mdn = 3), $p = 0.0482$. This result provides strong empirical evidence for the general effectiveness of AR as a tool for learning mathematical contents.

Finally, we analyzed the differences in test scores for the short-term (i.e., directly after the exhibition visit) and long-term (i.e., two months after the museum visit) posttests. Interestingly, we could not find significant differences in the performance of participants related to questions about augmented and non-augmented exhibits ($p = 0.6912$). This finding suggests that although participants remembered significantly more about AR exhibits than about non-AR exhibits directly after visiting the exhibition, AR seems to have no positive effect on long-term learning when comparing it with traditional, non-digital learning materials.

7.5 Discussion and outlook

To the best of our knowledge, the here presented study is the first experiment on the effect of AR on learning mathematical contents that distinguished short-term and long-term retention effects.

The empirical evidence we gathered provides support for the proposition that AR has the potential to be an effective tool for acquisition and retention of formal contents in informal learning environments – at least for short-term learning. At the same time, our results suggest that AR is not necessarily more effective than traditional non-digital learning materials when it comes to long-term learning. One potential explanation for this surprising finding can be derived from the multi-store model of human memory. As outlined before, information first enters STS through increased and continued attention to stimuli from the preattentive sensory stores and is then stored in STS in phonetic form. By embodying the design principles of CTML, AR seems to effectively support both processes. The transfer of information from STS to LTS, in contrast, largely depends on a subject's ability to rehearse, semantically organize, and integrate the newly acquired information with prior knowledge. It seems that – at least in our setting (i.e., informal learning of abstract mathematical content) – the AR materials designed by us did not effectively support these processes.

Even if we could add external validity in our field experiment through its realistic setting of the present experiment, many field experiments have to consider threats to internal validity. For example, we did not control the participants or their activities within their 90 minutes museum visit, which can be interpreted as potential confounding factor having influence on our results. Especially in self-directed learning settings as we used in our experiment, participants interactions with the exhibits differ in time and no of trials and we observed that some visitors paid more attention to either augmented or non-augmented exhibits issued in the pretest and posttest. However, this could be understood more as a positive effect derived from the use of technology rather than a threat.

Furthermore, another confounding factor results from the fact that a complete information equivalence of AR and non-AR materials could not be ensured. Thus, the AR materials were another representation of the exhibit's information. Finally, we asked the participants whether they visited the exhibition again between the two posttest activities which was denied by them, but we could not proof this. Hence our results could be effected by multiple exhibition visits.

Since the participants are from the same class and the museum visit and the posttests were part of their lessons, further influence could result from the effect of learning from test situations (Dimitrov and Rumrill, 2003; Johnson, 2010).

In conclusion, the present study contributes new insights on short-term and long-term learning with AR applications. In particular, it adds to the small but growing number of studies exploring the effective design for AR in teaching and learning. In order to confirm our findings, further studies on long-term retention should be carried out, particularly in other informal learning environments, for example, at workplaces.

Fact sheet P.3

Title	"Augmented Reality for Teaching and Learning – A Literature Review on Theoretical and Empirical Foundations"
Authors	**Peter Sommerauer** and Oliver Müller
Publication type	Conference
Publication outlet	Twenty-Sixth European Conference of Information Systems (ECIS2018), Portsmouth, UK, 2018
Year	2018
Status	Published
Full citation	Sommerauer, P., & Müller, O. (2018, June). Augmented Reality for Teaching and Learning-a literature Review on Theoretical and Empirical Foundations. In *ECIS* (p. 31).

8 Augmented Reality for Teaching and Learning – A Literature Review on Theoretical and Empirical Foundations

Abstract

Augmented Reality (AR) based teaching and learning has evolved rapidly over the past years. Researchers have shown that AR has the potential to deliver persuasive learning experiences in formal teaching (e.g., in classrooms) and in informal learning environments (e.g., museums). However, comparatively little extant research is firmly grounded in learning theories and applies rigorous empirical methods to evaluate the effect of AR on learning performance. In order to build a cumulative body of knowledge on AR-based instructional design and its effectiveness, it is necessary to consolidate both the theoretical foundations of and empirical evidence for using AR for teaching and learning. Against this background we conducted a focused systematic literature review on theoretical and empirical foundations of AR in education. We identify theory-based design elements and empirical measures for developing and applying AR teaching and learning applications and consolidate them in a design framework.

Keywords: augmented reality, learning theory, empirical studies, design framework

8.1 Introduction

Augmented reality (AR) refers to technologies that dynamically blend real-world environments and context-based digital information (Azuma, 1997). Recent advancements in mobile computing made AR systems affordable for the broad public. Today, mobile AR applications use head mounted displays, cameras, GPS sensors, and Internet access of smartphones and tablets to overlay real-world environments with dynamic, context-based, and interactive digital content.

With the publication of Billinghurst's AR compendium (2015), the breadth, depth and variety of extant research on AR was documented. A number of recent literature reviews have summarized the evolution of technology trends in the area of AR, have evaluated student learning with AR, have investigated the affordances of AR, and have reported about opportunities and challenges of AR in education (Bacca et al., 2014; Billinghurst et al., 2014; Diegmann et al., 2015; Chen et al., 2017) and in industry (Palmarini et al., 2018). However, although Chen et al. (2017) reported "that the number of AR studies in education has significantly increased", there is still a lack of re-search that is firmly grounded in learning theory and provides solid empirical evidence on how AR applications need to be designed and applied to improve learning outcomes.

On the one hand, AR is an emerging technology with high relevance for teaching, learning, and creative inquiry, thus it is expected to find broad adoption in education in the near future (Wu et al., 2013; Johnson et al., 2016). On the other hand, although in some domains (e.g., military) AR technologies have been in use for more than 50 years, it has been stated that "[d]ue to the nascent and exploratory nature of AR, it is in many ways a solution looking for a problem" (Dunleavy & Dede, 2014, p. 26) and that "relatively few research and development teams are actively exploring how mobile, context-aware AR could be used to enhance K-20 teaching and learning" (Dunleavy & Dede, 2014, p. 8).

Wu et al. (2013) aligned different instructional approaches and notions of using AR in education and emphasized the importance of learners' roles, locations, environments, and tasks. Yet, not all existing empirical studies on AR learning design are firmly grounded in learning theories. We argue that in order to build a cumulative body of knowledge on AR-based instructional design and its effectiveness, it is necessary to consolidate both the theoretical foundations of and empirical evidence for using AR for teaching and learning. Against the background of this tension between vision and reality, we conducted a systematic literature review focusing on empirical and theoretically grounded studies about the use and effects of AR for teaching and learning. Based on the findings of this review and inspired by Anderson's (2016) lens of theory applied on learning design, we propose a design framework for effective AR-supported teaching and learning.

The remainder of this paper is structured as follows: We first present the method used for our literature search and provide theoretical background on learning theories. We then outline the selection and analysis processes we applied to translate our findings into a reference framework for effective AR-based instructional design. After discussing our results, we conclude with a brief summary and directions for future research.

8.2 Literature search

Our systematic review was based on a database-driven literature search at the IT University of Copenhagen between May and October 2017, including the scientific databases ACM, Business Source Complete, IEEE/IEE Electronic Library (IEL), Lecture Notes in Computer Science (LNCS) and Lecture Notes in Artificial Intelligence (LNAI), SAGE Journals, Springer, and Taylor & Francis. We followed the stages provided for systematic reviews by Gough et al. (2017), consisting of the following four key activities:

1. Propose a research question
2. Ascertain and qualify relevant research
3. Critically evaluate research articles using a systematic and comprehensible process
4. Run a conclusive analysis and draw a final claim

In our focused literature review, we aimed at finding and analyzing research articles that (a) document empirical studies in which AR was used for supporting teaching and learning and (b) are grounded in learning theories. Therefore, we defined the following research question:

RQ1) Which learning theories are used as the basis for designing AR applications for teaching and learning?

8.2.1 Search strategy

Based on the above defined research questions, we used the search term "augmented reality" AND "theory" AND ("learn* OR teach* OR educat*") to retrieve relevant literature. This choice was driven by the focused nature of our literature review. We specifically aimed to identify studies that grounded the design and evaluation of AR applications in theories from the field of education. We are aware that through this strategy we may have missed some studies that only implicitly refer to extant theory (i.e., they do not explicitly contain the word "theory"), but as the search term produced a relatively large number of hits, this strategy seemed plausible. We limited the search to peer-reviewed scientific articles and initially found 325 database entries in various languages. Focusing on articles available in English, the database showed 291 results. After skipping duplicates and erroneous entries, like articles still not in English language or that are not peer reviewed (although this was indicated in the meta data), 184 sources remained.

8.2.2 Article selection

In the next step, the 184 articles were selected for cataloguing. Each article was represented by an internal serial number, the title of the article, the subjects provided by the database search tool or, when no information were given, the article's keywords. Additionally, we collected information about the source (i.e., journal or conference article), the year of publication, and added information about the main topics covered by the article. When provided in the abstract or introduction section of the article, we added information about the learning topic, the learning theory, the target group, and the number of participants in the study.

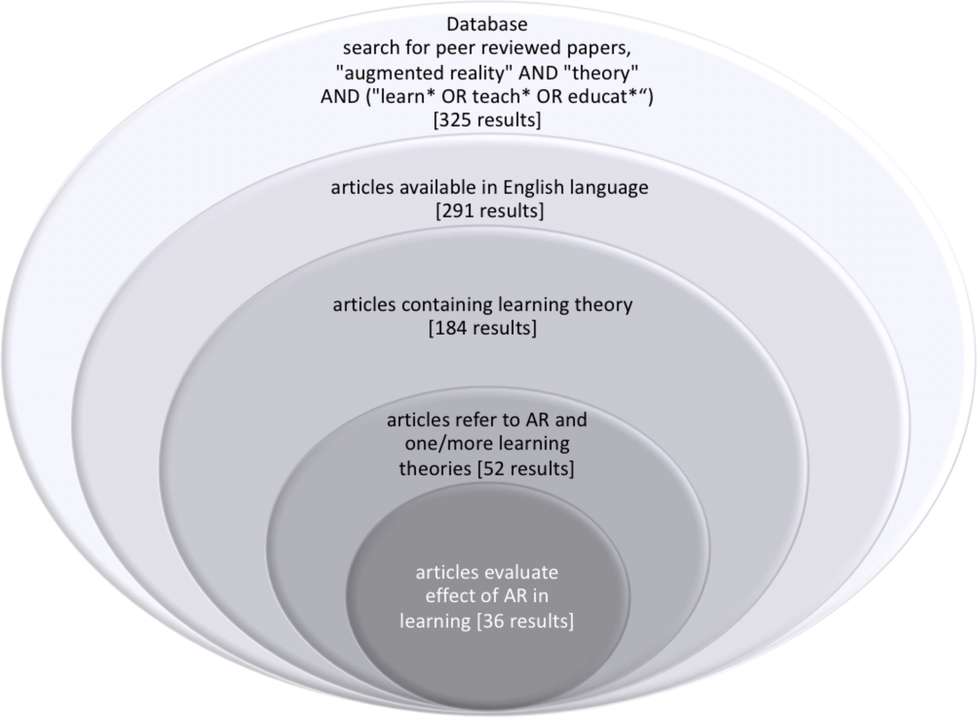


Figure 8.1: A visual representation of the literature search process

We continued the selection process by browsing through the theoretical and empirical sections of the articles. Sources not providing any information about how learning theories were applied or not showing evidence for the design and use of AR technology (i.e., purely conceptual articles) were excluded. This reduced the list of articles to 52 articles. In a final screening step, we excluded articles that did not evaluate the effect of AR on learning, which left us with a final list of 36 relevant articles. Table 8.1 summarizes our inclusion and exclusion criteria.

Table 8.1: Exclusion and inclusion criteria

Exclusion Criteria	**Inclusion Criteria**
Not a scientific source	Article was peer-reviewed
Not in English language	Article was original research in English language
Missing theoretical foundation / learning theory	Article refers to a learning theory
Not related to development of AR	Article focuses on AR development
Purely conceptual paper	Article contains a section on AR in practical use
Do not provide evaluation results	Article contains empirical results of testing AR

8.3 Data analysis

Our data analysis process was supported by the application of an analysis and synthesis method using a concept matrix to extract the findings (Whittemore & Knafl, 2005). We followed a concept-centric approach (Webster & Watson, 2002) and categorized the sources primarily with regards to learning theories (see Section 3.1) and learning performance measures (see Section 3.2). In addition, we extracted information about learning topics, learning environments, research methods, and standard article metadata.

8.3.1 Learning theories used in AR for teaching and learning

To answer research question RQ1, we categorized all sources according to learning theories. Over the past 100 years, many complementary understandings and theories about teaching and learning have emerged (Illeris, 2009, p. 7). Hence, we applied the desk reference for learning theories published by Illeris (2009) to organize learning theories. Following this approach, learning theories can be presented from a cognitive, behavioral or constructivist perspective, constituting the cognitive, social, and emotional dimensions of learning (Illeris, 2009, p. 8 ff). Accordingly, we first pruned the variety of learning theories to these three main categories and later expanded them with subcategories representing specific learning theories identified in our review. The numbers in brackets refer to the sources from the literature review (e.g., [17]).

Behaviorism is a learning paradigm that basically follows the idea of controlling and modifying a learner's behavior and acquisition of basic facts and skills using stimulus-response pairs and selective reinforcement (Illeris, 2009). From a psychological perspective, Skinner (1974) emphasized that behavior is the basic element of learning and argued that learning theories impede the empirical research on behavior theory. Thus, radical behaviorism is rather a method of experimental analysis than a learning theory (Skinner, 1974). Even more restrictive, Jarvis (in Illeris, 2009, p. 32 ff) rejects the idea of behavior as a driving force for human learning because of the ignorance and absence of meaning in learning. In our literature analysis, we could not find a single article using behaviorist theories as the foundation for the design of AR learning evaluation.

Cognitivism is a learning paradigm that sees the mind as an information processor and focuses on internal cognitive structures to understand how information is received, organized, stored, and retrieved in the brain. Pedagogically seen, the processing and transmission of information can be executed through communication, explanation, recombination, contrast, inference, and problem solving, mainly to acquire external, existing knowledge (Anderson 1983; Wenger 1987; Hutchins 1995; Illeris, 2009, p. 203). In our data literature analysis, two cognitivist theories were identified, namely the cognitive theory of multimedia learning (CTML) by Mayer (2009) and the embodied cognitive dissonance theory [86]. In his CTML, Mayer presents three views of multimedia learning. The delivery-media view refers on the presentation of learning content using two or more devices for delivery. The presentation-modes view focus on the presentation of material using two or more presentation modes. Finally, the sensory-modality view posits that two or more sensory systems in the learner are involved. On this basis, he proposes twelve principles of multimedia design which are applicable on multimedia learning (Mayer, 2009). Not all of those principles seem to be applicable for AR learning, as shown by the studies that implemented CTML in their AR learning applications [15, 17, 19, 28, 45, 105, 172]. While seven (19%) research articles are referring to Mayer's CTML, only one study referenced the embodied cognitive dissonance theory, integrating the influence of embodied learning (e.g. effects and actions of the body and its movements) and grounded in cognition theory [86]. The authors posit, that learning takes place using a multimodal link between perception and action by coupling the environment and the brain.

Constructivism is a learning paradigm that focuses on the processes by which learners build their own mental structures when interacting with an environment (Richard, 2015; Illeris, 2009). Following a task-oriented pedagogical view, constructivism equates learning with creating meaning from experience, where learning is more meaningful to students when they are able to interact with a problem or concept. Constructivist learning theories emphasize the need to actively engage students in problem solving and motivate them through meaningful contexts. Constructivism utilizes interactive teaching strategies to create meaningful contexts that help students to construct knowledge based on their own experiences. Learning tasks are implemented by using high-order thinking skills and transferring knowledge into new situations, like used in simulated worlds, role-plays, for debating, cooperative learning, and self-directed task-based learning (Piaget, 1954; Papert, 1980; Illeris, 2009). About 28 (78%) articles were related to constructivist learning theories. In this group, we identified four subcategories, namely, situated learning, game-based learning and simulations, experiential learning, and other learning theories (e.g., variation, transformative, collaborative, and meaningful learning). Especially in terms of situated learning our categorization does not align with former assignments, i.e. Dunleavy & Dede (2014). Although they state in their review, that situated learning "… extends other learning theories …" and that "… learning is a co-constructed, participatory process …“, they categorized situated learning as a learning theory on the same level next to constructivist learning theory. We understand their first argument in the way that it can also be concluded from other learning theories in our subcategories, i.e. game-based learning or simulation. Following their second argument, we understand that learning based on constructivist theory rely on constructed processes which supports the learning experience. Consequently, we allocated the four identified subcategories to constructivist learning theory.

An unexpected finding was that nine papers were referencing to more than one theory from one paradigm [6, 26, 40, 41,52,109, 110, 112, 124], and that two are referencing both cognitivist and constructivist theories [19, 28]. Consequently, creating an effective learning experience may require to incorporate ideas from more than one learning theory.

8.3.2 Measuring the effect of AR on learning performance

In addition to answer research question RQ1, we worked through the articles identifying measures of learning success. The most applied measures were (the numbers in brackets refer to the sources):

- number of fulfilled tasks [1, 6, 7, 19, 20, 26, 29, 40, 51, 86, 104, 109, 157, 172, 173]
- number of right/wrong answers in a given time frame [1, 6, 7, 19, 20, 29, 40, 51, 86, 109, 172, 173]
- number of right or wrong answers [1, 6, 15, 45, 51, 52, 110, 124,125]
- answering time [45, 86, 105, 109, 172]

The answering time and number of fulfilled tasks were measured either by a human assessor or using the application itself to collect the data. To ask questions related to the learning content, most studies used tests and questionnaires to collect the number of right or wrong answers. In most cases where game-based learning and simulation were implemented, the target measures were number of fulfilled tasks given within a time frame.

Many studies also used measures of user perception to evaluate the apps:

- perceived usefulness [4, 7, 8, 15, 45, 51, 52, 104, 123, 124]
- perceived learning [7, 19, 20, 26, 40, 41, 51, 52, 110, 123]
- perceived satisfaction [15, 20, 26, 41, 51, 52, 125]
- application usability test (i.e. Shneiderman, 2010; Lewis & Sauro, 2017) [15, 27, 28, 41, 75, 123].

With regards to perceived usefulness, perceived learning, and perceived satisfaction, all studies applied a questionnaire to collect data. Some studies implemented a usability test to measure the usability of both, the application of AR and the learning content. In summary, we found no article within our literature findings that did not report any positive impact resulting from the application of AR.

Most participants in the target group were students in the age range starting from 8 years to 24 years. A wider variety within the participants were found in two studies: one field experiment in a math exhibition at a museum with 101 mixed participants between 15 and 78 years [15], and a laboratory experiment having technicians between 21 and 58 years receiving a technical training for robot programming [172]. We synthesized the learning subjects from the articles and organized them according to learning topics as shown in Table 8.2.

Table 8.2: Learning topics and subjects

Business & Mangement	Creative Arts & Media	Health & Psychology	History	Languages & Cultures	Nature & Environment	Science, Engineering & Maths	Study Skills	Tech & Coding
Economic development [173]	3D-objects [3, 123]	Healthcare training [8]	Buildings of tobacco warehouse [51]	Chinese cultural festivals [106]	Urban planning simulation [112]	Science learning and motivation [17, 19, 28, 54, 77, 104]	Career decisions [173]	Programming [6]
Tourism [110, 112, 123]	Architecture [44]	Tolerance [173]	History [29, 110, 123]	Cultural development [110]	Ecosystem, ecology [4]	Electrical engineering [41]	scientific thinking [77]	Robot's posture [172]
	Art - theory [157]	Solidarity [40]	Architecture [123]	English Filipino German Spanish vocabulary learning [1, 45, 124]; Multiculturalism [52]	Environmental science [77[	Mathematics [15]	Natural reading and writing [75]	Magnetic fields [54]
	Guitar playing [86]				Water circle and quality [7, 125]	Astronomy, solar system [26, 105]	Library knowledge [20]	Electronics [109]
					Botanical garden [4]	Microbial science [27]	Educational choices [3]	

8.4 Framework for an effective design of AR for teaching and learning

In the following, we will synthesize the results of the literature review in a design framework. To better link the theoretical and empirical foundations, on the one hand, with the form and function of concrete AR apps, on the other hand, we once again scanned the sources and extracted design elements that can be traced back to both abstract learning theories and concrete system features. Therefore, we examined features like learning requirements from theory, design aspects, measures, and measurement parameters. We also arranged learning theories and design elements in a logical and hierarchical structure. The extraction process started with identifying applied learning theories in each study. We categorized the research articles accordingly and analyzed the implemented design elements within each theory group, focusing on content preparation and evaluation of the learning activities and the measurement of the learning results. The details will be outlined in the following sections.

8.4.1 Design elements

Table 8.3 gives an overview of design elements organized by the theories they were derived from. The studies that have been based on CTML use different subsets of the twelve design principles outlined in this theory to translate its basic ideas into concrete features or AR apps. Sommerauer and Müller [15], for example, focused on four CTML principles by aligning physical objects and virtual content in space (spatial contiguity principle) and time (temporal contiguity principle), using trigger images in a math exhibition (signaling principle), and playing spoken words instead of displaying written words (multimedia principle) to reduce learners' extraneous cognitive load and thereby enhance their cognitive information processing processes. We found exactly the same implementation of CTML at Parhizkar et al. [28], and Santos et al. [19]. Santos et al. applied CTML as key learning theory to prepare the learning content in their Augmented Reality Learning Environment (ARLE) [19].

Design elements derived from constructivist learning theory, like proposed in Carlson & Gagnon's conceptual model [8] or in the three-phase learning model from Parhizkar et al. [28], point towards elements derived from interface design and mobile learning. Introducing mobile aspects in learning design bursts the boundaries of a single location and enables the integration of location awareness, e.g. as described in Tseng et al. (2001) and Wu et al. (2010). Such design elements would be, e.g., to include maps and features indicating objects of interest nearby, or simply to invite students to move in class, like Furió et al. [125], or to visit a specific place, as Kamarainen et al. [7] implemented in their study. We found further support for this approach in other studies [1, 20, 27, 52], but the features were described in less detail.

Furthermore, a central aim of applying AR in learning environments is to turn simple learning into a motivational learning experience [1, 3, 4, 7, 8, 17, 19, 20, 26, 27, 41, 45, 52, 77, 104, 106, 110, 123, 124, 173]. In our literature review, nearly a third of the studies implemented design elements from game-based learning and simulation, and a third from experiential learning, to achieve this goal. Prensky (2001) provided a comprehensive list of features, e.g., motivating aims that affect enjoyment, intense and passionate involvement, structure by rules, goals, interactivity and variability, feedback and gratification, competition and emotion (p. 6), supported by design elements like storytelling, accomplishing missions, or implementing variation using mini-games between learning steps. Chen & Tsai [20] used a database for game story and learning process data to analyze learner's performance and their gaming skills. Furthermore, design elements as leader-boards or badges, points and rewards are implemented to "encourage students to have fun and perform a learning" (Vizent et al., 2015, p. 1087).

Experiential learning is constructed in a rather process-oriented way than through single tasks. Following Kolb (2014), such a process starts with a concrete experience, which is deepened including an observation and reflection step, followed by abstract conceptualization to lead to further active experimentation. Studies from our literature search implemented, e.g., storytelling as a first instruction, followed by observations in real-world environments (e.g., botanical garden, museum, in nature), either in single or group-based tasks to later share their findings with their colleagues in class and gain further experiences in a reflection and discussion session, to finally experiment with their own and other's findings [4, 7, 27, 41, 52, 173]. situated learning incorporates learning at specific places, e.g., in a library [20], at home [45], at a botanical garden [4], in nature [7, 105, 125] or special areas in a town [29, 51, 77, 106]. The design elements therefore influence and derived from the environment itself, such as, atmosphere, impression, environmental and real-world experiences, and integrate discovered objects. In addition, situated learning entail several learning activities into a learning sequence. A main difference in implementing AR learning and between our literature findings is, that learning can be an isolated individual task or a collaborative activity. Thus, the requirements for single user learning environments are different from multi-user environments.

Learning theory	**Design elements for implementation**
CTML	Meyer (2009): 12 principles of multimedia design consider coherence, signalling, redundancy, spatial contiguity, temporal contiguity, segmenting, pre-training, modality, multimedia, personalization, voice, image; e.g. aligning physical objects and virtual content in space and time, using trigger images, play spoken words instead of displaying written words;
Mobile learning	Herrington et al. (2009): Design principles for mobile learning: real world relevance, mobile contexts, explore, blended, whenever, wherever, whomsoever, affordances, personalise, mediation, produse; e.g. to including maps and features indicating objects of interest nearby, inviting students to move in class or visiting a specific place;
Game-based learning & Simulation	Hirumi et al. (2010), Kiili (2005), Prensky (2001), van Eck (2006): interaction, navigation, drama and presentation, storytelling, 3-dimensional, HCI (human controller interface), programming, pattern analysis, visual content analysis; e.g. storytelling, accomplishing missions, implementing variation (e.g. using mini-games), leaderboards, badges, points, rewards;
Experiential learning	Kolb's cycle of experiential learning (2014): how information is understood and processed: diverging (feel and watch), assimilating (think and watch), converging (think and do), accommodating (feel and do); e.g. instruction, observation, reflection, experimentation using examples which connects to real-world;
Situated learning	Mc Lellan (1996): including stories, reflection, cognitive apprenticeship, collaboration, coaching, multiple practices, articulation of learning skills and technology [123]; e.g. environmental influence, atmosphere, impression, experiences

Table 8.3: Learning theories and design elements for implementation

Some researchers added group-work activities (e.g. interaction between students, exploring different aspects to combine findings) in their design of the AR learning app, but often split the group tasks into individual tasks to be fulfilled by different user roles, e.g. Koutromanos & Styliaras [29] and Lundblad et al. [40], to present their findings after the AR training as a group result. Those studies incorporated no functions or tools to support active collaboration and communication inside the AR learning experience. The group activities were later realized outside the AR experience in the real environment, e.g. by having a group discussion [4, 7, 29, 52, 125]. In studies were real collaborative learning was introduced, the communication between the learners and their activities was labelled as crucial [19, 124]. Therefore, Ibáñez et al. [124] used a developer environment for multiplayer games for their AR app development. They implemented their own interface and communication module to migrate objects and users between multiple virtual environments (and the real world). However, no study incorporated any standard in their AR app design, especially for internal and/or external communication and capturing learner's behavior, activities, and results. Consequently, to design a comprehensive and persuasive frame-work we conclude to incorporate design elements covering the internal and external communication in AR learning applications for both, user and object communication, based on a well-established standard.

8.4.2 A conceptual framework

Learning is a complex process and today's concept of learning goes far beyond a simple acquisition of knowledge and skills (Illeris, 2009, p. 1). Also, emotional and social dimensions need to be included for future learning, whereas "… all learning implies the integration of two very different processes, namely an external interaction process between the learner and his or her social, cultural, or material environment, and an internal psychological process of elaboration and acquisition" (Illeris, 2009, p. 8). With our modular conceptual framework, we target to support this complexity of learning.

Based on the results of our literature review we built a design framework that is based on learning theories and additionally considers the conceptual view of Anderson of how learning and learning designs can be enhanced using emerging technologies and applying learning theories (2016, p.47). Our design framework reflects that learning consists of one or more learning sequences, which contain one or more learning activities. On the one hand, with this notion we strive to support the application of teaching-learning sequences (Méheut & Psillos, 2004), on the other hand, we enable the implementation of expanding-seeding and contracting-soloing learning sequences, as explored in organizational knowledge management systems (Bingham & Davis, 2012). Figure 8.2 shows a graphical representation of the proposed design framework.

The core of a learning activity in our model is to impart information and knowledge, represented as learning content. At the content layer, the learning content should be prepared applying Mayer's CTML and following any subset of the twelve principles of multimedia design (Mayer, 2009). Parhizkar et al. [28] structure their system design according to a three-phase learning model, which is partly based on CTML's design principles for content design, interface design based on constructivist learning theory, and structure design based on mastery learning strategy. We argue that the major design elements for content creation should focus on the learning content itself and be supported by cognitive theories and, therefore, be independent from aspects derived from other, e.g. constructivist, learning theories. Studies in our review that successfully constructed their learning content in this way are [15, 19, 28, 45, 105].

The integration of mobile aspects in the mobile-layer should be considered right after the step of content creation. Since most AR learning applications use mobile devices, our model includes design principles for mobile learning (Herrington et al., 2009), as demonstrated by Furió et al. [52]. At this, it is crucial to consider whether a learning application is used immovable or mobile, because this affects further design elements at the motivation level. The motivational layer considers aspects of game-based learning, simulation-based learning, and experiential learning, especially in terms of interaction and navigation and communication within and collaboration between learning activities [6, 8, 77, 123, 124]. This also lays the basis to support Kolb's elements for experiential learning [4]. Therefore, this layer includes a communication interface to collect and exchange information about users' learning experiences within the learning activity.

Finally, one or more learning activities can be assembled into a learning sequence. We included design elements following situated learning theory (which also includes collaborative learning), as proposed by McLellan (1996) and applied in the study by Chang & Jen-ch'iang [123]. Inside a learning sequence, elements of coaching, collaboration, and reflection should be included, as well as the application of multiple practices, learning skills, and technology.

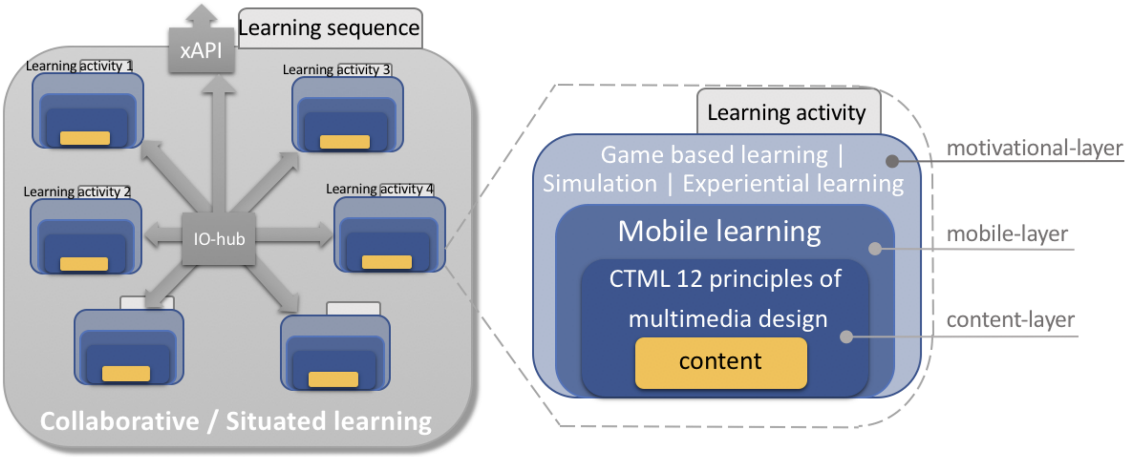


Figure 8.2: A graphical representation of the design framework

8.4.3 Internal and external information exchange about learning experience

Connecting two or more learning activities in a learning sequence enables learning to be treated as a process. In most of the studies we analyzed, the definition and management of such processes was handled in the app itself. As already mentioned in the introduction section, future trends for AR in teaching and learning focus on measuring and evaluating learning in personalized student learning experiences, which requires to collect information about learner's behavior, activities and results and to exchange this information with external systems. Both, the internal and external process management and the communication between learning activities and with learning management systems can then be supported by the implementation of a so-called input-output hub (IO-hub). Such an approach was not implemented in any study we have analyzed, but addresses an important issue that needs to be considered to use AR in learning successfully [3, 28, 44, 75]. Lundblad et al. [40] classified data exchange as a main advantage of using mobile AR, but in their application design they stored collected information locally and evaluated the results using interview-based questionnaires. When Ternier et al. [110] discussed issues encountered in terms of their mobile app's connectivity, they concluded that a network independent version would be a possible solution. However, communication between AR apps and external entities allows to collect data in form of activity statements describing the learner's behavior during the learning sequence. Thus, it addresses current challenges and future trends in AR teaching and learning.

Whenever data is collected and exchanged between entities, using established standards supports the integration of the application into various technology environments. According to the Experience Application Programming Interface (xAPI) specification, which is designed to support the information collection of formal and informal distributed learning activities (Kevan & Ryan, 2016), a data set describes single activity statements and is stored in a learning record store. In such way, series and different types of experiences can be collected to be analyzed later, including data, e.g., from wearables, mobile applications, workplace environments, and geo data (Silvers, 2017).

From a theory perspective, the xAPI specification is influenced by the socio-cultural framework Activity Theory (Silvers, 2017) and in close alignment with constructivist learning theory. For applying the xAPI standard in a learning sequence it is recommended to include it early in the design process. Thus, constructivist-aligned strategies are implemented from design through evaluation of the learning activity (Kevan & Ryan, 2016). Consequently, the data acquisition and data analysis then follow the main aspects of constructivist learning theory.

8.4.4 Practical application

To apply the framework in practice we mapped the design elements in retrospective to selected use cases from our literature review.

Studies that focused on content preparation implemented design elements derived from CTML [15, 17, 19, 28, 45, 105, 172]. Other studies documented their approach for content preparation less detailed but referred to CTML in a general way. For example, Santos et al. [19] stated that "multimedia learning theory provides a learning theory of how real-world annotation by AR can help students learn better based on human cognition and related processes in the brain".

Zhang et al. [26] created a 3D interactive learning environment for astronomical observation instruction using mobile AR for "3D depth and scale of astronomical and stellar configurations" for outdoor observations. Their key development principle was based on mobile learning theory, "because portable mobile devices are used, operations can be conducted and the device can be transported with no limitations on location, which reduces the influence of environmental variables on astronomical observation instruction". Moreover, they identified portability as a success factor lowering the limitations of the environment, since traditional learning software was mainly used on desktop computers or laptops, which "effectively exclud[ed] usage outdoors" [26]. In addition, they could include data derived from mobile device sensors and functions like geo positioning, compass, and angle finders "to engage in outdoor stargazing".

Studies which referred to game-based learning, like Chen & Tsai [20], Koutromanos & Styliaras [29], and Squire & Mingfong [77], implemented design elements, such as roles, stories and challenges, different places, interactive objects and tools, after they had prepared their curriculum. Similarly, Furió et al. [125] compared traditional learning with mobile game-based learning by means of mini-games. They built their lesson upon an already existing traditional classroom lesson to compare the student's performance between classroom learning and game-based learning. The learning content was prepared, e.g. by using an introduction video and several mini-games built upon mobile design elements (e.g. moving in classroom, finding QR-codes used as trigger images for specific objects), to create a game-based learning experience.

Those studies that used design elements from situated learning theory implemented these elements in addition and after the creation of single learning activities, e.g. Kamarainen et al. [7] (aid students in their understanding and interpretation of water quality measurements) and Koutromanos & Styliaras [29] (introduce buildings, their specific history and architecture). Both studies incorporated multiple learning and collaboration activities, which were designed sequentially, considering the learning content, mobile learning aspects, and game-based elements, and combined into a single situated learning experience. However, the information exchange between the single learning activities and the coordination (i.e. the sequence of the activities) were organized using traditional tasks lists, thus not implemented as a functionality within the learning application itself. Similar, elements for collaborative learning are introduced in traditional tasks, e.g., preparing group presentations in classroom.

8.4.5 AR learning and knowledge management

The integration of our proposed design principles into the organization and development of learning applications, e.g. for workplace trainings, contributes to the digital transformation of using and sharing organizational knowledge for both school and higher education and professional education. Thus, and addressing King's (2009) understanding of knowledge management, the design framework can be applied in education for the planning, organization, motivation and controlling of people, and to support educational processes and systems in organizations (e.g., companies and universities). Moreover, considering the framework's elements for designing a training application enables to identify and specify delivery points for submitting information during a learning activity and sequence. In this way, learning assessment becomes an integral component of the learning application design.

At this, the key to success lies in collecting the learner's data, so that it becomes possible to analyze this data to provide immediate feedback for the learner, but also for teachers and trainers at the organizational level. Integrated into learn management and knowledge management systems and combined with learning analytics tools it will potentially impact organizational learning and support knowledge management in education and training on several levels.

To apply the framework in practice, we illustrate the utilization in a use case from industry. A wide variety of companies present their organization, products, and services at international trade fairs. Operators of such fairs often need to ensure against the authorities that their staff has been instructed in safety and security issues, also considering local laws and regulations. Yet, such instructions are given either in preparation (homework) or at the office with the support of media-based trainings and followed by assessments based on multiple-choice tests. AR apps following the proposed design framework could support this task differently and more comprehensively. For example, the learning content could be provided in a cognitively efficient way (considering design elements from the content-layer), mobile features could enable learners to move around and investigate the environment from different perspectives (considering design elements from the mobile-layer), and in simulations, demonstration, and practice in the real environment (e.g., game-based elements at the motivational-layer). Furthermore, applying internal processes, tools, and equipment that is effectively utilized in practice (e.g. in learning sequences) leads finally to constructivist learning. E.g. in such simulations, safety and security procedures can be trained in that way and to ensure that people behave correctly in case of emergency. The AR learning app reports the results from learning activities to external entities in order to contribute to the need for documentation of the learning results. In such a way, the AR app also supports the organization's knowledge management system.

8.5 Conclusion and discussion

As shown in the previous sections, we identified the most frequently applied learning theories for designing and using AR for teaching and learning. Furthermore, we examined the remaining set of research articles concerning learning effects in relation with learning theories. However, most studies in our analysis did not apply design elements derived from learning theory in their measurement of learning. These measures can be broken down to the two, needed time and number of correct answers. In terms of analyzing the effect of AR in learning, most of the studies uses external scales and questionnaires for evaluation.

Resent publications, e.g. Bacca et al. (2014), Billinghurst & Lee (2015), Chen et al. (2017), Diegmann et al. (2015), Dunleavy & Dede (2014), Palmarini et al. (2018) and Wu et al. (2013), have illustrated that AR research spans a broad spectrum of objectives and methods. Yet, in our literature review we found only few studies that systematically rely on learning theories and empirical measures to design and apply AR for teaching and learning. Reflecting on our results shows that very few papers are reporting negative effects [172] or no positive effects [3] of using AR for learning. If negative findings are reported, they are referring to side aspects of using AR, e.g. cognitive overload [17, 109, 110], but they found a positive effect on learning [4, 20, 104,172]. However, developers of AR learning applications could learn also from negative research experiences, thus we would appreciate seeing more research results showing both, positive and negative effects.

With our proposed design framework, we aim to support the development of effective AR learning applications and which are applicable for various target groups, educational branches, and settings. Moreover, we considered the need of integrating AR learning into a manageable environment by adding design elements supporting internal and external communication and assisting instructional and organizational governance. This is strongly needed, as current and future developments on a technology level will provide new AR devices, e.g., full hands-free AR, gesture recognition, recognition of facial expressions, audio and video capturing and analysis.

Of course, our study is not free of limitations. Since we used a strict approach for identifying relevant research articles, we have probably missed articles that would have fitted to our search profile. However, we attempted to compensate for this by also relying on related recent literature reviews. A further limitation of our study and the proposed design framework could be that it only includes design elements derived from learning theories and does not incorporate additional design aspects theories not directly related to learning. Nevertheless, real world annotation, contextual visualization and vision-haptic visualization are the main strengths of AR, which are all sup-ported by learning theories (Santos et al., 2014).

Currently, we are developing AR learning applications based on our design framework. Our future research will demonstrate the effectiveness and possible constraints of the framework and its application. Since we did not explicitly include elements for assessing the learning performance, design elements considering the measurement of learning could be a future extension, especially for measuring learning results. However, the inclusion of the xAPI in the framework already supports the collection of a learner's data, thus enables the evaluation and analysis of learning results.

Appendix: Table and overview for coding

	ID	Source	Year	pos. / neg. effect	formal / informal	Empirical method		Cognitivist		Constructivist			
							Behaviorist	CTML	grounded cognition	situated	game-based learning, simulation	experiential learning	other
1	1	conference	2017	pos.	formal	Classroom experiment				1			
2	3	conference	2013	none	informal	Field experiment						1	
3	4	journal	2016	pos.	informal	Field experiment						1	
4	6	journal	2015	pos.	formal	Laboratory experiment					1	1	
5	7	journal	2013	pos.	formal	Field experiment				1		1	
6	8	journal	2016	pos.	informal	Laboratory experiment				1	1		
7	15	journal	2014	pos.	informal	Field experiment		1					
8	17	journal	2017	n. a.	n. a.	Literature review		1					
9	19	journal	2014	pos.	n. a.	Literature review		1				1	
10	20	journal	2012	pos.	informal	Field experiment				1	1		
11	26	journal	2014	pos.	formal	Field experiment					1		
12	27	conference	2017	pos.	formal	Laboratory experiment						1	
13	28	conference	2012	pos.	informal	Literature review		1					1
14	29	conference	2015	pos.	informal	Field experiment				1	1		
15	40	journal	2012	pos.	informal	Field experiment					1		1
16	41	journal	2015	pos.	formal	Laboratory experiment						1	1
17	44	diss/thesis	2015	pos.	n. a.	Observation				1			
18	45	journal	2016	pos.	formal	Laboratory experiment		1					
19	51	journal	2015	pos.	informal	Field experiment				1			
20	52	journal	2013	pos.	formal	Laboratory experiment						1	1
21	54	journal	2012	pos.	formal	Observation							1
22	75	journal	2013	pos.	formal	Laboratory experiment				1			
23	77	journal	2007	pos.	informal	Field experiment					1		
24	86	journal	2014	pos.	informal	Field experiment			1				
25	104	diss/thesis	2003	pos.	n. a.	Questionnaire							1
26	105	conference	2012	pos.	informal	Observation		1					
27	106	conference	2014	pos.	informal	Field experiment				1			
28	109	journal	2010	pos.	informal	Laboratory experiment					1		
29	110	journal	2012	pos.	informal	Laboratory experiment				1			1
30	112	journal	2005	pos.	informal	Observation					1		
31	123	report	2013	pos.	informal	Field experiment				1			
32	124	journal	2012	pos.	informal	Field experiment				1		1	
33	125	journal	2015	pos.	formal	Laboratory experiment					1		
34	157	journal	2013	pos.	informal	Field experiment					1		
35	172	conference	2010	neg.	formal	Laboratory experiment		1					
36	173	journal	2010	pos.	formal	Survey						1	
∑								7	1	12	11	10	7

Fact sheet P.4

Title	"The Effect of Marker-less Augmented Reality on Task and Learning Performance"
Authors	**Peter Sommerauer**, Oliver Müller, Leonard Maxim & Niels Østmann
Publication type	Conference
Publication outlet	14th International Conference on Wirtschaftsinformatik (WI2019) (nominated for best paper award)
Year	2019
Status	Published
Full citation	Sommerauer, P., Müller, O., Maxim L. & Østman N. (2019). The Effect of Marker-less Augmented Reality on Task and Learning Performance. In *14th International Conference on Wirtschaftsinformatik (WI2019), February 24-27, 2019, Siegen, Germany,* pp.1696-1710.

9 The Effect of Marker-less Augmented Reality on Task and Learning Performance

Abstract

Augmented Reality (AR) technologies have evolved rapidly over the last years, particularly with regard to user interfaces, input devices, and cameras used in mobile devices for object and gesture recognition. While early AR systems relied on pre-defined trigger images or QR code markers, modern AR applications leverage machine learning techniques to identify objects in their physical environments. So far, only few empirical studies have investigated AR's potential for supporting learning and task assistance using such marker-less AR. In order to address this research gap, we implemented an AR application (app) with the aim to analyze the effectiveness of marker-less AR applied in a mundane setting which can be used for on-the-job training and more formal educational settings. The results of our laboratory experiment show that while participants working with AR needed significantly more time to fulfill the given task, the participants who were supported by AR learned significantly more.

Keywords: Augmented Reality, Learning, Mobile Application

9.1 Introduction

Augmented Reality (AR) is known as a technology which augments the real environment with relevant digital information (Azuma, 1997). Such information can be superimposed on recognized objects using smartphones, tablets or AR goggles as user interfaces between the real and the virtual world. Additionally, AR allows a full 3D view of virtual objects and enables users to interact with them.

AR's potential has been shown in many use cases and in various settings, such as informal and formal learning environments, workplaces, museums and natural environments (Akçayır and Akçayır, 2017; Bacca et al., 2014; Billinghurst et al., 2015; Chen et al., 2017, Diegmann et al., 2015; Dunleavy and Dede, 2014; Palmarini et al., 2018; Radu, 2014; Sommerauer and Müller, 2018; Van Krevelen and Poelman, 2010). In most settings which have been studied so far, trigger images or QR codes have been used for identifying objects in order to superimpose digital information on them (Bacca et al., 2014). Only few applications exist that use so-called marker-less AR (Bacca et al., 2014). Marker-less AR works in a way that the real environment itself and real objects therein are recognized by the app, which then augments digital information and adds functionality to the digitally enriched objects and environments (Van Krevelen and Poelman, 2010), without any pre-defined trigger images or QR codes.

In this study we investigate the application and effectiveness of marker-less AR to support both the execution of a specific task in a mundane setting and the learning about the underlying domain by executing the task (i.e., learning-by-doing). In particular, we intend to answer the following research questions:

- RQ1: How can marker-less AR be implemented in a real-world environment?
- RQ2: How does marker-less AR affect task and learning performance?

In the pursuit of answering our research questions, we developed a marker-less AR app, which enables the user to learn the names of objects from the real environment. We created a fictional learning situation with a given task and compared the results from two groups, one using an AR-based tool, the other using a traditional paper-based tool (Note that a direct comparison between marker-less and marker-based AR is not the aim of this study). Hence, our laboratory experiment uses a static group design with an experimental group and a control group. With this design we intent to investigate the differences in task and learning performance of the two groups by measuring task performance (i.e., time required for completing the task) and learning performance (i.e., answering a post-test questionnaire with questions about the task).

The remainder of this paper is structured as follows: To prepare the background, we first present related work and provide theoretical background on marker-less AR and its implementation. As our study was motivated by investigating task performance and learning performance, we also present associated performance metrics that are derived from learning theories. We then outline the app development process along with the embedding of a number of theory-ingrained design principles, followed by an introduction of the used dataset for image recognition and the setup and execution of our experiment. Next, we provide detailed insights into our data analysis, which prepares for the discussion of our results. Finally, we conclude with a brief summary and directions for future research.

9.2 Background

Our research background focuses on synthesizing the findings of published systematic literature reviews on AR learning and empirical studies about marker-less AR from the last decade. In order to identify relevant related work, we analyzed the most cited literature reviews on AR for education.

Most extant studies do not focus on using AR in real-life environments, but investigate its use for supporting a narrow and well-defined task in a controlled setting. Hence, it is not surprising that virtually all existing studies focus on the application of marker-based AR, which is easy to implement in a controlled laboratory setting, and that only few studies have investigated the use of marker-less AR so far (Bacca et al., 2014; Billinghurst et al., 2015; Palmarini et al., 2018; Zhou et al, 2008). Moreover, marker-less AR is one key aspect discussed for implementing hybrid tracking for ubiquitous AR (Billinghurst et al, 2015; Palmarini et al., 2018; Van Krevelen and Poelman, 2010; Zhou et al, 2008).

What most studies have also in common is that they emphasize the need of further research on the features, use, advantages, and limitations of AR in educational settings (Akçayır and Akçayır, 2017; Bacca et al., 2014; Chen et al., 2016; Diegmann et al., 2015). Reported advantages of AR in educational settings include learning gains, higher motivation, facilitated interaction, better collaboration, lower cost, better user experiences, just-in-time information, enabling of situated learning and student-centered approaches, increase of students' attention, enjoyment, exploration, increased capacity for innovation, creation of positive attitudes, more awareness, anticipation, and authenticity (Akçayır and Akçayır, 2017; Azuma, 1997; Bacca et al., 2014; Billinghurst et al., 2015; Chen et al., 2016; Diegmann et al., 2015; Dunleavy et al., 2009; Palmarini et al., 2018; Radu, 2014; Sommerauer and Müller, 2014, Van Krevelen and Poelman, 2010; Zhou et al., 2008). In contrast, repeatedly reported limitations of AR in education include the observation that AR apps are mostly designed for only one specific knowledge field (Bacca et al., 2014), that teachers cannot create new learning content (Akçayır and Akçayır, 2017; Bacca et al., 2014; Diegmann et al., 2015; Radu, 2014), that there are difficulties maintaining superimposed information, that learners pay too much attention to the virtual information, that evaluation focused on short-term instead of long-term learning (Bacca et al., 2014), and that AR can be perceived as an intrusive technology (Bacca et al., 2014; Palmarini et al., 2018; Zhou et al., 2008). Still, most studies found positive evidence for the effectiveness of AR in education, for example, in the form of enhanced learning performance, higher learning motivation, improved perceived enjoyment, decreased cost, as well as adding creating positive attitudes towards education and fostering students' commitment (Akçayır and Akçayır, 2017; Azuma, 1997; Bacca et al., 2014; Billinghurst et al., 2015; Chen et al., 2016; Diegmann et al., 2015; Dunleavy et al., 2009; Palmarini et al., 2018; Radu, 2014; Sommerauer and Müller, 2014; Van Krevelen and Poelman, 2010; Zhou et al., 2008).

In Bacca et al.'s review of AR for education, the authors report about 19 studies that use marker-based AR, 4 studies with marker-less AR, and 7 studies covering location-based AR (Bacca et al., 2014). They discuss challenges around the improvement of recognition algorithms (e.g., for human forms) in the process of achieving more immersive and not intrusive AR learning experiences. Furthermore, they recommend vocational educational training (VET) classes as target groups for future studies.

In their literature survey of AR, Billinghurst et al. [2015] additionally focus on technology for user activity tracking considering input and interaction. They provided first design guidelines and interface patterns for AR development tools, starting with considering physical objects, virtual content and interaction metaphors and their connection. Additionally, they suggest future research directions as user tracking, user interaction, AR displays, and social acceptance of AR.

Dunleavy & Dede provide insights in AR teaching and learning, focusing on AR utilizing mobile, context-aware technologies (e.g. smartphones, tablets), thus enabling AR users interacting with digital information which is embedded within physical environments and in both, formal and informal learning environments (2009). They additionally investigate affordances and limitations for AR related to teaching, learning and instructional design and see AR as primarily aligned with situated and constructivist learning theory, stating, that AR positions learners within a real-world physical and social context while guiding, scaffolding and facilitating participatory and metacognitive learning processes (e.g. authentic inquiry, active observation, peer coaching, reciprocal teaching). Since AR legitimate users in peripheral participation with multiple modes of representation, they distinguish between location-aware and vision-based AR. In this context, AR has some limitations regarding student cognitive overload and managing level of complexity, which is a key instructional issue. Therefore, they recommend to decrease cognitive load by creating a simplified experience structure initially and increasing complexity as the experience progresses, thus scaffolding each experience explicitly at every step to achieve the desired experience or learning.

When Radu states that the educational community remains unclear regarding the educational usefulness of AR and regarding contexts in which this technology is more effective than other educational mediums, he refers to 26 publications comparing student learning with AR vs. non-AR apps (Radu, 2014). Radu (2014) observed some negative consequences, such as attention tunneling, usability difficulties, ineffective classroom integration, and learner differences. His table of factors influencing learning in AR covers content representation, multiple representations that appear at appropriate time and space, learners are physically enacting educational concepts, attention is directed to relevant content, learners are interacting with 3D simulations, interaction and collaboration are natural.

Still, the benefits of AR in educational environments and the value of AR apps applied in educational environments has not yet been investigated in its entirety (Dalal and Trigg, 2005). The different directions of AR apps differ regarding their potential benefits. In their systematic literature review to synthesize a set of 25 publications, Diegmann et al. (2015) identified 14 different benefits clustered in six different groups. They considered dimensions like state of mind (e.g. increased motivation, increased attention, increased concentration, increased satisfaction), teaching concepts (e.g. student-centered learning, collaborative learning), presentation (e.g. increased details, information accessibility, interactivity), learning type (e.g. improved learning curve, increased creativity), content understanding (e.g. improved development of spatial abilities, memory), and reduction of costs (Diegmann et al., 2015). They then mapped the benefits to five directions of AR in educational environments (discovery-based learning, objects modeling, AR books, skills training, AR gaming) and indicated that specific directions of AR apps are more likely to lead to certain benefits, such as increased motivation. Especially, they emphasize that future research is needed to investigate the causality between benefits and directions of AR.

In their review of AR in education from 2011 to 2016, Chen et al. focused on research which includes the uses, advantages, features, and effectiveness of AR in educational settings (2016). They recommended to undertake more studies considering the difference of cognitive process and psychological immersion between AR and reality settings, individual interaction, sense of identity, adaptive application in AR, AR classroom design and evaluation research, teacher's role model in AR educational setting, design and implementation of AR learning resources in K-12.

The literature review by Akçayır & Akçayır focuses on current advantages and challenges of AR education. Although AR promotes enhanced learning achievement, they experienced a discrepancy for AR in terms of cognitive load and/or cognitive overload, and AR ease of use vs. challenges for AR app usability (Akçayır and Akçayır, 2017). Since research studies report both, they advise AR developers to develop and consequently implement empirically proven design principles, focusing on AR use and educational outcomes, and AR apps designed for diverse populations (e.g. kids, students, lifelong learners). They emphasize the need to investigate students' satisfaction, motivation, interaction, and commitment, and provide insights from research and development comprising explanations of development processes and factors being considered in design.

Dunleavy, Dede, and Mitchell document in their review covering AR simulations for teaching and learning, how teachers and students describe and comprehend ways of participation in AR simulation, to aid or hinder teaching and learning (Dunleavy et al., 2009). By means of qualitative case studies across two middle schools they demonstrate that AR supports multi user environments and immersive collaborative simulation.

For professional education and training, Palmarini et al. focused on the state of the art of AR apps applied in maintenance (Palmarini et al., 2018). Based on 30 primary studies between 1997-2017, they unveil most relevant technical limitations for AR and propose results indicating a high fragmentation among hardware, software and AR solutions which lead to a high complexity for selecting and developing AR systems, thus identifying areas where AR technology still lacks maturity (e.g. marker-less AR).

Further limitations for AR which are still present today were depicted by Zhou et al. for tracking techniques, interaction techniques, user interfaces, and AR displays, especially for head mounted displays (HMD) (Zhou et al., 2008). Although the development of AR hardware became more sophisticated in the past decade, the major technical issues are not sufficiently dissolved and need to be overcome, like low sensitivity trigger to recognition (Akçayır and Akçayır, 2017).

9.3 Methodology

In our study we followed the advice from Dunleavy & Dede (i.e., decreasing cognitive load by creating a simplified experience structure) (2009), Diegmann et al. (i.e., causality between benefits of AR) (2015), Chen et al. (i.e., AR classroom design and evaluation research, design and implementation of AR learning resources) (Chen et al., 2016), Palmarini (i.e., use of marker-less AR) (Palmarini et al., 2018). In order to develop an AR app for both school and professional education (VET), we applied design principles from Billinghurst et al. (i.e., real physical objects/virtual elements to be displayed, linking interaction metaphor) (Billinghurst et al., 2015) and Sommerauer & Müller (i.e., design elements derived from learning theories) (2018).

For the evaluation of the effect of marker-less AR applied in a learning scenario we chose to design a controlled laboratory experiment to compare the support of AR with traditional, paper-based material inside a classroom. In this, we aimed to ensure that no or hardly any differences in information equivalence (Larkin and Simon, 1987) could affect the results of our study. Finally, our research design aimed to support and control exactly those research design elements which were the key subject of investigation.

With the experiment we examined the usability of AR, its effectiveness and the potential for teaching and learning. The evaluation covered measures for perceived usefulness, perceived learning and students' motivation as well as objective performance in terms of time to completion for the task and number of mistakes made in a recall and retention test administered as a post-test. In addition, we employed the Systems Usability Scale (SUS) (Lewis & Sauro, 2017) to evaluate the usability of the applied AR system.

In our app development, we considered design elements from Billinghurst et al., who proposed to focus on physical objects, virtual content, the interaction metaphor, and their connections (2015). Additionally, we applied the conceptual framework by Sommerauer & Müller (2014), which is inspired by Anderson's work on how learning can be enhanced using emerging technologies and applying learning theories (Anderson 2016). At the heart of this framework are one or more learning sequences, each consisting of one or more connected learning activities. At the center of a single learning activity stands the learning content. This content should be designed according to different learning theories, indicated by the different concentric layers surrounding the learning content. At the first layer, it is proposed to apply the 12 design principles of the cognitive theory of multimedia learning (CTML) (Mayer, 2009). In the second layer, design elements from mobile learning (e.g., Herrington et al., 2009) shall be considered for application design. Finally, it is proposed to implement design elements from game-based learning (e.g. leaderboard, mission) (Hirumi et al., 2010), simulations (e.g., storytelling, drama), experiential learning theory (e.g., diverging, assimilating) (Kolb, 2014), and situated learning (McLellan, 1996). Additionally, collaborative learning elements can be introduced at the learning stage, where multiple learning activities are combined into a learning sequence (Sommerauer and Müller, 2014).

We instantiated the above described conceptual framework by developing an AR learning app prototype. It supports the task of learning names related to physical objects used in a particular professional domain – in our case, the florist industry. More specifically, the app combines machine learning techniques for image recognition and machine translation to identify objects that are in the focus of the mobile phone camera in real-time and superimpose information such as the object's name in different languages onto the object. As a training application, the app can be used in any workplace environment and the trainee can select between exploration mode or quiz mode. In both, the user needs to focus the particular object using the device's camera (e.g. smartphone, tablet, any head-mounted device). Once the object is recognized, the app provides a selection of labels, comprising the three most likely names of the object using a percentage scale and colors. In quiz-mode, the app shows the most likely label and two randomly selected labels and the trainee has to pick the correct one. Figure 9.1 shows screenshots of the application and show the explore and quiz modes.

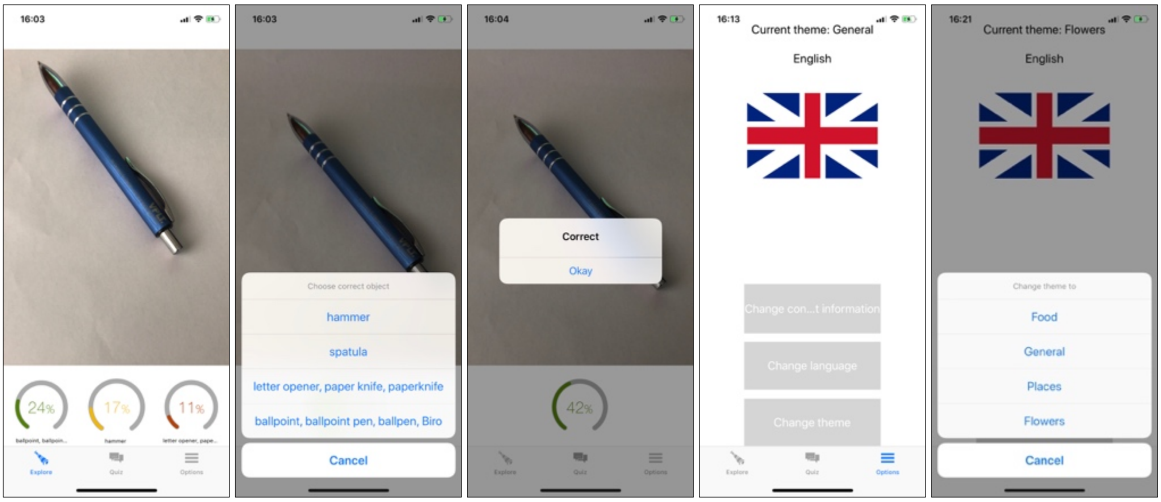


Figure 9.1: App in explore mode, quiz mode and selection of language and theme

The app design integrates design elements from CTML (i.e., the multimedia principle, the spatial contiguity principle, the temporal contiguity principle, and the signaling principle) with elements from the theory of mobile learning (i.e., users can use the app across space and time) and game-based elements. From a technical perspective, the app is based on Apple's ARKit framework[12] for implementing mobile AR experiences, Google's MobileNets model[13], a convolutional neural network for efficient image recognition on mobile phones, and the Google Translate API[14] for automated translation of texts into multiple languages.

As a foundation for our flower identification app we used the flowers dataset by Nilsback and Zisserman (2008) implemented as a selectable theme in our app. The flowers dataset consists of 8,189 images of flowers commonly occurring in the United Kingdom. The images are divided into 103 classes and each class consists of between 40 and 250 images. The images are scaled so that the smallest dimension is 500 pixels. The flowers are identified by different features describing different properties, e.g., color (HSV values of pixels), histogram of gradient orientations (HOG) (Dalal and Trigg, 2005), and distinctive image features (DIFT) (Lowe, 2004) on foreground region and foreground boundary. In prior studies the recognition accuracy was measured at 72.8 percent.

The instructional design for the learning situation applied in the experiment contained elements from cognitive and constructivist learning theories. While the learning content was prepared based on CTML principles, elements of constructivist theory were implemented in the learning activity, such as, task orientation, mobile learning, and situated learning, by sending learners on missions including storytelling.

The laboratory experiment was based on a sequential quantitative method research (Creswell, 1994) applying a static group design. The aim of the experiment was to identify differences in the application of AR vs. traditional learning. While the experiment group was supplied with mobile devices (iPhone 8+ and X) running the AR app, the control group received a traditional, paper-based tool (catalogue) to fulfil their task. Both groups received the same instructions and were required to fulfil the same task. At the end of the experiment both groups received a post-test questionnaire covering the same topics and questions. The questionnaire contained three sections. The first covered aspects for perceived usefulness, perceived learning, and students' motivation. The second part was a multiple-choice test asking for the names of five flowers shown as pictures. For each, participants could choose between three given names. The number of correctly identified flowers was used as an objective measure for learning performance. The third section of the questionnaire contained ten questions from the System Usability Scale (SUS), which was only available for the group using AR in the experiment.

9.4 Experimental Setup

The laboratory experiment followed a static group design comprising an experimental group and a control group. With this design we intended to investigate the differences in task and

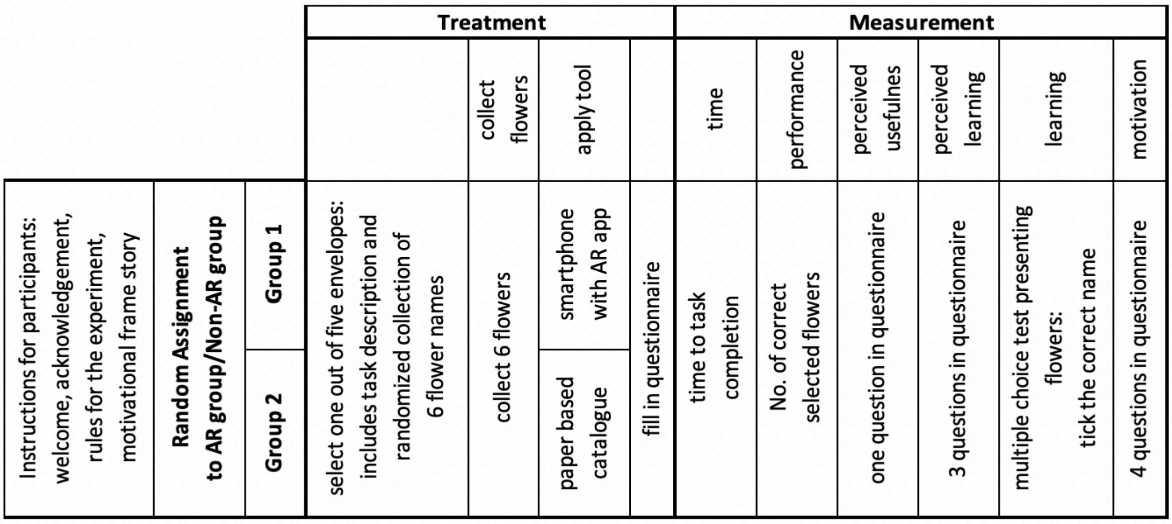
---

[12] https://developer.apple.com/arkit/

[13] https://github.com/tensorflow/models/tree/master/research/slim/nets/mobilenet

[14] https://github.com/tensorflow/models/tree/master/research/slim/nets/mobilenet

learning performance of two groups: one supported by an AR tool and one using traditional tools (i.e. a catalogue). Following similar studies (Akçayır and Akçayır, 2017; Azuma, 1997, Bacca et al., 2014; Billinghurst et al., 2015; Chen et al., 2016; Dunleavy et al., 2009; Radu, 2014; Sommerauer and Müller, 2014) and in line with our research questions, we used the item "time for task completion" as a measure for task performance and "No. of correctly identified flowers" from the questionnaire after the treatment as an indicator for learning performance. Figure 9.2 gives an overview of the randomized field experiment.



			Treatment				**Measurement**					
				collect flowers	apply tool		time	performance	perceived usefulnes	perceived learning	learning	motivation
Instructions for participants: welcome, acknowledgement, rules for the experiment, motivational frame story	**Random Assignment to AR group/Non-AR group**	**Group 1**	select one out of five envelopes: includes task description and randomized collection of 6 flower names	collect 6 flowers	smartphone with AR app	fill in questionnaire	time to task completion	No. of correct selected flowers	one question in questionnaire	3 questions in questionnaire	multiple choice test presenting flowers: tick the correct name	4 questions in questionnaire
		Group 2			paper based catalogue							

Figure 9.2: Overview of the randomized field experiment

We prepared two flower meadow, each consisting of 100 fake flowers composed of four different flower pictures per flower species and covering a selection of 25 different flower species from the flower dataset. The pictures were printed on paper and mounted on skewers. On the back side, the fake flowers were numbered according to an internal reference list to allow internal identification without the need for labels.

As a traditional tool for supporting participants in the experiment, we prepared a flowers catalogue covering exactly the 25 different flower species from the flower meadows. The flower pictures in the catalogue were different from those in the flower meadow and the catalogue was ordered alphabetically.

The questionnaire in the first section used a Likert scale containing five values from strongly disagree (1) to strongly agree (5) and covering eight questions:

- Perceived Usefulness:
  A. The AR app / catalogue was helpful to fulfil the task.
- Perceived Learning:
  B. With this activity I have learned something.

  C. I have learned about flowers.
  D. I can put together a bouquet on my own.
- Motivation: What do you think about the experiment and its setup?
  E. The introductory story was motivating.
  F. The task was simple and understandable.
  G. It was exciting to fulfill the task.
  H. The activity was entertaining.

Both rooms for the experiment where prepared in the same way. We set up the flower meadow with the fake flowers sticking in carton boxes and grouped by flower type. The carton boxes were placed on three tables in the center of the room. There was enough space to walk around the tables and to reach the flowers easily.

The main task for the participants was to collect six flowers from the meadow, which were named in form of a word-cloud on the instruction sheet in an envelope. We prepared five envelopes and the selection of the flower names for the word-cloud was done by a randomization process. Such, we used a webtool (www.randomizer.org) to collect 5 sets of 6 unique numbers per set within the range from 1 to 25. To arouse student attention and motivation, we narrated a story to send them on a mission, thus following design principles from game-based learning and simulation: "You fell in love with another person and have learned that you can break the ice between you and your crush with a smoothly arranged bouquet of flowers. Since you are absolutely unfamiliar with how to create a convincing flower bouquet, you ran a data analysis on your partner's Facebook account and received a list of preferred flowers presented in the word-cloud below". The mission to accomplish was formulated in the way, that "You know that love is like a little bird which flies away after some time and since you have just this one chance to score, give your best and collect the flowers as listed in the word cloud from the "self-service shop" as accurately and as fast as you can!".

While the AR group could use a prepared iPhone (we used four iPhone 8+ and one iPhone X) to complete their mission, the control group (non-AR group) was provided with the aforementioned flowers catalogue. As noted earlier, we used different pictures for the catalogue and the production of the fake flowers.

The experimental process was designed in a way that after listening to the initial instruction participants were assigned an envelope with further instructions, the story, the mission, and either an iPhone or a flowers catalogue. Then the researcher started a timer and the students needed to collect the flowers as fast as possible. Afterwards, they came back to the researcher who recorded the collected flower numbers and asked students to complete the questionnaire. Since the students received a participant number, this number was noted on the questionnaire for later analysis. Once the participants completed all tasks, the fake flowers were put back to the flower meadows and the room was prepared for the next group.

9.5 Implementation

We invited 71 students from a Masters course in Information Technology at a technical university in northern Europe to participate in the experiment, but only 44 attended. The students were already divided into working groups from their course and we assigned them to sessions with a maximum of ten students per session and a duration of approximately 15 minutes. Participating students received a voucher from the university's coffee shop as a reward right after the experiment.

The experiment started with a short introduction to welcome and thank the students for their participation. The participants were given some motivational instructions and were told to not chat with each other during the experiment or tell others about the experiment afterwards to not influence other students attending later. To split the group into the AR group (participants interacting with AR app during the experiment) and non-AR group (control group working with catalogue instead of AR app), students were told to choose between one of the two rooms by having equal numbered groups.

Participants could choose one of the five envelopes and when they started reading the instructions, a timer was set. After collecting the flowers, the students had to move to the research assistant and hand over their flower bouquet and all provided materials. To document the selected flowers and the required time to completion for the task, participants received a number to record their results for analysis. They then received the questionnaire to be answered on their own, marked with their participants number. After the students completed the questionnaire, they could leave the experiment.

Both experiment groups were treated in the same way, except of having different tools (AR app and paper catalogue) to fulfill the main task. There were no a priori time restrictions given, but students in the AR group were asked to terminate the collecting of flowers after 15 minutes.

9.6 Data Analysis

A participants' data record contained participant ID, group (AR, non-AR), gender (female, male), envelope number, IDs of the collected flowers, time to task completion, and the answers to the questions of the post-test questionnaire. In a first analysis, we assessed the number of correct flowers collected and the answers from the questionnaire. Overall, 18 female and 27 male students took part in the experiment, where 20 were assigned to the AR group and 25 the non-AR group.

In the AR group, 6 female and 14 male participants required from 510 to 1200 seconds to complete the given task (median 858.5 seconds, mean 864 seconds). They collected between 2 and 6 correct flowers from the given bouquet (median 4, mean 4.45). In terms of learning performance, the number of correct named flowers in their post-test questionnaire reached from 0 to 5 (median 2, mean 2.55).

In the non-AR group, 12 female and 13 male participants needed between 68 and 330 seconds to complete the task (median 171 seconds, mean 182.24 seconds). They collected 0 to 6 correct flowers from the given bouquet (median 5, mean 5.16) and the number of correctly named flowers in the post-test questionnaire reached from 0 to 5 (median 2, mean 1.96). Between the two groups there was no difference in the distribution of envelopes, which was tested by performing a Kolmogorov Smirnov test.

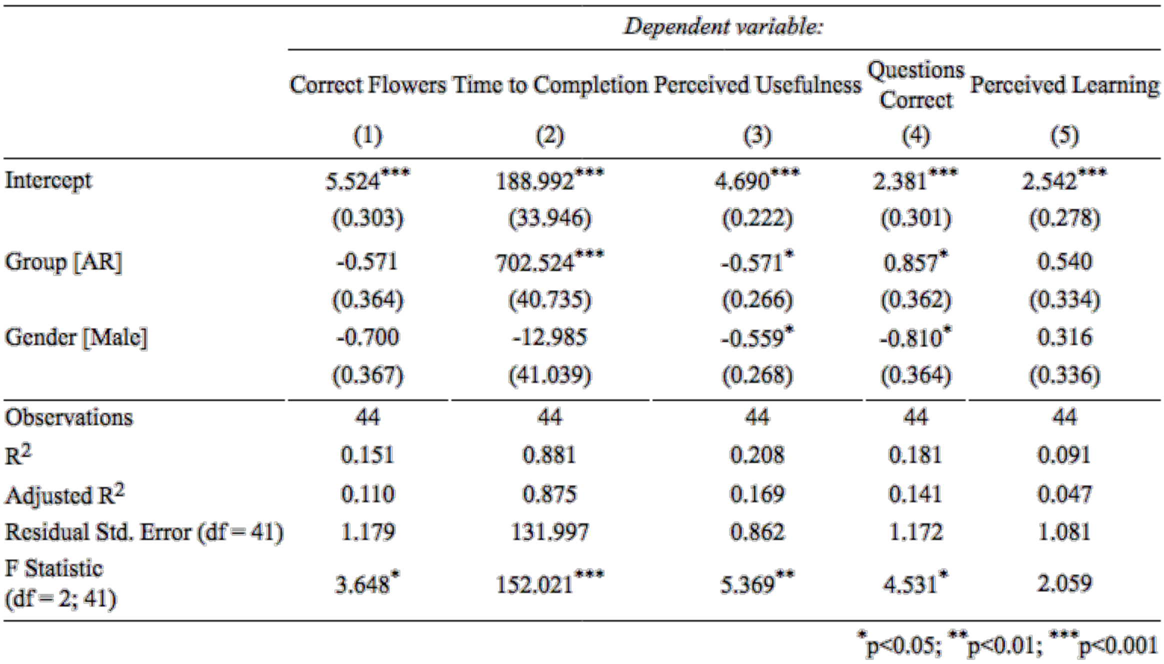
Table 9.1: Correlation matrix

		Group	Gender	Envelope	t2compl	NoCorrFl	QuizRes	QA	QB	QC	QD	QE	QF	QG	QH
Group	Pearson Correlation	1	-,183	,113	-,927**	,287	-,229	,382**	-,241	-,352*	-,079	-,026	,099	-,014	,117
	Sig. (2-tailed)		,230	,458	,000	,056	,130	,010	,111	,018	,607	,866	,519	,930	,444
Gender	Pearson Correlation	-,183	1	-,220	,139	-,325*	-,283	-,363*	,154	,149	,150	,000	-,011	-,115	-,158
	Sig. (2-tailed)	,230		,146	,364	,029	,059	,014	,312	,329	,325	1,000	,942	,452	,300
Envelope	Pearson Correlation	,113	-,220	1	-,053	,122	,099	-,001	,065	-,098	,058	,231	-,091	,241	,235
	Sig. (2-tailed)	,458	,146		,732	,425	,516	,994	,670	,522	,703	,126	,552	,110	,121
t2compl	Pearson Correlation	-,927**	,139	-,053	1	-,306*	,295*	-,326*	,242	,324*	,120	,052	-,034	-,009	-,172
	Sig. (2-tailed)	,000	,364	,732		,041	,049	,029	,109	,030	,433	,733	,824	,951	,259
NoCorrFl	Pearson Correlation	,287	-,325*	,122	-,306*	1	,107	,176	-,227	-,281	-,053	-,037	-,013	,232	,201
	Sig. (2-tailed)	,056	,029	,425	,041		,485	,249	,133	,061	,730	,812	,933	,126	,186
QuizRes	Pearson Correlation	-,229	-,283	,099	,295*	,107	1	,166	-,248	-,119	,021	,230	,196	,183	,199
	Sig. (2-tailed)	,130	,059	,516	,049	,485		,276	,101	,435	,890	,128	,197	,228	,189
QA	Pearson Correlation	,382**	-,363*	-,001	-,326*	,176	,166	1	-,188	-,069	-,290	,257	,354*	,306*	,328*
	Sig. (2-tailed)	,010	,014	,994	,029	,249	,276		,216	,652	,053	,089	,017	,041	,028
QB	Pearson Correlation	-,241	,154	,065	,242	-,227	-,248	-,188	1	,709**	,428**	,109	,034	,241	,148
	Sig. (2-tailed)	,111	,312	,670	,109	,133	,101	,216		,000	,003	,476	,824	,111	,331
QC	Pearson Correlation	-,352*	,149	-,098	,324*	-,281	-,119	-,069	,709**	1	,322*	,148	-,031	,274	,238
	Sig. (2-tailed)	,018	,329	,522	,030	,061	,435	,652	,000		,031	,331	,842	,068	,116
QD	Pearson Correlation	-,079	,150	,058	,120	-,053	,021	-,290	,428**	,322*	1	,018	-,173	,109	,060
	Sig. (2-tailed)	,607	,325	,703	,433	,730	,890	,053	,003	,031		,908	,256	,475	,696
QE	Pearson Correlation	-,026	,000	,231	,052	-,037	,230	,257	,109	,148	,018	1	,175	,506**	,545**
	Sig. (2-tailed)	,866	1,000	,126	,733	,812	,128	,089	,476	,331	,908		,250	,000	,000
QF	Pearson Correlation	,099	-,011	-,091	-,034	-,013	,196	,354*	,034	-,031	-,173	,175	1	,221	,127
	Sig. (2-tailed)	,519	,942	,552	,824	,933	,197	,017	,824	,842	,256	,250		,144	,407
QG	Pearson Correlation	-,014	-,115	,241	-,009	,232	,183	,306*	,241	,274	,109	,506**	,221	1	,674**
	Sig. (2-tailed)	,930	,452	,110	,951	,126	,228	,041	,111	,068	,475	,000	,144		,000
QH	Pearson Correlation	,117	-,158	,235	-,172	,201	,199	,328*	,148	,238	,060	,545**	,127	,674**	1
	Sig. (2-tailed)	,444	,300	,121	,259	,186	,189	,028	,331	,116	,696	,000	,407	,000	

Next, we ran an exploratory correlation analysis between all relevant pairs of variables in our dataset (Table 9.1). We found statistically significant correlations between group assignment and time to completion (mean of AR/non-AR: 864sec/182sec), perceived usefulness (QA) (mean of AR/non-AR: 3.65/4.40), and one of the questions related to perceived learning (QC) (mean of AR/non-AR: 3.4/2.8). Interestingly, we also found a significant correlation between gender and the number of correctly collected flowers (mean of female/male: 5.33/4.52, $p<0.01$), and perceived usefulness (mean of female/male: 4.50/3.77, $p<0.01$).

As our pseudo random assignment of students to groups did not produce an even distribution of males and females between the AR and non-AR group and because the correlation analysis indicated that gender is correlated with some of our dependent variables of interest, we decided to use regression models to test the main hypotheses of our experiment, namely that AR has a positive impact on (perceived) task performance and (perceived) learning performance. The advantage of a regression model over t-tests or ANOVA is in the ability to model the influence of multiple independent variables (in our case group and gender) on one dependent variable. Table 9.2 summarizes the results of this analysis.

Table 9.2: Regression results



	Dependent variable:				
	Correct Flowers	Time to Completion	Perceived Usefulness	Questions Correct	Perceived Learning
	(1)	(2)	(3)	(4)	(5)
Intercept	5.524^{***}	188.992^{***}	4.690^{***}	2.381^{***}	2.542^{***}
	(0.303)	(33.946)	(0.222)	(0.301)	(0.278)
Group [AR]	-0.571	702.524^{***}	-0.571^{*}	0.857^{*}	0.540
	(0.364)	(40.735)	(0.266)	(0.362)	(0.334)
Gender [Male]	-0.700	-12.985	-0.559^{*}	-0.810^{*}	0.316
	(0.367)	(41.039)	(0.268)	(0.364)	(0.336)
Observations	44	44	44	44	44
R^2	0.151	0.881	0.208	0.181	0.091
Adjusted R^2	0.110	0.875	0.169	0.141	0.047
Residual Std. Error (df = 41)	1.179	131.997	0.862	1.172	1.081
F Statistic (df = 2; 41)	3.648^{*}	152.021^{***}	5.369^{**}	4.531^{*}	2.059

$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

According to the regression results, participants in the AR group did not perform significantly better in terms of correctly identifying flowers than participants in the paper catalogue group. With regard to time needed to complete the task, participants in AR group even performed significantly worse than participants in the paper catalogue group. Hence, we did not find any empirical evidence that the AR app increased participants' objective task performance in terms of task accuracy and task time. Consistent with this finding, participants in the AR group evaluated the perceived usefulness of their tool (i.e. the AR app) significantly worse than participants in the non-AR group working with the paper catalogue.

However, when looking at objective learning performance, measured by the number of questions answered correctly in the post-test questionnaire, we found that participants in the AR group performed significantly better. This finding provides empirical support for the effectiveness of AR as a tool to enhance students' objective learning performance. With regard to perceived learning (measured by the average scores of questions B-D), we did not find a significant difference between the groups.

9.7 Discussion

In our experiment students achieved an observably better learning performance when using the AR flower identification app instead of a comparable paper catalogue, a result that is similar to prior research results comparing AR-based training to traditional paper-based training methods (Bacca et al., 2014; Dunleavy and Dede, 2014; Radu, 2014). Therefore, and to answer RQ2, we conclude that AR can support students' learning performance. However, it may also be that the learning performance for the AR group was influenced by their longer task completion times, thus students were more engaged with the learning content and more motivated (Akçayır and Akçayır, 2017). This can either be seen as a potential confounding factor which has to be controlled for in future studies (e.g. by predefining the available time for conducting a task), or as a positive side effect of using AR for teaching and learning (Bacca et al., 2014; Billinghurst et al., 2015; Chen et al., 2016; Sommerauer and Müller, 2014; Van Krevelen and Poelman, 2010). One could argue that when using AR students voluntarily spend more time with the learning materials, as compared to using traditional paper-based tools.

Considering participants' behavior during the experimental task, we noted that students in the AR group acted differently than those in the non-AR group. While participants in the AR group needed to investigate the flowers sequentially (because the app can only identify one object at a time) and thus examined nearly all flowers from the meadow, participants in the non-AR group selected a flower's name from the task description, searched for the name in the catalogue, and then located the flower by scanning the flower meadow with their eyes and matching the picture from the catalogue with the pictures on the meadow. On the one hand, this resulted in much shorter task times, as the human eye can focus on multiple objects at the same time (or at least can change focus much more quickly than AR technology), in comparison to the participants in the AR group who additionally had to perform the task of hand-eye coordination when using the app. On the other hand, when filling out the post-test questionnaire students realized that they had not inspected all flowers from the meadow and catalogue in sufficient detail in order to answer the questions correctly (the flowers students had to name in the post-test were different from those they had to collect).

A further observation related to the above point was that as participants in the AR group were forced by the app's functionality to look at each flower and since the app showed the three most likely names for identifying a flower and the elated confidence levels, students required more attempts to select the correct flower. We are convinced that this was a main driver behind the longer time needed to complete the task. Additionally, students from the AR group confirmed that it is more fun to look at the flowers with the app instead of just learning from a book.

It is remarkable that while the perceived learning of the AR group is not significantly higher compared to the non-AR group, their objective learning performance was significantly higher. The better objective learning performance may be explained by the different ways participants approached the task in the two groups. While students in the non-AR group focused on finding the flower picture for the given flower name and selecting a similar flower from the meadow, students in the AR group pointed their smartphone upon every single flower in the meadow to see its name. A single flower was represented multiple times in the flower meadow and students from the AR group visualized a particular flower more often. This finding corresponds to results from other studies, where AR is more effective than using traditional media (Akçayır and Akçayır, 2017; Bacca et al., 2014; Chen et al., 2016; Diegmann et al., 2015; Radu, 2014; Sommerauer and Müller, 2014).

Since the paper catalogue prepared for the experiment was ordered alphabetically and only contained few pages covering the presented 25 flowers, students in the non-AR group had an advantage when matching flower names between the task description and catalogue. This could be a major limitation in our study in regard to the results for participants task performance times compared with participants from the AR group. Using a flower identification book with hundreds of pages ordered by species instead of alphabetically would have been more realistic for our comparison and would probably have led to different results, at least in terms of task completion times. However, this observation indicates that the prepared catalogue was designed to support task completion.

Our app is technically able to identify up to 60 pictures per second, comparing it with several thousands of pictures from the database. Thus, the setup of the experiment with only a handful of flowers did not challenge the full potential of the app, which is a further limitation in terms of system performance in comparison of traditional tools with AR based tools. Nonetheless, with our study we could contribute to the discussion about improvement of AR recognition and marker-less AR [4]. For future research and practical application, the AR app can be utilized in any other learning environment just by exchanging the underlying image recognition machine learning model. This represents a cost-efficient alternative to integrate AR into classroom trainings (Radu, 2014).

9.8 Conclusion

With the app development and its application in the experiment we could answer our RQ 1 and demonstrate how marker-less AR can be implemented for education in a real-world environment. Thus, we followed recommendations for further research in the directions of implementing AR in real-life settings (Chen et al., 2016) and applying image-based tracking (Billinghurst et al., 2015) and marker-less AR (Bacca et al., 2014; Palmarini et al., 2018; Zhou et al., 2008) for ubiquitous learning (Akçayır and Akçayır, 2017). Moreover, with our study we investigated how marker-less AR affects task and learning performance in a mundane setting, for example in our simulation of a florist's job. Our results showed that from a learning aspect, students using the AR app performed better when it comes to recalling the learning content, similar to prior studies (Akçayır and Akçayır, 2017; Chen et al., 2016; Diegmann et al., 2015; Sommerauer and Müller, 2014). Although students in both groups achieved the same level of accuracy in fulfilling the given task, those students in the AR group needed more time. Since the experimental setup unintentionally supported the control group in faster task completion time, which points towards the finding that tasks processed with AR need to be designed differently.

Relying on the predefined dataset and machine learning model from Nilsback and Zisserman (2008) was an efficient decision and guaranteed a consistent recognition rate for each object in the experiment. However, participants had some troubles with finding the correct focus for the fake flowers because of reflections, shadows and different illumination caused by the changing daylight which is also mentioned in prior studies and therefore a limitation which should be investigated in future research (Akçayır and Akçayır, 2017; Dunleavy et al., 2009; Palmarini et al., 2018).

Students from the non-AR group benefited from the reduced catalogue to accomplish their mission. Since the AR app is able to recognize up to 60 pictures in a second from a dataset containing 8,189 pictures, the comparison of both tools in the experiment and for the given task, to search and identify a flower by its given name, was not really fair. However, with our study we demonstrate a content application of AR in association with its benefits and directions, particularly its scalability in a mundane situation.

Conducting an experiment just with students is not always satisfying. However, in our larger research program this was only a first test to demonstrate the use of the marker-less AR app and to collect and analyze first empirical data to investigate its effectiveness. In fact, we are beyond this now and are currently testing the app with a target group of low-threshold skilled employees.

From the aspect of using marker-less AR in educational settings we have ascertained that the recognition sometimes lacks due to optical influences, which is still a common issue for AR applications (Akçayır and Akçayır, 2017; Bacca et al., 2014; Billinghurst et al., 2015; Dunleavy and Dede, 2014; Palmarini et al., 2018; Sommerauer and Müller, 2014; Van Krevelen and Poelman, 2010; Zhou et al., 2008). Hence, future technological development should focus on recognition algorithms and the preparation of large and validated datasets in order to support the implementation of marker-less AR in education and in various real-life situations. Furthermore, the application of object detection instead of image recognition inside AR applications provides potential for new findings about how full 3D support for such AR apps assists learning and a better understanding. First results from our continuing research already confirm that object detection facilitates the recognition of a series of objects in one single viewpoint.

Fact sheet P.5

Title	“Collaboration in Augmented Reality Supported Workplace Training.”
Authors	**Peter Sommerauer**, Oliver Müller & Leonard Maxim
Publication type	Conference paper
Publication outlet	
Year	
Status	Manuscript submitted for publication
Full citation	Sommerauer, P., Müller, O. & Maxim L. (n.d.), Collaboration in augmented Reality Supported Workplace Training. *(Manuscript submitted for publication)*

10 Collaboration in Augmented Reality Supported Workplace Training

Abstract

Augmented Reality (AR) is widely used in various training and learning settings, like schools, universities, and workplaces. The effects of individual AR learning apps on learning performance in formal learning environments have been examined in detail in various studies. However, up to today only few empirical studies have investigated AR's potential for supporting learning at workplaces and in a collaborative setting. In this study we target this research gap by using an AR application to support collaborative learning in technically oriented workplace trainings. We conducted an observational field study with 67 professionals from the event technology industry and the results of our qualitative and quantitative data analysis suggest that integrating collaborative elements into AR trainings at workplaces has the potential to enable training experiences which can hardly be simulated with traditional media and which are perceived as beneficial for motivation and learning performance.

Keywords: Augmented Reality, collaboration, workplace training

10.1 Introduction

One of the most significant advantages of Augmented Reality (AR) is to enable immersive learning experiences by connecting digital information and physical objects in the real environment (Dunleavy et al., 2009). It has been argued that collaborative AR learning experiences especially focus on perception and performance; communication, interaction and collaboration; and the development and expansion of critical thinking and problem-solving skills (Bacca et al, 2014; Dünser et al., 2008; Dunleavy et al., 2009; Swan and Gabbard, 2005). For example, AR has the ability to support collaboration in face-to-face and remote settings by sharing a common space and having multiple people view, discuss, and interact with 3D models simultaneously (Billinghurst et al., 2015; Dunleavy et al., 2009; Van Krevelen and Poelman, 2010).

AR applications can be designed to address special educational needs of students in VET institutions and support expert learning (Meyer et al, 2014). The positive impact of computer supported collaborative learning (CSCL), has been shown in many previous studies (Dillenbourg, 2002). To create a compelling and effective design for AR learning applications, the integration of students, teachers, education technology experts, and software developers into a collaborative creation process is beneficial (Bacca et al., 2015). To measure learning success in such collaborative AR environments, process measures and subjective measures have been proposed as meaningful metrics for experiments and other empirical studies (Dunleavy et al., 2009).

However, the majority of existing studies investigate AR in learning and training environments with students (Akçayır and Akçayır, 2017; Bacca et al, 2014; Radu, 2012; Radu, 2014) and focus solely on usability or student motivation. In the context of vocational training, studies that have investigated AR-based learning were mainly interested in the development and application of tools and in aspects related to user interfaces and hardware. These studies also provided insights about how AR supports learning, in particular by providing multiple means of presentation, expression and engagement.

Measurement and evaluation of learning in personalized learning experiences, e.g., considering the acquisition of skills, competencies and critical thinking, is essential for the evolution of the digitization of education (Adams Becker et al., 2017). While most studies use metrics addressing perceived usefulness, learning, satisfaction, and dimensions of task performance, e.g. number of fulfilled tasks or correct answers (Jetter et al., 2018; Sommerauer and Müller, 2018), they nonetheless often lack a solid grounding in learning theory (Sommerauer and Müller, 2018).

With our study, we aim at gaining knowledge about designing AR applications for learning in a collaborative vocational training setting and thus contribute to the existing research gap regarding collaborative AR. Hence, we follow Billinghurst et al., who argued that collaborative AR supports collaboration on real world tasks and it is a particularly promising area for future AR user studies (2015). The central aim of our field study is to investigate the implementation of collaborative AR in a workplace training environment and to evaluate its application with domain experts. More specifically, we try to answer the following research questions: Are collaborative AR learning applications beneficial for workplace training? What is the added value of collaborative AR in workplace trainings? How can collaboration be integrated to support learning in AR trainings at workplaces?

The reminder of this study covers the introduction of the related background, followed by the description of the applied research methodology, and the development of the AR application. Subsequently, the test setup, the implementation, and the findings are depicted. In the last section, we discuss our findings and limitations and highlight some aspects for future research.

10.2 Background

Many studies compared learning with AR and non-AR applications. They suggest that AR can have a positive impact on learning in terms of, for example, increased content understanding, learning of spatial structures, language associations, long-term memory retention, improved collaboration, and motivation (Radu, 2012; Radu, 2014). However, aspects like attention tunneling, usability difficulties, ineffective classroom integration and learner differences have been identified as potential negative impacts of applying AR for teaching and learning (Radu, 2012; Radu, 2014). Hence, an effective integration of AR into teaching and learning implies the ability to create learning experiences that are aligned with general classroom pedagogy and curriculum (Radu, 2014).

AR demonstrated its potential not only in schools, but also at workplaces for technically and process-oriented hands-on training. In this context, is has been argued that future AR applications should focus on supporting ubiquitous, informal, and collaborative learning (Akçayır and Akçayır, 2017). Collaborative learning is widely understood as a situation in which two or more people learn or attempt to learn something together (Dillenbourg, 2002), either in pairs, small groups, classes, or communities. Mostly, the learning session follows a predefined course, covering the study of course material or performing learning activities such as problem solving. The collaborative aspect in such learning refers furthermore to the kind of interaction in such constructed learning situations, e.g. either introduced face-to-face or mediated by computer-based systems, and includes aspects of synchronous or asynchronous collaboration across time and space (Dillenbourg, 2002).

Interaction and collaboration are natural (Dillenbourg, 2002) and AR experiences potentially cause improvements in group collaboration (Billinghurst et al., 1996; Billinghurst et al., 1997; Szalavári et al., 1998), e.g. in mobile learning environments, using shared displays (Radu, 2014) or face-to-face collaboration in the same location to interact with shared AR content (Billinghurst et al., 1996; Billinghurst et al., 1997; Szalavári et al., 1998). In this context, new concepts like Tangible AR (TAR) using Tangible User Interface (TUI) were introduced (Billinghurst et al., 2008). Furthermore, location-based AR on mobile devices enables trainees to immerse themselves in the learning process and increase their collaboration skills (Bacca et al., 2015).

Various training situations have been identified where a user requires collaboration on a real-world task, e.g. in public management, crisis situations, urban planning, or to support remote maintenance in various industries (Billinghurst et al, 2015; Peyton, 1998; Radu, 2014). Moreover, AR supports remote and co-located activities in unique ways that would otherwise be almost impossible. Hence, AR has the potential to seamlessly integrate multiple users with display devices in multiple contexts, enhancing telepresence (Billinghurst et al., 2008).

In a direct comparison, AR performs better in collaborative maintenance sessions than traditional phone assistance (Havard et al., 2015). Measures used in collaborative AR experiments were performance time, game scores, object counts, and performance quality (Billinghurst et al, 2015). Therefore, the development and evaluation of collaborative AR interfaces constitutes a particularly promising area for further research in this direction (Billinghurst et al, 2015). Also, the applied data collection methods to evaluate the effect of AR for interaction and collaboration in training situations as in focus groups or conversational analysis needs to be further explored in future research studies (Bacca et al, 2014).

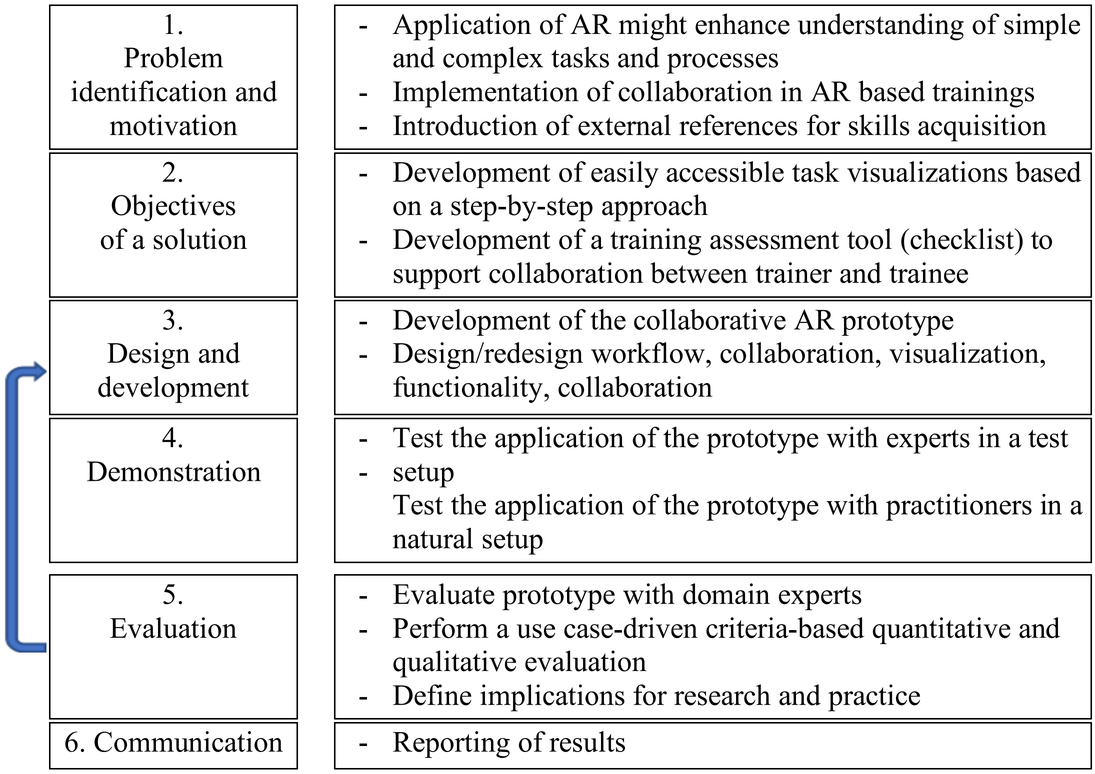
It is worth noting that previous studies also reported constraints for collaborative AR training experiences in comparison to non-technological settings. Examples include slower task performance and usability issues, which might explain identified disadvantages in collaboration (Radu, 2014). For example, collaboration might invite trainees to compete in a training session, which can lead to rushing and skipping over critical training steps (Dunleavy et al., 2009). Additionally, collaborative AR applications require a stable network connectivity to enable collaboration on virtual content, monitoring students' activities, and controlling the learning experience (Radu, 2014; Van Krevelen and Poelman, 2010).

10.3 Methodology

In our research, we followed Nunamaker et al.'s multi-methodological approach to IS research (Nunamaker et al., 1991) and the therefrom derived DAGS framework from Adams and Courtney (2004), which both focus on the integration of design science, systems development, and action research. Using design science and systems development methods, we built an AR prototype that embodies different learning theories, represents a proof-of-concept, and allows us to collect empirical data from the field. Subsequently, in the spirit of action research, we used the prototype to intervene into a real-world training setting and evaluate the usefulness, usability, and learning support of the prototype through quantitative (survey) and qualitative methods (participant feedback and observation).

For the design and development of the prototype we applied the design science research (DSR) process model following Peffers et al. (2006), illustrated in Figure 10.1. We developed a collaborative AR-based prototype for smartphones and tablets and filled it with content from an existing competence requirements catalogue, developed in collaboration with training experts from industry.

We collaborated with domain experts from the event technology industry in all stages of the development and evaluation process. The app followed a predefined storyline and considered the training requirements derived from a catalogue covering competence requirements, which should be implemented in the training session. Furthermore, the design of the app was based on design elements for AR learning app design suggested by Sommerauer and Müller (2018).



1. Problem identification and motivation	- Application of AR might enhance understanding of simple and complex tasks and processes - Implementation of collaboration in AR based trainings - Introduction of external references for skills acquisition
2. Objectives of a solution	- Development of easily accessible task visualizations based on a step-by-step approach - Development of a training assessment tool (checklist) to support collaboration between trainer and trainee
3. Design and development	- Development of the collaborative AR prototype - Design/redesign workflow, collaboration, visualization, functionality, collaboration
4. Demonstration	- Test the application of the prototype with experts in a test - setup Test the application of the prototype with practitioners in a natural setup
5. Evaluation	- Evaluate prototype with domain experts - Perform a use case-driven criteria-based quantitative and qualitative evaluation - Define implications for research and practice
6. Communication	- Reporting of results

Figure 10.1: Applied DSR process

Figure 10.2 shows the conceptual setup. For logging the user data, we implemented so-called experience-statements recorded on a learning record store (LRS), applying the xAPI (Kevan et al., 2016). For this purpose, we used Learning Locker® and shaped the xAPI statements to log a user's data to fit the requirements of the performance measurement catalogue.

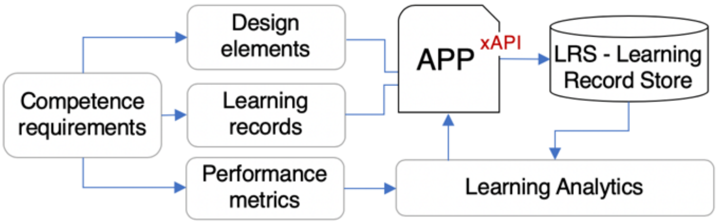


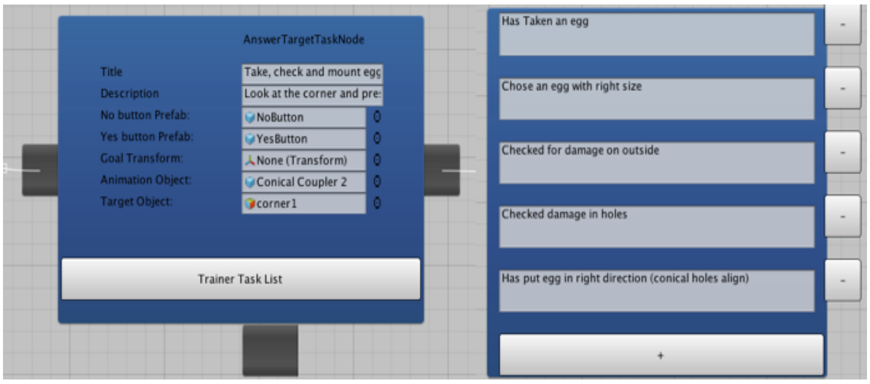
Figure 10.2: Conceptual setup for app development

Aspects derived from the competence requirements catalogue (Table 10.1) determine the applied design elements for the app design (Sommerauer and Müller, 2018) and the learning records implemented in the app and define the investigated performance metrics Sommerauer and Müller, 2018), e.g. task performance. The app in practice sends user data to the LRS to provide it for further processing, i.e. for learning analytics and to prepare a feedback for users (e.g. on a dashboard).

Table 10.1: Checklist items related to competences

Competence reference		Checklist item description
10 Fit up and rig performance equipment	10.02 Inspect the technical performance equipment visually for damage	Checked for damage on outside
		Checked damage in holes
		Checked pivot for damage
		Checked if pivot fits
		Checked if split pen closes properly
	10.03 Choose the right mounting accessories	Has taken an egg
		Chose an egg with right size
		Took the right pivot
		Chose correct split pen
	10.04 Choose the right mounting methods	Wobbled or hammered (if needed)
	10.05 Mount and rig technical performance equipment according to instructions and/or plans	Has put egg in right direction (conical holes align)
		Placed the pivot in right direction (conical hole)
	10.08 Secure technical performance equipment and accessories	Put split pen in hole of pivot
	10.10 Take action if something goes wrong	Disposed damaged pivot (if needed)

The truss connection application was developed using the Unity3D game engine (Helgason et al., 2004) accompanied by the built-in Vuforia Augmented Reality and HLAPI (The Multiplayer High Level API) frameworks. In order to quickly adapt the prototypes to the project needs, a node editor tool for generating final applications was implemented. This allows the developer to create, edit, and connect nodes (Figure 10.3) - each node represents one task that the trainee has to perform (e.g. "Take, check and mount egg on corner 1") which is further subdivided into subtasks (e.g. "Has taken an egg", "Chose an egg with right size", etc).



Title: Take, check and mount egg
Description: Look at the corner and pre
No button Prefab: NoButton
Yes button Prefab: YesButton
Goal Transform: None (Transform)
Animation Object: Conical Coupler 2
Target Object: corner1

Has Taken an egg
Chose an egg with right size
Checked for damage on outside
Checked damage in holes
Has put egg in right direction (conical holes align)

Figure 10.3: Generic node editor tool

10.4 Test Setup

Our setup of the field study was built upon the theory for action research (Adams and Courtney, 2004). To simulate a realistic workplace environment, we built a 6m x 4.5m black-box room to simulate a setting similar to a theater, re-enacting the scenery of a stage background with reduced lighting. The black box was equipped with a SD square heavy steel truss element which was fixed on the floor, a table with tools and components required for truss connection and electric power supply with a busbar. The truss had a square profile and thus four corners.

The setup of the training session followed industrial training instructions for connecting a truss and covered the identification of the items and tools used in the activity and the preparation of a truss element to prepare the connection of a further truss element. First, a so-called egg with a conical drilled hole needs to be put correctly into the hole of the longitudinal member at the first edge. Then, a pin needs to be mounted in the correct direction to fix the egg. For this activity, a hammer is used to ensure a strong connection. The final action is to secure the pin with a spigot. All four corners of the truss need to be prepared this way.

In our simulation, we designed the app following a learning process based on Peyton's four learning steps: demonstration, deconstruction, comprehension, and execution (Peyton, 1998). Trainees need to start the app and point the camera of their device towards the truss. At the first corner, the app superimposes a 3D model of the egg to demonstrate how to install the egg into the hole and the trainee is requested to follow the instruction. To get to the next step, the trainee needs to push a button and the app shows with 3D animations how to correctly put the pin into the holes of the truss. Once the trainee is ready for the next task, the app displays a 3D animation demonstrating how the spigot should be placed correctly. During the session, the trainer is invited to give verbal feedback and to intervene, if corrections are necessary. At the end of each step and the activities on the first corner, the trainer is requested by the system to send feedback based on a predefined checklist and starts the training on the second corner. The information provided by the trainer's checklist is stored at the LRS and contain aspects for evaluating a trainee's task performance linked to the checklist item descriptions (Table 10.1).

At the second corner, the app provides a single 3D animation combining all three steps of the activity in one animation. Again, the trainee is invited to replicate the given visual instructions and the trainer checks afterwards, whether the tasks were fulfilled correctly. Again, the trainer is allowed to give verbal instructions and also to reject the evaluation; in this case, the trainee needs to start again with the activities for the current corner.

For the remaining two corners of the truss, the trainee receives the instructions to assemble the corner on his/her own. Since we want to measure the task performance in these two steps, the app requires pressing a button after successfully assembling corner three and four. Thus, the trainees' task performance, i.e. time to completion, can be measured. The training session itself ends with feedback given by the trainer and the app shows the participant's performance analysis via a dashboard, based on the analysis of the recorded users' data from the LRS.

To measure a participant's overall performance, we focused on three aspects:

- Number of correctly and incorrectly fulfilled tasks
- Completion time for corners 3 and 4
- Time taken for the whole training scenario (Corners 1-4)

Furthermore, we prepared a questionnaire to receive a participant's feedback based on closed and open questions. The questionnaire included questions addressing a participant's impression of the system's usefulness (perceived usefulness), perceived learning, and motivation. The answers were given according to a five-level Likert scale (strongly disagree, disagree, neutral, agree, totally agree). Perceived usefulness is defined as "the degree to which a person believes that using a particular system would enhance his or her job performance" (Dunleavy and Dede, 2014, p. 320). Perceived learning describes the degree to which a student observes to obtain knowledge in a particular learning situation (Radu, 2014). Motivation is considered to be intent and engagement as action; both terms are often used interchangeably (Fredricks et al., 2011; Reschly and Christenson, 2012). Since our study is of exploratory nature, we used the term motivation as a measure for a learner's interest and engagement in a particular learning activity. In that sense, we address a learner's self-perception on this dimension. We formulated the questions according to these definitions:

- Perceived Usefulness: Q1: The App was helpful to fulfill the task.
- Perceived Learning: Q2: With this activity I have learned something.
  Q3: I have learned about truss connection.
  Q4: I can connect trusses correctly and safely.
- Motivation: Q5: The introductory story was motivating.
  Q6: The task was simple and understandable.
  Q7: It was exciting to experiment with the app.
  Q8: The activity was entertaining.

Finally, we also included questions regarding the system usability scale (SUS) (Bangor et al., 2008).

10.5 Implementation

We conducted the field study at a trade fair for the event technology industry. Visitors of the fair were mostly trained people and experts from the field of event technology, which were asked by research assistants outside the black box room to participate in our field study and were then directed into the black box. Inside the black box, the trainer took over and started with the training session.

At the beginning, participants were instructed along a storyline: the rigging crew needs help for setting up the stage and is looking for outstanding performers in quick-and-safe-rigging. Thus, the participants first need to successfully complete the training for connecting a truss and to test their skills supported by the new AR training app. The training covered the following activities:

- In step 1 and 2, prepare a truss following the instructions;
- Correctly identify damaged parts;
- Follow the advices of the assessor;
- Fulfill the tasks in step 3 and 4 on your own;
- Complete a questionnaire afterwards;

Then the trainer handed over an iPad where the training app was already started and connected with the trainer's app. The app provided a unique session code for each training session which was used as a session identifier and to link sessions and questionnaires. Participants were allowed to ask any questions and the trainer was instructed to support the activities in step 1 and 2 (preparing corner 1 and 2 of the truss). Figure 10.4 shows a participant in action.

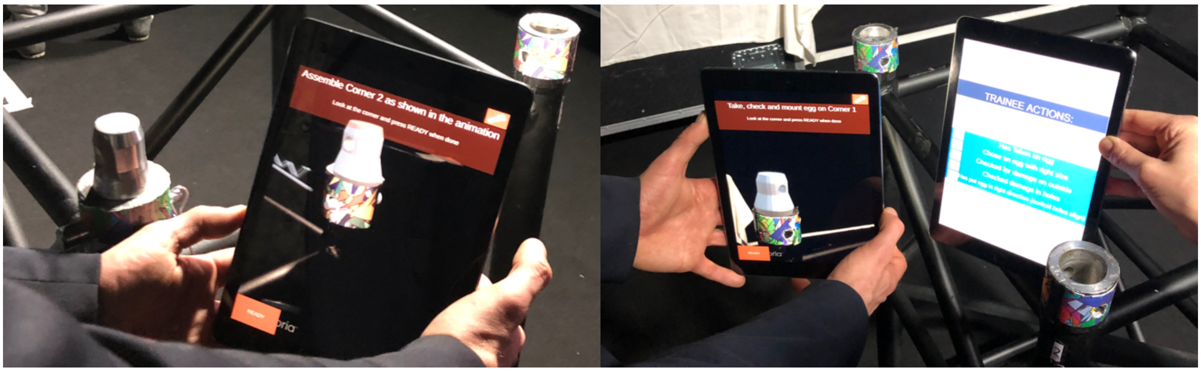


Figure 10.4: Collaborative AR in the field study

10.6 Data Analysis

We analyzed the collected data in two ways. The first part covered a manual examination of the participants' qualitative feedback provided in the questionnaire as described in chapter 10.4. In the second part, we asked participants closed questions regarding perceived usefulness (Q10: "What do you think are the strengths, benefits and Q11 added values of using the app in such a training scenario?") and perceived learning (Q12: "Introducing AR in trainings, do you think that people learn more and/or faster and/or with higher motivation?"). Finally, we asked for further use cases (Q13: "Which use cases do you think are applicable for implementing AR in workplace training?"). For the data analysis of the second part we applied the innovative synthesis method and a systematic analysis method to support this process, following Whittmore & Knafl (2005).

In a first step, we collected the answers from the questionnaires and assigned them according to the layers of the design framework proposed by (Sommerauer and Müller, 2018). These are the content layer, mobile layer, motivational layer, and situated/collaborative layer. Hence, we aim to provide the findings from this study for future development, i.e. to contribute to the research field of AR application design requirements. Furthermore, we counted the given answers to emphasize and quantify the users' perceptions. The analysis shows that most of the participants' answers were given in terms of motivational aspects (51 answers), followed by content-oriented answers (35 answers), answers reflecting situated/collaborative learning (27 answers), and only a few referred to mobile aspects (3 answers). Table 10.2 shows an extract of the collected qualitative feedback.

We continued by analyzing the participants feedback concerning strengths, added value, constraints and use cases, and categorized them. In summary, the most cited positive aspects for introducing collaborative AR in workplace trainings were visualization (22), independence (15), efficiency (13), language independence (10), process-orientation (9), realism (8), collaboration (6), motivation (5), and mobile and generic aspects (3). Applicable use cases suggested by participants are trainings introducing more complex tasks, safety training, training situations where you have no/low access to tools, and trainings in flexible environments.

The second step covered the statistical analysis of the quantitative answers of the questionnaire in combination with the user data logged by the system. First, we evaluated the results for Q1, perceived usefulness (pu) with N=57, min=1, max=5, mean=4. For perceived learning (pl) we aggregated the answers of questions Q2-Q4 (N=6, min=1, max=5, mean=3.7) and for motivation/engagement (m) we aggregated the answers of questions Q5-Q8 (N=61, min=1, max=5, mean=4.20). In addition, we calculated the SUS for each participant to learn how participants perceived the usability of the AR app. The mean value in our evaluation was 72.80 (minimum 25, maximum 100), which is comparable with SUS values of good products (Bangor et al., 2008).

Furthermore, we analyzed the user data collected by the AR app and investigated metrics like task performance regarding the overall training process (d0-6), and especially for the completion of the tasks in corner 3 (d4-5) and 4 (d5-6) of the truss preparation activity. Figure 10.5 shows a visual representation of the data. We conducted a paired t-test to compare the time to completion for corner 3 and 4. There was a significant difference in the scores for time to completion of corner 3 (M=39.03, SD=20.95) and corner 4 (M=23.37, SD=15.44), with t(66)=5.707, $p<0.001$). These results suggest that participants were faster at corner 4 than corner 3. We furthermore identified a weak positive correlation (r=0.268, n=67, p=0.029) based on a Pearson correlation analysis which suggests that participants who needed less time to completion for corner 3 were also faster at corner 4.

Table 10.2: Categorized qualitative feedback

Layer	Aspects	Quotation
Content	language	Imparts without language barrier what needs to be done. Language and nationalities independent.
	descriptive	The visual representation is very helpful because it is easy to understand what needs to be done.
	self-explaining	Trainer may need to correct only minor issues and it is self-explanatory.
	understanding	The visual, very clear presentation / instruction makes it easy for everyone to understand how to proceed.
	complexity	Learning about complex tasks. Ability to combine a series of steps.
	clear	Intuitive operation and clearly defined activities.
	interactive	Step by step instruction on the object.
	multimedia	Higher memorability through multisensory learning.
Mobile	independence	Learning …, time and place independent.
		Can be used for several people on a construction site.
		For trainees and interns as an exercise in the storage.
Motivation / Engagement	simple	simple handling; simple to learn;
	quick	You quickly learn how to handle the traverse.
	entertaining	It's quick and entertaining. Hands-on approach.
	costs	No need to travel. Cost efficient.
	safety	Training with no danger.
	pace	You can train multiple students on their own tempo.
	fun	Have fun, enjoy the work.
Situated Learning, Collaboration	collaborative	You learn together and make no mistakes.
	complexity	Learning about complex tasks. Ability to combine a series of steps.
	realistic	Realistic, simple and descriptive training.

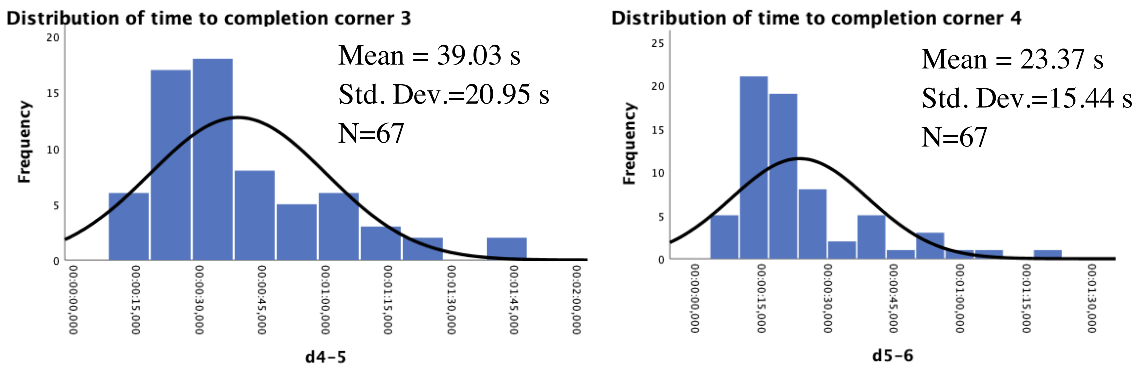


Figure 10.5: Histogram time to completion corner 3, 4

For the quantitative evaluation of the questionnaires we focused on perceived usefulness and perceived learning and the questions whether participants learned more, faster, and with higher motivation. We performed a Pearson correlation analysis between perceived usefulness (pu) and the other variables.

The first investigation addresses the interdependence between perceived usefulness (pu) and the SUS (uscore). The results show a significant and strong correlation (r=0.617, n=57, p=0.000) between pu (mean=4.18) and uscore (mean=72.80). This suggests that participants who valued the app as helpful assessed its usability on a higher level, or participants who uprated the system's usability perceived the app as helpful. In terms of perceived usefulness and motivation (m, mean=4.20) in the training session, we found a medium positive correlation between those variables (r=0.486, n=52, p=0.000). This indicates that participants who found the app helpful to fulfill the task were also more motivated. Figure 10.6 shows the corresponding scatterplots.

Investigating the correlation between perceived usefulness and perceived learning (pl, mean=3.69), we found a medium positive correlation (r=0.405, n=51, p=0.003). This suggest that participants who found the app helpful to fulfill the task also perceived that they have learned more. Alternatively, participants who perceived that they have learned more found the app more helpful.

In addition, we rated the answers for Q3 ("… people learn more / faster / with higher motivation with AR in trainings") from our qualitative feedback section of the questionnaire and awarded 2 points for a clear and positive answer (yes), 1 point for a positive answer with reservations, 0 points for an unbiased answer, -1 point for a negative answer with restrictions and -2 points for a clear and negative answer (no). In this way we calculated and defined a variable as "learnvalue" for the whole participant group with N=59 and mean=1.02.

Examining the value interpreted from participants' qualitative feedback in regards to learning, the computed Pearson correlation coefficient showed a medium positive correlation between perceived usefulness and the examined learn-value (r=0.402, n=57, p=0.002). Participants who agreed that the app is useful also valued the effectiveness of AR in trainings with higher approval. This is shown in the following Figure 10.6.

Since motivation was an aspect addressed in Q3 of the questionnaire too, we also tested if a correlation between the variables (m and learnvalue) is verifiable. The calculated Pearson coefficient shows a strong positive correlation (r=0.633, n=57, p=0.000) which indicates that participants with higher motivation felt more confident about AR's effectivity. Again, the scatterplot in Figure 10.6 summarizes the results.

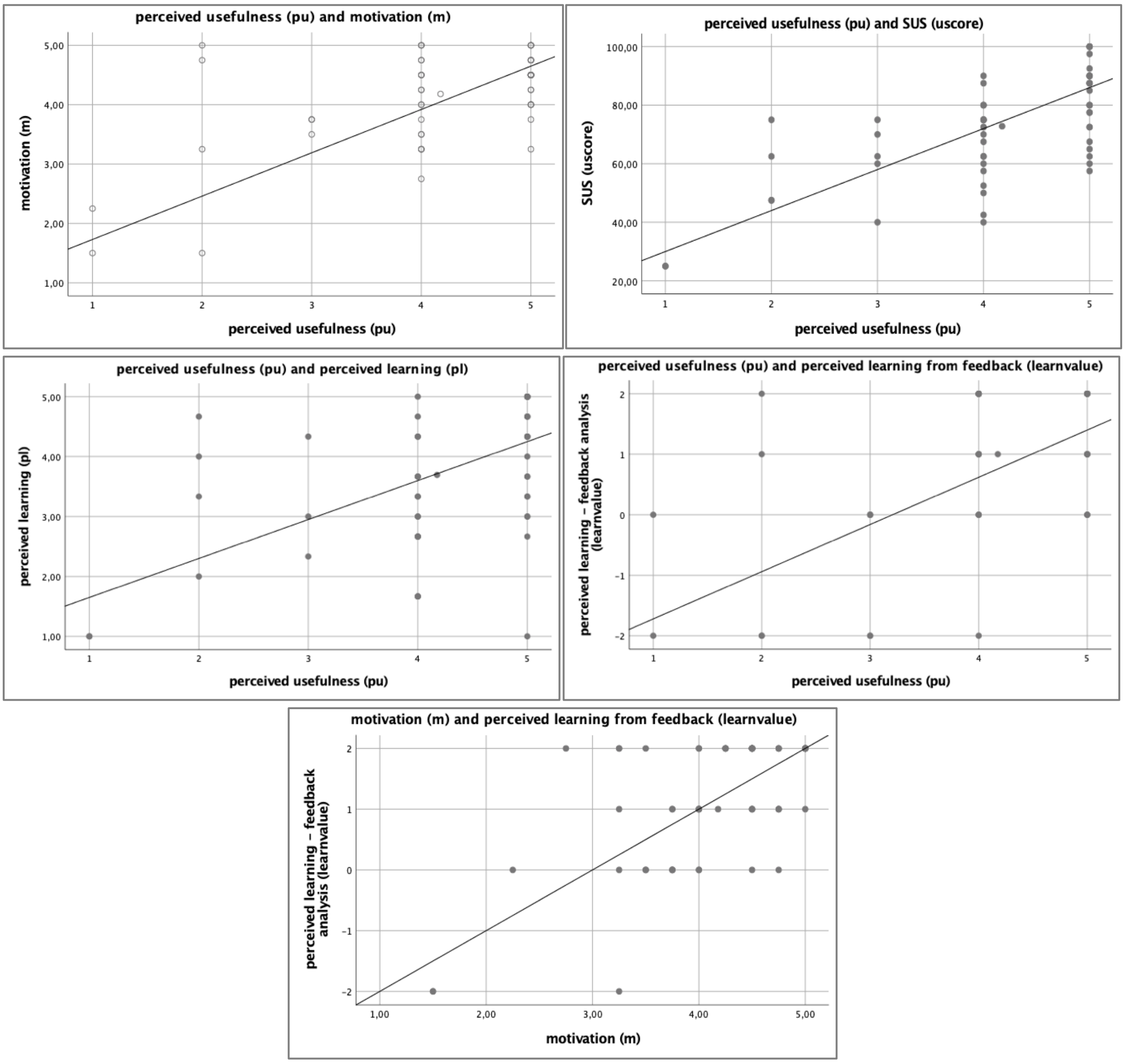


Figure 10.6: Scatterplots from the quantitative evaluation of the questionnaire

10.7 Discussion and Conclusion

Our findings from the field study provide considerable insights for the implementation of collaborative AR in a workplace training environment for answering the research questions. Especially in terms of the received feedback from subject matter experts and the evaluation of the targeted variables, i.e. usability, perceived usefulness, perceived motivation, and perceived learning, collaborative AR was largely approved by the participants. We successfully implemented a process-oriented set of instructions in our AR app development, which was aligned with the requirements derived from a standardized training curriculum from industry, and introduced features for collaboration in the given setup. That way, we were able to demonstrate the application of collaborative AR in a workplace training and provide details for the AR design framework in return. Since trainers and trainees benefitted from the motivational, collaborative and realistic training setting and appreciated aspects derived from the design, e.g. the interactive, intuitive and safe application, we identified independency, efficiency, and process-orientation as added value of AR in workplace trainings.

However, with the results of the field study we could confirm Billinghurst stating that process and subjective measures may be more important than quantitative outcome measures in collaborative AR experiments (Bacca et al, 2014, p. 198). Mainly in terms of collaboration and the applied step-by-step approach in the training, experts mentioned that collaborative AR opens up new opportunities to structure trainings individually, e.g. to first let trainees interact with the environment to overcome a trainee's inhibitions and in a second step, discuss training aspects in detail. Some experts argued in the direction that particularly in the first training sequence the trainer should add fundamental details to supply the trainee with important information. That way collaboration in trainings supports individual learning paths.

We noticed from our own observation and the participants' feedback that the task performance was slowed due to the AR application, similar to other studies (Palmarini et al., 2018; Radu, 2014) and mostly because of usability issues. Especially in the first consecution of the training, participants required to get used to the system, i.e. pointing the camera in a good angle to the trigger image and starting the AR visualization.

Since we focused our research on receiving qualitative and expert feedback, we utilized age and gender issues to a lesser extent, or considered the educational background of the individual participants. In terms of gender we could identify that 11 female and 56 male participants took part in our field study. This distribution reflects the gender situation of the industry (higher number of male employees) and therefore the strong male dominance in the professional field of event technology. However, we could not find any meaningful evidence for gender related differences for our study which could be caused by the small group of female participants in our field study. On the other hand, this finding could serve as representation for the equal skills between men and women in the particular industry.

We did not record the training situations to identify any differences in the individual training sessions. Since we worked with different trainers during the field study, this situation could have possibly influenced the results. Although we instructed the involved staff according to the structured process for the field study, personal aspects could have influenced participants' motivation and behavior in the training session (e.g. sympathy, level of details explained). However, an analysis towards differences in the collected data considering session dates and times did not show any effect.

Many participants noted that the task was too simple and referred to more complex tasks which would be interesting to investigate the application of collaborative AR at workplaces. Thus, the task simplicity could have had an influence on a participant's motivation and the perceived aspects like learning and usefulness in our study. However, we prepared the app in a way that makes it possible to map more complex tasks, which we intend to implement in our further research.

In contrast to this work, very few user studies report on collaborative AR for workplace training and almost none that examined communication process measures. Since communication is a key aspect in collaborative environments, future research should investigate in how active collaboration can be explored and how collaborative AR applications can be designed and implemented to support communication processes and their measuring. One of our next steps will be to evaluate the AR app in contrast to traditional trainings in a control-group design. We are currently working on an AR training simulation for a series of workplace trainings where the truss app is one training scenario and which refers to a definite skills set, thus leads to a qualification for stage technicians based on industry standards.

References

Adams, L.A. & Courtney, J.F. (2004). Achieving relevance in IS research via the DAGS framework. In: *Proc. of the 37th Ann. Hawaii Int. Conf. System Sciences*, 1-10.

Adams Becker, S., Cummins, M., Davis, A., Freeman, A., Hall Giesinger, C., & Ananthanarayanan, V. (2017). *NMC Horizon Report: 2017 Higher Education Edition*. Austin, Texas: The New Media Consortium.

Akçayır, M., & Akçayır, G. (2017). Advantages and challenges associated with augmented reality for education: A systematic review of the literature. *Educational Research Review*, *20*, 1-11.

Albert, W., & Tullis, T. (2013). *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics*. Newnes: Boston, MA, USA.

Anderson, J.R. (1983). *The Architecture of Cognition*. Cambridge, MA: Harvard University Press.

Anderson, T. (2016). *Theories for learning with emerging technologies*. Emerging technologies in distance education.

Aottiwerch, N., & Kokaew, U. (2017, July). Design computer-assisted learning in an online Augmented Reality environment based on Shneiderman's eight Golden Rules. In *14th International Joint Conference on Computer Science and Software Engineering (JCSSE), 2017*, 1-5. IEEE. [27]

Ariso, J. M. (2017). *Augmented reality*. Berlin: De Gruyter.

Azuma, R. T. (1997). A survey of augmented reality. *Presence: Teleoperators & Virtual Environments*, *6*(4), 355-385.

Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., & MacIntyre, B. (2001). Recent advances in augmented reality. *IEEE computer graphics and applications*, 21(6), 34-47.

Bacca, J., Baldiris, S., Fabregat, R., & Graf, S.: Augmented reality trends in education: a systematic review of research and applications. In: *Journal of Educational Technology & Society. 17*(4), 133.

Bacca, J., Baldiris, S., Fabregat, R., & Graf, S. (2015). Mobile augmented reality in vocational education and training. *Procedia Computer Science*, *75*, 49-58.

Bangor, A., Kortum, P.T. & Miller, J.T.: An empirical evaluation of the system usability scale. In: *International Journal of Human–Computer Interaction. 24*(6), 574-594.

Becker, S. A., Brown, M., Dahlstrom, E., Davis, A., DePaul, K., Diaz, V., & Pomerantz, J. (2018). *NMC horizon report: 2018 higher education edition*. Louisville, CO: EDUCAUSE.

Becker, J., & Niehaves, B. (2007). Epistemological perspectives on is research: a framework for analysing and systematizing epistemological assumptions. *Information Systems Journal, 17* (2), 197–214.

Beckett, K. L., & Shaffer, D. W. (2005). Augmented by reality: The pedagogical praxis of urban planning as a pathway to ecological thinking. *Journal of Educational Computing Research, 33*(1), 31-52. [112]

Berg, S. A., & Chyung, S. Y. (2008). Factors that influence informal learning in the workplace. *Journal of workplace learning*, *20*(4), 229-244.

Bhattacherjee, A. (2012). *Social science research: Principles, methods, and practices*.

Billinghurst, M., Clark, A., & Lee, G. (2015). A survey of augmented reality. *Foundations and Trends® in Human–Computer Interaction*, *8*(2-3), 73-272.

Billinghurst, M., & Duenser, A. (2012). Augmented reality in the classroom. *Computer, 45*(7), 56-63.

Billinghurst, M., & Kato H..: Out and about—real world teleconferencing. In: *BT technology journal. 18*(1), 80–82.

Billinghurst, M., Kato, H., & Poupyrev, I. (2008). Tangible augmented reality. *ACM SIGGRAPH ASIA*, 7.

Billinghurst, M., Weghorst, S., & Furness T. (1996). Shared space: Collaborative augmented reality. In: *Proceedings of CVE' 96 Workshop*.

Billinghurst, M., Weghorst, S., & Furness T. (1997). Wearable computers for three dimensional cscw. In: *2012 16th International Symposium on Wearable Computers, IEEE Computer Society*. 39-39.

Bingham, C. B., & Davis, J. P. (2012). Learning sequences: their existence, effect, and evolution. *Academy of Management Journal, 55*(3), 611-641.

Brühl, R. (2017). *Wie Wissenschaft Wissen schafft: Wissenschaftstheorie und-ethik für die Sozial-und Wirtschaftswissenschaften*. UTB.

Burgess, R. G. (1984). *In the field: An introduction to field research*.

Burrell, G., & Morgan, G. (2017). *Sociological paradigms and organisational analysis: Elements of the sociology of corporate life*. Routledge.

Cabero Almenara, & J., Barroso J. (2016). The educational possibilities of Augmented Reality. In *New Approaches in Educational Research. 5*(1), 44-50.

Carlson, K. J., & Gagnon, D. J. (2016). Augmented reality integrated simulation education in health care. *Clinical simulation in nursing, 12*(4), 123-127. [8]

Chang, K.-E., Chang, C.-T., Hou, H.-T., Sung, Y.-T., Chao, H.-L., & Lee, C.-M. (2014). Development and behavioral pattern analysis of a mobile guide system with augmented reality for painting appreciation instruction in an art museum. *Computers & Education*, *71*, 185–197.

Chang, Y. H., & Jen-ch'iang, L. I. U. (2013). Applying an AR technique to enhance situated heritage learning in a ubiquitous learning environment. *TOJET: The Turkish Online Journal of Educational Technology, 12*(3). [123]

Chang, Y. H., & Lin, Y. K. (2014b). A kinect-based system for learning Chinese cultural festivals. In *International Conference on Machine Learning and Cybernetics (ICMLC), 2014* (Vol. 2), 462-468. IEEE. [106]

Chang, Y. L., Hou, H. T., Pan, C. Y., Sung, Y. T., & Chang, K. E. (2015). Apply an augmented reality in a mobile guidance to increase sense of place for heritage places. *Educational Technology & Society, 18*(2), 166-178. [51]

Chen, C. M., & Tsai, Y. N. (2012). Interactive augmented reality system for enhancing library instruction in elementary schools. *Computers & Education, 59*(2), 638-652. [20]

Chen, P., Liu, X., Cheng, W., & Huang, R. (2017). A review of using Augmented Reality in Education from 2011 to 2016. In *Innovations in smart learning*. 13-18. Springer, Singapore.

Chen, Y., Zhou, D., Wang, Y., & Yu, J. (2017). Application of Augmented Reality for Early Childhood English Teaching. In *International Symposium on Educational Technology (ISET), 2017*, 111-115. IEEE. [1]

Cheng, K.-H., & Tsai, C.-C. (2013). Affordances of augmented reality in science learning: Suggestions for future research. *Journal of Science Education and Technology, 22*(4), 449–462.

Christenson, S. L., Reschly, A. L., & Wylie, C. (Eds.). (2012). *Handbook of research on student engagement*. Springer Science & Business Media.

Creswell, J. W. (2003). Research Design: *Qualitative, Quantitative, and Mixed Methods Approaches* (2nd ed.), Thousand Oaks, CA: Sage Publications.

Creswel, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches*. Los Angeles: University of Nebraska–Lincoln.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159.

Colley, H., P. Hodkinson, & J. Malcolm. (2003). *Informality and Formality in Learning: a report for the Learning and Skills Research Centre*. London: LSRC.

Cooper H. (1998). *Synthesizing Research: A Guide for Literature Reviews*. 3rd edn. Sage Publications. Thousand Oaks, CA.

Da Hora Macedo, S., Fernandes, F. A., de Lima, J. V., & Biazus, M. C. V. (2012). Learning object to teach the interaction between two magnetics using Augmented Reality. *Journal of Educational & Instructional Studies in the World, 2*(4). [54]

Davis, F. D. (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology. In: *MIS quarterly*, 319-340.

Dede, C. (2009). Immersive interfaces for engagement and learning. *Science*, *323*(5910), 66–69.

De Freitas, S., Rebolledo-Mendez, G., Liarokapis, F., Magoulas, G., & Poulovassilis, A. (2010). Learning as immersive experiences: Using the four-dimensional framework for designing and evaluating immersive learning experiences in a virtual world. *British Journal of Educational Technology, 41*(1), 69-85. [173]

Del Bosque, L., Martinez, R., & Torres, J. L. (2015). Decreasing Failure in Programming Subject with Augmented Reality Tool. *Procedia Computer Science, 75*, 221-225. [6]

Dey, A., Billinghurst, M., Lindeman, R. W., & Swan, J. (2018). A systematic review of 10 years of augmented reality usability studies: 2005 to 2014. *Frontiers in Robotics and AI*, 5, 37.

Diegmann, P., Schmidt-Kraepelin, M., Van den Eynden, S., & Basten, D. (2015). Benefits of Augmented Reality in Educational Environments - A Systematic Literature Review. In: *Wirtschaftsinformatik, vol. 3*(6), 1542-1556

Dillenbourg, P. (2002). *Over-scripting CSCL: The risks of blending collaborative learning with instructional design*. Heerlen, Open Universiteit Nederland, 61-91.

Dillenbourg, P., Schneider, D., & Synteta, P. (2002). Virtual learning environments. In *3rd Hellenic Conference" Information & Communication Technologies in Education"*, Kastaniotis Editions, Greece, 3-18.

Dimitrov, D., & Rumrill, P. J. (2003). Pretest-posttest designs and measurement of change. *Work*, *20*, 159–165.

Drljevic, N., Wong, L. H., & Boticki, I. (2017). Where does my Augmented Reality Learning Experience (ARLE) belong? A student and teacher perspective to positioning ARLEs. *IEEE Transactions on Learning Technologies*.

Dunleavy, M., Dede, C., & Mitchell, R. (2009). Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning. *Journal of science Education and Technology*, *18*(1), 7-22.

Dunleavy, M., & Dede, C. (2014). Augmented reality teaching and learning. In J. Spector, M. Merrill, J. Elen, & M. Bishop (Eds.), *The Handbook of Research for Educational Communications and Technology*, New York: Springer, 735–745

Dünser, A., Grasset, R., Billinghurst, M.: A survey of evaluation techniques used in augmented reality studies. In *Human Interface Technology Lab New Zealand*, 5. (2008)

Dünser, A., Steinbügl, K., Kaufmann, H., & Glück, J. (2006). Virtual and augmented reality as spatial ability training tools. In *Proceedings of the ACM SIGCHI New Zealand Chapter's International Conference on Computer-Human Interaction*, 125–132.

Echeverría, A., Améstica, M., Gil, F., Nussbaum, M., Barrios, E., & Leclerc, S. (2012). Exploring different technological platforms for supporting co-located collaborative games in the classroom. *Computers in Human Behavior*, *28*(4), 1170–1177.

Elias, T. (2011). Learning analytics. In: *Learning*. 1-22.

European Commission (EC). (2001). *Communication: making a European area of lifelong learning a reality*. Retrieved August 13, 2019, from:http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2001:0678:FIN:EN:PDF

Fredricks, J., McColskey, W., Meli, J., Mordica, J., Montrosse, B., & Mooney, K. (2011). Measuring Student Engagement in Upper Elementary through High School: A Description of 21 Instruments. *Issues & Answers. REL 2011-No. 098*. Regional Educational Laboratory Southeast.

Fonseca, D., Martí, N., Redondo, E., Navarro, I., & Sánchez, A. (2014). Relationship between student profile, tool use, participation, and academic performance with the use of Augmented

Reality technology for visualized architecture models. *Computers in Human Behavior, 31*, 434–445.

Furió, D., GonzáLez-Gancedo, S., Juan, M. C., Seguí, I., & Rando, N. (2013). Evaluation of learning outcomes using an educational iPhone game vs. traditional game. *Computers & Education, 64*, 1-23. [52]

Furió, D., Juan, M. C., Seguí, I., & Vivó, R. (2015). Mobile learning vs. traditional classroom lessons: a comparative study. *Journal of Computer Assisted Learning, 31*(3), 189-201. [125]

Garrison, C., & Ehringhaus, M. (2007). *Formative and summative assessments in the classroom.*

Garzón, J., & Acevedo, J. (2019). A Meta-analysis of the impact of Augmented Reality on students' learning effectiveness. *Educational Research Review.*

Gopalan, V., Bakar, J. A. A., & Zulkifli, A. N. (2017). A brief review of augmented reality science learning. In *AIP Conference Proceedings* (Vol. 1891, No. 1, p. 020044). AIP Publishing. [17]

Gough, D., Oliver, S., & Thomas, J. (Eds.). (2017). *An introduction to systematic reviews.* Sage.

Greenfield, P. M. (2009). Technology and informal education: What is taught, what is learned. *Science*, *323*(5910), 69–71.

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS quarterly*, 337-355.

Harrison, G., & List, J. (2004). Field Experiments. *Journal of Economic Literature*, *42*(4), 1009–1055.

Havard, V., Baudry, D., Louis, A., & Mazari, B. (2015) Augmented reality maintenance demonstrator and associated modelling. In: *2015 IEEE Virtual Reality Conference. VR 2015 - Proceedings*, 329–330.

Helgason D., Francis N., & Ante J. (2004). Unity3D. In: *Copenhagen: Unity Technologies.*

Herrington, Jan & Oliver, Ron. (1995). *Critical characteristics of situated learning: Implications for the instructional design of multimedia.*

Herrington, A., Herrington, & Mantei, J. (2009). *Design principles for mobile learning.*

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004) Design Science in Information Systems Research, *MIS Quarterly (28:1)*, 75-105.

Hevner, A., & Chatteijee, S. (2010). *Design Research in Information Systems*. New York: Springer Publishing

Hirumi, A., Appelman, B., Rieber, L., & Van Eck, R. (2010). Preparing instructional designers for game-based learning: Part 2. *TechTrends, 54*(4), 19.

Huang, T. C., Chen, C. C., & Chou, Y. W. (2016). Animating eco-education: To see, feel, and discover in an augmented reality-based experiential learning environment. *Computers & Education, 96*, 72-82. [4]

Hulin, T., Schmirgel, V., Yechiam, E., Zimmermann, U. E., Preusche, C., & Pöhler, G. (2010). Evaluating exemplary training accelerators for programming-by-demonstration. In *RO-MAN, 2010 IEEE*, 440-445. IEEE. [172]

Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, MA: MIT Press.

Ibáñez, M. B., Maroto, D., Rueda, J. J. G., Leony, D., & Kloos, C. D. (2012). Architecture for Collaborative Learning Activities in Hybrid Learning Environments. *J. UCS, 18*(15), 2187-2202. [124]

Ibáñez, M. B., Di Serio, Á., Villarán, D., & Delgado Kloos, C. (2014). Experimenting with electromagnetism using augmented reality: Impact on flow student experience and educational effectiveness. *Computers & Education*, *71*, 1–13.

Iivari, J. (2015). Distinguishing and contrasting two strategies for design science research. *European Journal of Information Systems*, 24(1), 107-115.

Illeris, K. (Ed.). (2009). *Contemporary theories of learning: learning theorists... in their own words*. Routledge.

Jansen, H. (2010). The logic of qualitative survey research and its position in the field of social research methods. In *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* (Vol. 11, No. 2).

Jayaram, S., Connacher, H. I., & Lyons, K. W. (1997). Virtual assembly using virtual reality techniques. *Computer-aided design*, *29*(8), 575-584.

Jetter, J., Eimecke, J., & Rese, A. (2018). Augmented reality tools for industrial applications: What are potential key performance indicators and who benefits?. In: *Computers in Human Behavior. 87*, 18-33.

Johnson, D. (2010). Crossover experiments. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(5), 620–625.

Johnson, L., Adams, S., & Cummins, M. (2012). *The NMC Horizon Report*: *2012 Higher Education Edition*. Austin, Texas: The New Media Consortium

Johnson, L., Adams Becker, S., Estrada, V., & Freeman, A. (2015). *NMC Horizon Report*: *2015 Higher Education Edition*. Austin, Texas: The New Media Consortium.

Johnson, L., Adams Becker, S., Cummins, M., Estrada, V., Freeman, A., & Hall, C. (2016). *NMC Horizon Report: 2016 Higher Education Edition*. Austin, Texas: The New Media Consortium.

Kamarainen, A. M., Metcalf, S., Grotzer, T., Browne, A., Mazzuca, D., Tutwiler, M. S., & Dede, C. (2013). EcoMOBILE: Integrating augmented reality and probeware with environmental education field trips. *Computers & Education, 68*, 545-556. [7]

Kaufmann, H. (2003). Collaborative augmented reality in education. *Institute of Software Technology and Interactive Systems, Vienna University of Technology*.

Keebler, J. R., Wiltshire, T. J., Smith, D. C., Fiore, S. M., & Bedwell, J. S. (2014). *Shifting the paradigm of music instruction: implications of embodiment stemming from an augmented reality guitar learning system.* Frontiers in psychology, 5. [86]

Kevan, J. M., & Ryan, P. R. (2016). *Experience API: Flexible, decentralized and activity-centric data collection*. Technology, Knowledge and Learning, 21(1), 143-149.

Kiili, K. (2005). Digital game-based learning: Towards an experiential gaming model. *The Internet and higher education, 8*(1), 13-24.

King W.R. (2009) Knowledge Management and Organizational Learning. In: *King W. (eds) Knowledge Management and Organizational Learning. Annals of Information Systems*, vol 4. Springer, Boston, MA

Kolb, D. A. (2014). *Experiential learning: Experience as the source of learning and development*. FT press.

Koutromanos, G., & Styliaras, G. (2015, July). " The buildings speak about our city": A location based augmented reality game. In *6th International Conference on Information, Intelligence, Systems and Applications (IISA)*, 1-6. IEEE. [29]

Kuechler, W., & Vaishnavi, V. (2008). "On Theory Development in Design Science Research: Anatomy of a Research Project," *European Journal of Information Systems*

K. Kumpulainen, L. Krokfors, L. Lipponen, V. Tissari, J. Hilppö, & A. Rajala. (2009). *'Learning Bridges –Toward Participatory Learning Environments'.*

Larkin, J., & Simon, H. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, *11*(1), 65–100.

Lee, K. (2012). Augmented reality in education and training. *TechTrends*, *56*(2), 13-21.

Levy, Y., & Ellis, T. J. (2006). A systems approach to conduct an effective literature review in support of information systems research. *Informing Science*, 9.

Levy, Y., & Ellis, T. J. (2011). A guide for novice researchers on experimental and quasiexperimental studies in information systems research. *Interdisciplinary Journal of information, knowledge, and management*, *6*(1), 151-161.

Lewis, J. J. R., & Sauro, J. (2017). Revisiting the Factor Structure of the System Usability Scale. *Journal of Usability Studies, 12*(4).

Liu, T., Tan, T., & Chu, Y. (2009). Outdoor Natural Science Learning with an RFID-Supported Immersive Ubiquitous Learning Environment. *Educational Technology & Society*, *12*(4), 161–175.

Lundblad, T., Malmberg, C., Areskoug, M., & Jönsson, P. (2012). Simulating real-life problems in secondary science class: a socio-scientific issue carried through by an Augmented Reality simulation. *HUMAN IT 12.2*(2012), 1–41. [40]

Lune, H., & Berg, B. L. (2016). *Qualitative research methods for the social sciences*. Pearson Higher Ed.

Manuti, A., Pastore, S., Scardigno, A. F., Giancaspro, M. L., & Morciano, D. (2015). Formal and informal learning in the workplace: a research review. *International journal of training and development*, *19*(1), 1-17.

Margetis, G., Zabulis, X., Koutlemanis, P., Antona, M., & Stephanidis, C. (2013). Augmented interaction with physical books in an Ambient Intelligence learning environment. *Multimedia tools and applications, 67*(2), 473-495. [75]

Marshall, B., Cardon, P., Poddar, A., & Fontenot, R. (2013). Does sample size matter in qualitative research?: A review of qualitative interviews in IS research. *Journal of Computer Information Systems.*

Martin, A. J. (2007). Examining a multidimensional model of student motivation and engagement using a construct validation approach. *British Journal of Educational Psychology, 77*(2), 413-440.

Martín-Gutiérrez, J., Luís Saorín, J., Contero, M., Alcañiz, M., Pérez-López, D. C., & Ortega, M. (2010). Design and validation of an augmented book for spatial abilities development in engineering students. *Computers & Graphics*, *34*(1), 77–91.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist*, *32*(1), 1–19.

Mayer, R. E. (2005). Cognitive theory of multimedia learning. *The Cambridge handbook of multimedia learning, 41*, 31-48.

Mayer, R. E. (2009). *Multimedia Learning* (2nd ed.). Cambridge: Cambridge University Press.

McGivney, V. (2002). *Spreading the Word reaching out to new learners*. National Institute of Adult Continuing Education (NIACE), Leicester.

McLellan, H. (1996). *Evaluation in a situated learning environment, situated Learning perspectives*, Englewood Cliffs, NJ: Educational Technology Publications.

Meyer, A., Rose, D.D., & Gordon, D. (2014). *Universal Design for Learning, theory and practice*. CAST Professional Publishing, Wakefield, MA

Méheut, M., & Psillos, D. (2004). Teaching–learning sequences: aims and tools for science education research. *International Journal of Science Education, 26*(5), 515-535.

Mikalef, K., Giannakos, M. N., Chorianopoulos, K., & Jaccheri, L. (2013). Does informal learning benefit from interactivity? The effect of trial and error on knowledge acquisition during a museum visit. *International Journal of Mobile Learning and Organisation 11*, 7(2), 158-175. [157]

Mills, E., Chan, A., Wu, P., Vail, A., Guyatt, G., & Altman, D. (2009). Design, analysis, and presentation of crossover trials. *Trials*, *10*(27), 1–6.

Milgram, P., & Kishino, F. (1994). A Taxonomy of Mixed Reality Visual Displays. *IEICE Transactions on Information Systems, 77*(12).

Milgram, P., Takemura, H., Utsumi, A., & Kishino, F. (1995). Augmented reality: A class of displays on the reality-virtuality continuum. In *Telemanipulator and telepresence technologies* (Vol. 2351), 282-292. International Society for Optics and Photonics.

Munoz-Cristobal, J. A., Jorrin-Abellan, I. M., Asensio-Perez, J. I., Martinez-Mones, A., Prieto, L. P., & Dimitriadis, Y. (2015). Supporting teacher orchestration in ubiquitous learning environments: A study in primary education. *Learning Technologies, IEEE Transactions on Learning*, *8*(1), 83-97.

National Research Council. (2009). *Learning science in informal environments: People, places, and pursuits.* National Academies Press.

NMC. (2012). *NMC Horizon Report: 2012 Higher Education Edition*. Retrieved from http://www.nmc.org/publications/horizon-report-2012-higher-ed-edition

Nunamaker, J. F. Jr., Chen, M., & Purdin, T.D. (1991). Systems Development in Information Systems Research. In: *Journal of Management Information Systems.* 251-266.

Odeh, S., Abu Shanab, S., & Anabtawi, M. (2015). Augmented Reality Internet Labs versus its Traditional and Virtual Equivalence. *International Journal of Emerging Technologies in Learning, 10*(3). [41]

OECD. (n.d.). *Recognition of Non-formal and Informal Learning*. Retrieved January 12, 2014, from http://www.oecd.org/education/skills-beyond-school/recognitionofnon-formalandinformallearning-home.htm

Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford: Oxford University Press.

Palmarini, R., Erkoyuncu, J. A., Roy, R., & Torabmostaedi, H. (2018). A systematic review of augmented reality applications in maintenance. In: *Robotics and Computer-Integrated Manufacturing, vol. 49*, 215-228

Papagiannakis, G., Singh, G., & Magnenat-Thalmann, N. (2008). A survey of mobile and wireless technologies for augmented reality systems. *Computer Animation and Virtual Worlds*, *19*(1), 3-22.

Papert, S. (1980). *Mindstorms*. New York: Basic Books.

Parhizkar, B., Obeidy, W. K., Chowdhury, S. A., Gebril, Z. M., Ngan, M. N. A., & Lashkari, A. H. (2012). Android mobile augmented reality application based on different learning theories for primary school children. *In 2012 International Conference on Multimedia Computing and Systems*, 404-408. IEEE. [28]

Parhizkar, B., Islam, M., Lashkari, A., Abdullas-Al-Jubair, M., Gebril, Z., & Zarrabi, M. (2012b). Simulation of Investigating the Earth and Universe Using Interactive Mobile Augmented Reality Based On Smart Evaluation. *In Proceedings of the International MultiConference of Engineers and Computer Scientists, Vol. 1*, 656-659). [105]

Parsons, J., & Cole, L. (2005). What do the pictures mean? Guidelines for experimental evaluation of representation fidelity in diagrammatical conceptual modeling techniques. *Data & Knowledge Engineering*, *55*(3), 327–342.

Peffers, K., Tuunanen, T., Gengler, C. E., Rossi, M., Hui, W., Virtanen, V., & Bragge, J. (2006). The Design Science research process: a model for producing and presenting information systems research. *Proceedings of the First International Conference on Design Science Research in Information Systems and Technology* (DESRIST 2006). Claremont, 83–106.

Peyton, J.W.R. (1998). Teaching in theatre. In: *Teaching and learning in medical practice*. Manticore, 171-180.

Piaget, J. (1954). *The Construction of Reality in the Child*. New York: Basic Books.

Pinsonneault, A., & Kraemer, K. (1993). Survey research methodology in management information systems: an assessment. *Journal of management information systems*, *10*(2), 75-105.

Prensky, M. (2001). *Digital Game-Based Learning*. McGraw-Hill.

Radu, I. (2012). Why should my students use AR? A comparative review of the educational impacts of augmented-reality. In: *2012 IEEE International Symposium on Mixed and Augmented Reality* (ISMAR), 313-314.

Radu, I. (2014). Augmented reality in education: a meta-review and cross-media analysis. *Personal and Ubiquitous Computing*, *18*(6), 1533-1543.

Reschly, A. L., & Christenson, S. L. (2012). Jingle, jangle, and conceptual haziness: Evolution and future directions of the engagement construct. In *Handbook of research on student engagement*, 3-19. Springer, Boston, MA.

Richard, J. A. (2015). Understanding Theories of Learning. *International Journal of Multidisciplinary Research in Modular Education, 1*, 343-347.

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Sage Publications.

Santos, M. E. C., Chen, A., Terawaki, M., Yamamoto, G., Taketomi, T., Miyazaki, J., & Kato, H. (2013). Augmented reality X-Ray interaction in K-12 education: theory, student perception and teacher evaluation. In *IEEE 13th International Conference on Advanced Learning Technologies* (ICALT), 141-145. IEEE. [3]

Santos, M. E. C., Chen, A., Taketomi, T., Yamamoto, G., Miyazaki, J., & Kato, H. (2014). Augmented reality learning experiences: Survey of prototype design and evaluation. *IEEE Transactions on Learning Technologies, 7*(1), 38–56. doi:10.1109/TLT.2013.37 [19]

Santos, M. E. C., Taketomi, T., Yamamoto, G., Rodrigo, M. M. T., Sandor, C., & Kato, H. (2016). Augmented reality as multimedia: the case for situated vocabulary learning. *Research and Practice in Technology Enhanced Learning*, 11(1), 4. [45]

Screven, C. (1993). Museums and Informal Learning. *CMS Bulletin*, 1(1).

Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and Quasi-experimental Designs for Generalized Causal Inference*, Boston: Houghton-Mifflin.

Shneiderman, B. (2010). *Designing the user interface: strategies for effective human-computer interaction*. Pearson Education India.

Shukri, S. A. I. A., Arshad, H., & Abidin, R. Z. (2017). The design guidelines of mobile augmented reality for tourism in Malaysia. In *AIP Conference Proceedings Vol. 1891*, No. 1, 020026. AIP Publishing.

Silvers, A. (2017). *Going with xAPI.* Retrieved from http://makingbetter.us/portfolio/going-xapi-aaron-silvers/. Accessed November 12, 2017.

Skinner, B.F. (1974). *About Behaviorism*. New York: Knopf.

Snape, D., & Spencer, L. (2003). The foundations of qualitative research. In J. Richie & J. Lewis (Eds.), *Qualitative Research Practice,* 1-23. Los Angeles: Sage.

Sommerauer, P., & Müller, O. (2014). Augmented reality in informal learning environments: A field experiment in a mathematics exhibition. *Computers & Education, 79*, 59-68. [15]

Sommerauer, P., Müller, O. (2018). Augmented Reality for Teaching and Learning – a Literature Review on Theoretical and Empirical Foundations. In: *Proceedings of the 26th European Conference on Information Systems* (ECIS). Portsmouth, UK.

Sotiriou, S., & Bogner, F. X. (2008). Visualizing the invisible: Augmented reality as an innovative science education scheme. *Advanced Science Letters*, 1(1), 114-122.

Squire, K. D., & Jan, M. (2007). Mad City Mystery: Developing scientific argumentation skills with a place-based augmented reality game on handheld computers. *Journal of Science Education and Technology*, 16(1), 5-29. [77]

Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of communication*, *42*(4), 73-93.

Stork, S., & Schubö, A. (2010). Human cognition in manual assembly: Theories and applications. *Advanced Engineering Informatics, 24*(3), 320-328. [109]

Straub, D., Boudreau, M.-C., & Gefen, D. 2004. "Validation Guidelines for IS Positivist Research," *Communications of the AIS* (13:24), 380-427.

Sutcliffe, A. (2016). Designing for User Experience and Engagement. In *Why Engagement Matters*, 105-126. Springer International Publishing.

Syberfeldt, A., Holm, M., Danielsson, O., Wang, L., & Brewster, R. L. (2016). Support systems on the industrial shop-floors of the future–operators' perspective on augmented reality. *Procedia CIRP*, *44*, 108-113.

Swan, J.E., & Gabbard J.L. (2005). Survey of user-based experimentation in augmented reality. In: *Proceedings of 1st Int. Conference on Virtual Reality*. 22, 1-9.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York: Springer.

Szalavári, Z., Schmalstieg, D., Fuhrmann, A., & Gervautz, M. (1998). Studierstube: An environment for collaboration in augmented reality. In: *Virtual Reality*, 3(1), 37–48.

Ternier, S., Klemke, R., Kalz, M., Van Ulzen, P., & Specht, M. (2012). ARLearn: Augmented Reality Meets Augmented Virtuality. *Journal of UCS*, 18(15), 2143-2164. [110]

Tan, H. S. (2003). *The use of a virtual environment in the education of engineering students* (Doctoral dissertation, Durham University). [104]

Tseng, Y. C., Wu, S. L., Liao, W. H., & Chao, C. M. (2001). Location awareness in ad hoc wireless mobile networks. *Computer, 34*(6), 46-52.

Vaishnavi, V., & Kuechler,W. (2007). *Design science research methods and patterns: innovating information and communication technology*. Boston: Auerbach Publications.

Van Eck, R. (2006). *Digital game-based learning: It's not just the digital natives who are restless.* EDUCAUSE review, 41(2), 16.

Van Krevelen, D.W., Poelman, R. (2010). A survey of augmented reality technologies, applications and limitations. In: *International journal of virtual reality, 9*(2), 1.

Van Raaij, E. M., & Schepers, J. J. (2008). The acceptance and use of a virtual learning environment in China. *Computers & Education*, *50*(3), 838-852.

Veliz Reyes A. (2015) *Augmented pedagogies.* Available from: British Library EThOS, Ipswich, MA. Accessed November 12, 2017. [44]

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science, 46*(2), 186-204.

Vicent, L., Villagrasa, S., Fonseca, D., & Redondo, E. (2015). Virtual learning scenarios for qualitative assessment in higher education 3D arts. *Journal of universal computer science, 21*(8), 1086-1105.

Wagner, D., & Schmalstieg, D. (2009, July). History and future of tracking for mobile phone augmented reality. In *2009 International Symposium on Ubiquitous Virtual Reality*, 7-10. IEEE.

Webster, J. & Watson, R.T. (2002). Analyzing the Past to Prepare For the Future: Writing a Literature Review. *MIS Quarterly, 26* (2), xiii-xxiii.

Wenger, E. (1987). *Artificial Intelligence and Tutoring Systems: Computational and Cognitive Approaches to the Communication of Knowledge*. San Francisco: Morgan Kaufmann.

Whittemore, R., & Knafl, K. (2005). The integrative review: updated methodology. *Journal of advanced nursing, 52*(5), 546-553.

Wittrock, M. (1992). Generative learning processes of the brain. *Educational Psychologist, 27*(4), 531–541.

Wu, H. K., Lee, S. W. Y., Chang, H. Y., & Liang, J. C. (2013). Current status, opportunities and challenges of augmented reality in education. *Computers & education*, *62*, 41-49.

Wu, T. T., Sung, T. W., Huang, Y. M. & Yang, C. S. (2010). Location Awareness Mobile Situated English Reading Learning System. *Journal Of Internet Technology, 11*(7), 923-933.

Zender, R., Weise, M., von der Heyde, M., & Söbke, (2018). Lehren und Lernen mit VR und AR – was wird erwartet? Was funktioniert?. In *Proceedings of DeLFI Workshops 2018*.

Zhang, J., Sung, Y. T., Hou, H. T., & Chang, K. E. (2014). The development and evaluation of an augmented reality-based armillary sphere for astronomical observation instruction. *Computers & Education, 73*, 178-188. [26]

Zhou, F., Duh, H. B. L., & Billinghurst, M. (2008). Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR. In *Proceedings of the 7th IEEE/ACM international symposium on mixed and augmented reality*, 193-202). IEEE Computer Society.