Augmented Reality for Learning – The Role of Contextuality, Interactivity, and Spatiality for AR-based Learning Experiences

Von der Fakultät für Ingenieurwissenschaften
Abteilung Informatik und Angewandte Kognitionswissenschaft
der Universität Duisburg-Essen

zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften (Dr. rer. nat.)

genehmigte kumulative Dissertation

von

Jule Marleen Krüger
aus
Recklinghausen

1. Gutachter: Prof. Dr. Daniel Bodemer
2. Gutachter: Prof. Dr. Eleni Kyza

Tag der mündlichen Prüfung: 10.08.2023

Table of Contents

Acknowledgments

Finally. With this dissertation I am finishing my doctoral project, building on the work of the last five years, and combining it all into one text, providing an integrated picture of everything I have done. The conclusion of this chapter is an important step in my academic career and I'm incredibly thankful for all the people who supported me during this time – I could not have done it without you!

In the academic context, first and foremost my thanks go out to my supervisor, Daniel Bodemer, who encouraged me and supported my ideas and wishes, helping to cultivate them into concepts and insights. Thank you for the many discussions we had on work-related and -unrelated topics, I enjoyed them all immensely. Thank you for your guidance throughout the whole process of my doctoral project, I could not have done this without you.

I would also like to thank my PsychMeth colleagues who were a big part of the process of this dissertation. Thank you for carrying me through this, with all its ups and downs. Thank you for lots of stimulating conversations and inspiring interaction. Thank you for all the joy you brought me and the many late nights. I'm sorry for all the bad AR puns. Stay ARwesome!

Additionally, I want to thank all my co-authors and everyone who contributed to the creation of the learning material, data collection, data analysis or other steps of the research process of the respective papers and studies. So much effort has been put into all the research steps, and the support through Bachelor's and Master's theses candidates and student assistants was essential for its success.

Outside of the academic context, I would also like to thank the other people in my life who accompanied me during this time, for all of the way or part of it – my family and friends, acquaintances from all contexts, and strangers who smiled at me. Thank you for your constant support, even if my answer to “How is your doctoral project coming along?” may have often been a death stare.

Furthermore, I want to thank everyone who finds this work interesting enough to read it, may it be to gather insights for their own research, to see how a dissertation should or should not be written, or just for fun – I hope it helps you on your way and you enjoy reading it!

I am looking forward to doing much more work in the field of learning with AR in the future. This is just the beginning. You know what they say: life is a highway – and I don't know how to stop.

Abstract

Augmented reality (AR) is a form of presenting information by combining virtual and physical elements. This combination can be leveraged for unique learning scenarios, providing learners with information through interactive and spatial representations contextualised in a physical environment. While the technology necessary for this is already commonly used and outcomes of research generally suggest positive effects on learning processes and outcomes, the specific mechanisms that play a role for learning in AR are not yet fully established and examined. The aim of this dissertation is to fill this gap and provide insights into specific characteristics of AR-based learning and how they can be leveraged to support learning processes and outcomes. Based on three subgoals, a theoretical framework on educational AR is introduced, the results of systematic empirical studies are analysed, and recommendations for practical implementations are made. The theoretical ARcis framework developed as part of the dissertation explores the unique features of AR and elaborates the three characteristics contextuality, interactivity, and spatiality. Contextuality describes the integrated perception of virtual and physical elements, interactivity describes the manipulation of these elements in different ways, and spatiality describes the perception of spatial elements in 3D space. These characteristics can have an influence on learning processes and outcomes and can be leveraged for the design of systematic research. Furthermore, they can be used to develop AR experiences in a goal-oriented way. First empirical insights on the three characteristics and their influence on learning were collected in seven studies that are part of the five papers included in this dissertation. The studies focus on specific aspects of the AR-based learning experience, and most were designed as value-added studies with one of the three characteristics in mind. The study outcomes suggest a positive influence of the implementation of combined virtual and physical elements, guided mental and physical interactivity, and spatial representations in educational AR on cognitive and motivational processes and outcomes. In addition to the theoretical framework and empirical studies, practical design implications are described and analysed. It is proposed that the design of learning material should be aligned with the learning goals of the situation, leveraging the three ARcis characteristics for the design of purposeful AR experiences. All in all, the definition of the unique characteristics of AR, the outcomes of the empirical studies, and recommendations for practical application can inform research and practice on learning with AR.

Zusammenfassung

Augmented Reality (AR) ist eine Form der Informationsdarstellung, bei der virtuelle und physische Elemente kombiniert werden. Diese Kombination kann für einzigartige Lernszenarien genutzt werden, in denen Lernenden Informationen durch interaktive und räumliche Repräsentationen kontextualisiert in einer physischen Umgebung vermittelt werden. Während die dafür notwendige Technologie bereits weit verbreitet ist und Forschungsergebnisse im Allgemeinen auf positive Auswirkungen auf Lernprozesse und -ergebnisse hindeuten, sind die spezifischen Mechanismen, die beim Lernen mit AR eine Rolle spielen, noch nicht vollständig bekannt und erforscht. Das Ziel dieser Dissertation ist es, diese Lücke zu füllen und Einblicke in die spezifischen Merkmale des AR-basierten Lernens zu geben und zu zeigen, wie diese zur Unterstützung von Lernprozessen und -ergebnissen genutzt werden können. Auf Basis von drei Teilzielen wird ein theoretischer Rahmen für AR in der Bildung vorgestellt, die Ergebnisse systematischer empirischer Studien werden analysiert und Empfehlungen für die praktische Umsetzung gegeben. Das im Rahmen der Dissertation entwickelte theoretische ARcis Framework untersucht die besonderen Merkmale von AR und arbeitet die drei Eigenschaften Kontextualität, Interaktivität und Räumlichkeit heraus. Kontextualität beschreibt die integrierte Wahrnehmung virtueller und physischer Elemente, Interaktivität beschreibt die Manipulation dieser Elemente auf unterschiedliche Weise und Räumlichkeit beschreibt die Wahrnehmung räumlicher Elemente im 3D-Raum. Diese Eigenschaften können einen Einfluss auf Lernprozesse und -ergebnisse haben und für die Gestaltung systematischer Forschung genutzt werden. Darüber hinaus können sie genutzt werden, um AR-Erfahrungen zielgerichtet zu entwickeln. Erste empirische Erkenntnisse zu den drei Eigenschaften und ihrem Einfluss auf das Lernen wurden in sieben Studien gesammelt, die Teil der fünf in dieser Dissertation enthaltenen Arbeiten sind. Die Studien konzentrieren sich auf spezifische Aspekte der AR-basierten Lernerfahrung, und die meisten wurden als Mehrwertstudien mit einer der drei Eigenschaften im Hinterkopf konzipiert. Die Studienergebnisse deuten auf einen positiven Einfluss der Implementierung kombinierter virtueller und physischer Elemente, angeleiteter mentaler und physischer Interaktivität und räumlicher Repräsentationen in bildungsbezogener AR auf kognitive und motivationale Prozesse und Ergebnisse hin. Neben dem theoretischen Rahmen und den empirischen Studien werden auch praktische Implikationen für die Gestaltung beschrieben und analysiert. Es wird vorgeschlagen, dass die Gestaltung des Lernmaterials auf die Lernziele der Situation abgestimmt werden sollte, indem die drei ARcis Eigenschaften für die Gestaltung zielgerichteter AR-Erfahrungen genutzt werden. Alles in allem können die Definition der einzigartigen Eigenschaften von AR, die Ergebnisse der empirischen Studien und die Empfehlungen für die praktische Anwendung die Forschung und Praxis zum Lernen mit AR informieren.

List of Included Papers

The following papers are included in this cumulative doctoral dissertation. Paper 1, 2, 3, and 5 have been published in various peer-reviewed scientific journals and conference proceedings. Paper 4 has been submitted for publication.

Paper 1 (ARcis Framework, Studies 1 and 2)

Krüger, J. M., Buchholz, A., & Bodemer, D. (2019). Augmented reality in education: three unique characteristics from a user's perspective. In M. Chang, H.-J. So, L.-H. Wong, F.-Y. Yu, & J. L. Shih (Eds.), *Proceedings of the 27th International Conference on Computers in Education, Volume 1* (pp. 412-422). Asia-Pacific Society for Computers in Education. https://apsce.net/icce/icce2019/04_Proceedings.html

Paper 2 (Extension of ARcis Framework concerning interactivity, Study 3)

Krüger, J. M., & Bodemer, D. (2020). Different types of interaction with augmented reality learning material. In D. Economou, A. Klippel, H. Dodds, A. Peña-Rios, M. J. W. Lee, D. Beck, J. Pirker, A. Dengel, T. M. Peres, & J. Richter (Eds.), *2020 6th International Conference of the Immersive Learning Research Network (iLRN)* (pp. 78–85). Immersive Learning Research Network. https://doi.org/10.23919/iLRN47897.2020.9155148

Paper 3 (Extension of ARcis Framework concerning spatiality, Study 4)

Krüger, J. M., Palzer, K., & Bodemer, D. (2022). Learning with augmented reality: Impact of dimensionality and spatial abilities. *Computers and Education Open, 3*, Article 100065. https://doi.org/10.1016/j.caeo.2021.100065

[Krüger, J. M., Palzer, K., & Bodemer, D. (2023). Corrigendum to 'Learning with augmented reality: Impact of dimensionality and spatial abilities' [Computers and Education Open, Volume 3 (December 2022), Article 100065]. *Computers and Education Open, 4*, Article 100127. https://doi.org/10.1016/j.caeo.2023.100127]

Paper 4 (Extension of ARcis Framework concerning contextuality, Study 5)

Krüger, J. M., & Bodemer, D. (subm.). *Positioning augmented reality information for learning in nature: An exploratory pilot study* [Manuscript submitted for publication].

Paper 5 (Extension of ARcis Framework concerning specific applications, Studies 6 and 7)

Krüger, J. M., & Bodemer, D. (2022). Application and investigation of multimedia design principles in augmented reality learning environments. *Information, 13*(2), Article 74. https://doi.org/10.3390/info13020074

List of Tables

List of Figures

1 Introduction

Learning technologies used in various topic areas, in formal and informal learning settings, and on different educational levels become more and more sophisticated in their hard- and software and their potential to convey information. One development that is becoming increasingly popular in education is augmented reality (AR). AR describes the visualisation of virtual information in the form of an integrated overlay onto the real world, specifically through systems combining real[1] and virtual elements, including real-time interactivity and registration in three-dimensional (3D) space (Azuma, 1997). On the reality-virtuality continuum reaching from fully real to fully virtual environments, AR has been placed within the in-between area of mixed reality (MR) leaning towards real environments (Milgram et al., 1994). While research on AR in education is becoming more mainstream and is implemented increasingly often (see, for example, reviews by Buchner & Kerres, 2023; Fidan & Tuncel, 2018; Garzón, 2021), there are still a lot of research gaps, especially when it comes to experimental research including value-added and learner-technology interaction studies instead of media comparisons (Buchner & Kerres, 2023). For systematic and experimental research, a more detailed definition of learning-related attributes of educational AR is helpful, so that a focus on mechanisms concerning theses specific attributes is possible. In this doctoral dissertation, I present the ARcis framework, which I describe as a basis for more systematic research examining specific components of AR-based learning experiences that can be leveraged to support the achievement of different learning goals: contextuality, interactivity and spatiality. I will describe this framework, results from first empirical studies implemented on its basis and its potentials for application in this dissertation.

1.1 Technological Background of AR

From a technological perspective, AR systems have been defined through three characteristics: 1) combining real and virtual elements, 2) real-time interactivity and 3) registration in 3D space (Azuma, 1997). In order to achieve this, devices require three hardware components: sensors, a processor, and a display (Craig, 2013a). Sensors are especially relevant for the registration in 3D space described by Azuma (1997), which requires the tracking of the device's position and the physical environment so that the virtual elements can be placed in alignment with physical objects. This can be done by sensors like GPS, accelerometers, depth-sensors, or a camera implementing computer vision (see Craig, 2013b for a review of AR hardware). While not all devices have sensor-based forms of environmental tracking

[1] While AR is often described as having "real" (or "real-world") and "virtual" components, in the current dissertation I will mainly label this distinction as "physical" and "virtual". Because virtual elements can represent real(-world) elements, it would be wrong to say that they are generally "not real". As to not confuse these levels of represented and representing components, I use the "physical" property to distinguish the elements. Physical elements can be perceived in a non-mediated way and are often (but not exclusively) natural objects in an already existing environment that is either not designed by humans, or at least not designed with the current learning objective in mind. Virtual elements are only perceivable by mediation or translation through some sort of technological device, e.g., a screen or a speaker, and are often designed by humans specifically for some purpose, e.g., a learning objective. In some of the papers included in this dissertation, we still make the distinction between "real"/"real-world" and "virtual" elements, because my use of the labels changed over time after gaining more insights into the field of research.

beyond the camera available, an easier implementation of the registration in 3D space can be achieved through specific physical anchor-objects, e.g., physical objects or images functioning as AR markers called “fiducial markers” (Craig, 2013a, p. 41). These are easily recognised and tracked through computer vision, and virtual objects can be attached to them without the necessity to track the whole physical context. Real-time interactivity of the virtual elements also requires sensors for tracking user input and can technologically be achieved in different ways depending on the form of interaction. It can include body tracking for whole-body interaction, hand-tracking for hand-based interaction, touch-based interaction through a screen, or interaction through a form of controller (Craig, 2013a). All of these forms of interaction distinguish AR from just watching a movie with CGI elements (Azuma, 1997). Concerning the display requirement, which is crucial in transferring the information from the system to the user, the perception of co-present virtual and physical elements can be achieved through video see-through or optical see-through displays (Milgram et al., 1994). In video see-through, the virtual elements are overlayed onto a real-time stream of the camera view of a device, whereas in optical see-through displays, virtual elements are projected directly into the user’s field of view, for example through semi-transparent mirrors. The devices used for this can in both cases take the form of immersive displays like head-mounted displays (HMDs) or can be achieved as monitor-based options, although the latter are most often video-based in the form of tablets or smartphones.

While the required hardware for AR including sensors and displays shows that sophisticated technology is necessary for AR to be achieved, most smartphones and tablets are already equipped with sufficient features to implement video see-through AR: a camera to record the physical world and for optical tracking, a GPS sensor for location detection, accelerometer and gyroscope sensors for detection of device movement and a touchscreen for interaction and to display virtual elements integrated into the camera stream of the physical world (Craig, 2013b). This has also influenced the implementation of AR in educational settings. After the first generation of AR in education has been defined as being driven by access to specific hardware, the second generation from 2010 to 2020 was based on the use of AR applications on mobile devices (Garzón, 2021). In a systematic review on AR in K-12 education, it was shown that while from 2000 to 2013 only 17% of studies used video see-through in output devices, this increased to 62% in studies from 2014 to 2020 (Zhang et al., 2022). This can mainly be attributed to the increasing usage of tablets and smartphones from 2014, with over 80% of studies including these devices in 2018 to 2020. These data show that AR is already being widely used in education through smart mobile devices, which most people have access to in their daily life.

1.2 Research on AR in Education

Looking back at the last five years of horizon reports for technological developments that might become interesting for educational settings, AR was a highlighted technology under the umbrella term mixed reality (MR) with a time-to-adoption horizon of four to five years in 2018 (Becker et al., 2018) and two to three years in 2019 (Alexander et al., 2019). In 2020, the umbrella term extended reality (XR) was used to cluster AR, virtual reality (VR), MR and haptic technologies as a highlight in the emerging technologies and practices category (O'Brien, 2020). In 2021 (Pelletier et al., 2021) and 2022 (Pelletier et al., 2022), however, AR was not mentioned as a key technology or practice in the horizon reports, which may show a shift in focus away from XR, but may also show that AR and VR are becoming more usual in the educational technologies landscape, instead of being highlights to adopt in the future.

The growing usage of and research on AR in education is reflected in the growing number of systematic literature reviews and studies executed over the last years. When just looking at some reviews on AR in education in recent years, a steady increase of studies is apparent, for example shown in a systematic review by Garzón and colleagues (2019) with an increase from 33 studies in 2012 to 154 studies in 2018, with nearly 50% of the evaluated studies in the educational field 'natural sciences, mathematics and statistics', and primary (31%) and bachelor's level education (30%) as the most prevalent target groups. More recent systematic reviews still show a steady increase in research on AR in education over the last years, with studies in the Web of Science database nearly doubling from 2016 (220 studies) to 2019 (436 studies; Garzón, 2021), and a similar pattern of increase from 2016 (7 studies) to 2020 (20 studies) also found when specifically looking at top educational technology journals (Buchner & Kerres, 2023). In a recent meta-analyses by Chang and colleagues (2022), 134 (quasi-)experimental studies on AR in education from 2012 to 2021 were analysed. The subject area 'science' was included in almost 50% of studies in this review and most studies took either place at the elementary (39%) or the postsecondary (36%) level of education.

In addition to the general reviews of research on AR in education described above, there have also been an increasing number of systematic, scoping, mapping, or integrative reviews for specific areas of application and subject domains. Specific areas of application in which research on educational AR has been systematically reviewed include K-12 education (Law & Heintz, 2021; Zhang et al., 2022), higher education (López-Belmonte et al., 2019; Mystakidis et al., 2021), professional training (Han et al., 2022), vocational training (Chiang et al., 2022), patient education (Urlings et al., 2022), and informal learning sites (Goff et al., 2018). Even more systematic reviews on AR in education considering specific subject domains became available over the last years. Different subject domains in which research on educational AR has been systematically reviewed include science learning and education (Arici et al., 2019; Jdaitawi et al., 2022; Xu et al., 2022), STEM education (Ibáñez & Delgado-Kloos, 2018; Mystakidis et al., 2021; Sırakaya & Alsancak Sırakaya, 2020), chemistry education (Mazzuco et al., 2022), physics education (J. W. Lai & Cheong, 2022), programming education (Theodoropoulos & Lepouras, 2021), engineering education (Álvarez-Marín & Velázquez-Iturbide, 2021; Vásquez-Carbonell, 2022), architecture, construction and civil engineering education (Diao & Shih, 2019; Hajirasouli & Banihashemi, 2022), language learning (Cai et al., 2022; Parmaxi & Demetriou, 2020), health sciences (Rodríguez-Abad et al., 2021), healthcare education (Gerup et al., 2020), medical education (Parsons & MacCallum, 2021; Tang et al., 2020), anatomy education (Bölek et al., 2021; Chytas et al., 2020; McBain et al., 2022), and history education and heritage visualisation (Challenor &

Ma, 2019). This shows how broad the field of research on AR concerning different application areas and subject domains has recently become.

When looking at recent research on AR in education, most of the research is technology-driven with a focus on media comparisons, i.e. comparing AR-based implementations with other digital implementations or a traditional educational practice like a book or lecture. In a meta-analysis of 134 papers including (quasi-)experimental studies on AR in education from 2012 to 2021, H.-Y. Chang and colleagues (2022) found that 84% of 201 comparisons compared AR with non-AR instruction. Even when only looking at studies on AR in education published in top journals on educational technology, the picture looks very similar, with 80% of 92 studies published from 2009 to 2020 using a media comparison approach, even increasing over the last years although this approach has been highly criticised (Buchner & Kerres, 2023). Surry and Ensminger (2001) describe three major criticisms of media comparisons: 1) One medium is not inherently better than another, as it only delivers the information, and the instructional method is more important than the medium. This is based on the delivery truck argument described by R. E. Clark (1983). 2) More understanding of the specific attributes of media is necessary, so that better study designs can be developed with a focus on these attributes and how they can be used to support learning in different types of learners. 3) Due to the many differences between different media, there are a lot of confounding variables that make it difficult to figure out which factors led to an effect in an empirical study. This is based on the pseudoscience argument described by Reeves (1995).

Two alternatives to media comparison studies that still focus on the role of the technology proposed by Surry and Ensminger (2001) are intra-medium studies, comparing different designs of the same technology, and aptitude-treatment-interaction studies, looking at how the technology might be (dis)advantageous for different types of learners. Only 22% of the 92 studies on AR in education analysed by Buchner and Kerres (2023) were found to be value-added studies, and only 11% were studies focusing on learner characteristics. This shows a gap in the research when it comes to these types of studies. Buchner and Kerres (2023) also state the necessity for alternative studies concerning AR in education, proposing to execute more research concerning the "how" of using AR in the form of value-added studies (i.e. intra-medium studies) and concerning the "when" of using AR in the form of learner-treatment-interaction studies (i.e. aptitude-treatment-interaction studies) and value-added studies comparing different learning outcomes. To receive a more complete picture concerning the effective and efficient usage of AR for learning, these kinds of studies should be implemented instead of or in addition to media comparison studies (see also Section *1. Introduction* in Paper 3, Krüger et al., 2022).

Looking at the second point of criticism by Surry & Ensminger (2001), the authors state that more information is necessary about attributes of specific media and how these influence learning in order to design better studies on educational technology. Kozma (1994) makes a similar point, stating that the attributes or capabilities of a medium need to be defined by the symbol systems or forms of representations that it can leverage for information communication and its capabilities to process this information. He further states that these attributes should in turn be defined based on how they interact with and influence learners' construction of and operation on their internal representations so that it can be determined how the medium could be used effectively to support learning. The goal of the current doctoral dissertation is formulated based on those suggestions.

1.3 Goals of the Current Dissertation

The goal of the current dissertation can be summarised as: "Gaining insight into specific characteristics of AR-based learning and how they can be leveraged to support learning processes and outcomes". I further define three subgoals for reaching this goal, a theoretical, an empirical, and a practical subgoal:

1) theoretically defining characteristics of learning with AR and analysing how specific mechanisms may have an impact on learning
2) empirically examining how the characteristics of learning with AR and their specific mechanisms influence learning
3) practically applying the theoretical and empirical insights into the characteristics and mechanisms for the design of AR-based learning experiences

To work towards Subgoal 1, I will present three AR-specific characteristics based on the literature on AR-based learning and their potential for the support of effective and efficient learning on both a process and an outcome level in Section 3. These characteristics can be used for systematic empirical research on AR-based learning(Subgoal 2) and the purposeful design of AR experiences (Subgoal 3). Concerning Subgoal 2, I will describe empirical studies based on the three characteristics that are part of my doctoral project in Section 4. Furthermore, for Subgoal 3 I will take the step towards the application of the framework in the design of AR-based learning experiences. Because this is based on both the definition of the characteristics and the outcomes of the studies, it will be interwoven with all parts of the dissertation and summarised in the practical implications in Section 5.3.

The basis for reaching the overarching goal and subgoals are five papers that are included in this cumulative doctoral dissertation. The papers all include a theoretical part, constructing or expanding the theoretical ARcis framework guided by Subgoal 1 (see Section 3 for a description of the framework), and a total of seven empirical studies guided by Subgoal 2 (see Section 4 for an overview of all papers and studies). In Paper 1 (see Section 4.1), the construction and description of the underlying ARcis framework with the three characteristics contextuality, interactivity, and spatiality is provided, which is further extended in the other four papers. The empirical part in Paper 1 includes two studies firstly subsumed and integrated into the ARcis framework: Study 1 looks at a comparison of AR and non-AR simulations for inquiry-based learning, with differing interactivity and spatiality, and Study 2 compares close and far visualisations for group formation support, focusing on contextuality (Paper 1, Krüger et al., 2019). Paper 2 (see Section 4.2) focuses on the ARcis characteristic interactivity, theoretically expanding the ARcis framework concerning mental and physical interaction in AR. Empirically, Paper 2 includes Study 3, looking at the impact of higher and lower mental and physical interaction in an interactive AR-based learning setting (Paper 2, Krüger & Bodemer, 2020). Paper 3 (see Section 4.3) focuses on the ARcis characteristic spatiality, expanding the theoretical framework concerning the dimensionality of objects in AR. The empirical part includes Study 4, comparing a 3D and a 2D visualisation in AR (Paper 3, Krüger et al., 2022). In Paper 4 (see Section 4.4), contextuality is in focus with an expansion of the ARcis framework concerning the placement of virtual information in relation to contextually relevant physical objects. The included Study 5 compares placement of virtual information near to and far from corresponding physical objects (Paper 4, Krüger & Bodemer, subm.). In Paper 5 (see Section 4.5), the connection of spatiality and contextuality to the multimedia learning principles expands the focus of the ARcis framework to more specific design applications. Study 6 included in Paper 5 describes the application of the spatial contiguity principle, and Study 7 describes the application of the coherence principle (Paper 5, Krüger & Bodemer, 2022a). Data collection for Studies 1, 2, 6 and 7 (Study 5 partly) were conducted as part of advised Bachelor's and Master's theses. In Table 1, an overview of the five papers and their components, including the theoretical developments and the seven studies, is shown. These will be described in more detail in Section 4.

Table 1

Overview of Papers and Studies

Papers	Components
Paper 1 *Basic ARcis Framework* (Krüger et al., 2019)	▪ Framework: Basis of ARcis Framework ▪ **Study 1**: Inquiry-Based Learning in AR ▪ **Study 2**: Group Formation with AR Support
Paper 2 *Interactivity in AR* (Krüger & Bodemer, 2020)	▪ Framework: Extension of ARcis Framework concerning Interactivity ▪ **Study 3**: Physical and Mental Interaction in AR
Paper 3 *Spatiality in AR* (Krüger et al., 2022)	▪ Framework: Extension of ARcis Framework concerning Spatiality ▪ **Study 4**: Dimensionality and Spatial Abilities in AR
Paper 4 *Contextuality in AR* (Krüger & Bodemer, subm.)	▪ Framework: Extension of ARcis Framework concerning Contextuality ▪ **Study 5**: Position of Physical Context in AR
Paper 5 *Multimedia Learning in AR* (Krüger & Bodemer, 2022a)	▪ Framework: Extension of ARcis Framework concerning application of Multimedia Principles ▪ **Study 6**: Spatial Contiguity Principle in AR ▪ **Study 7**: Coherence Principle in AR

2 Learning with AR

AR-based learning is a growing research area with a lot of reviews on the usage of AR in education over the last years as already described in Section 1.2. In the AR-based settings used in this research, many different pedagogical approaches and strategies have been applied. Sommerauer and Müller (2018), for example, identified the cognitive theory of multimedia learning (CTML), mobile learning, game-based learning and simulation, experiential learning and situated learning as learning theories implemented in educational AR. Garzón and colleagues (2020) similarly identified CTML and situated learning, but also collaborative learning, inquiry-based learning, and project-based learning as relevant in this regard. Hanid and colleagues (2020) similarly found the learning strategies of game-based learning, collaborative learning, and experiential learning, and furthermore also interactive learning to be prevalent in AR-based education. In the more specific case of STEM learning in higher education, Mystakidis and colleagues (2021) also described the instructional strategies of collaborative learning, experiential and discovery learning, activity-based learning, but also a simple presentation of information in the reviewed literature. This shows that researchers try to incorporate pedagogical approaches into the AR applications that they examine, which include collaborative, experiential/discovery/inquiry-based learning, activity-based/interactive learning, game-based learning, mobile learning, project-based learning, situated learning, the application of the CTML, and the simple presentation of information. Goal of the implementation of these different pedagogical approaches is usually to improve learning processes and outcomes, although the focus can be very diverse for different application designs and empirical studies.

Different variables have been applied in empirical studies to examine learning processes and outcomes when learning in AR. Radu (2012) summarises positive effects of using AR in education in the form of increased content understanding, long-term memory retention, increased student motivation, and improved collaboration. Further reviews suggest that AR can facilitate skill and knowledge acquisition more effectively, support understanding and knowledge transfer, increase motivation and interest, improve spatial abilities and psychomotor-cognitive skills (Wu et al., 2013) and improve content understanding, memory retention, task performance, collaboration, and motivation (Radu, 2014). Enhancement of learner outcomes in the form of learning achievement, learning motivation or decrease of cognitive load, and pedagogical contributions in the form of increased enjoyment and level of engagement or interest have also been listed (Akçayır & Akçayır, 2017). Garzón and colleagues (2019) list learning gains, motivation, comprehension of abstract concepts, autonomy, sensory engagement, memory retention, collaboration, creativity, and accessibility as advantages of AR reported in the 61 studies they assessed. In total, there have thus been cognitive, motivational, and emotional variables in the form of learning processes and outcomes that have been part of research on AR. This shows a wide range of outcome variables, hinting at very different mechanisms that can play a role when learning with AR. In the following sections, I will describe in more detail how AR can have an influence on learning, starting with an analysis of which representations can be applied in AR. Afterwards, I will describe relevant constructs for AR-based learning and more specific insights into how AR has been found to influence learning processes and outcomes in the literature.

2.1 Multiple External Representations

As described above in Section 1.2, Kozma (1994) states that media should be defined by their attributes, including their capabilities of using specific symbol systems or forms of representations, and in turn define how these attributes can support learning processes. Representations can be defined as elements that represent something else, involving a represented and a representing world (Ainsworth, 2006). In AR, various different forms of representations can be implemented due to its combined virtual and physical nature. On the physical side, placing AR-based learning experiences in real-world environments includes accessibility to natural objects (e.g., plants or landscapes) that can represent more general concepts and objects in a learning experience. These natural elements are limited to existing environments and objects that are available to and accessible by the learners. They include 3D objects, spatial environments, and sounds, but also the possibility to touch and feel things, smell, and taste. Natural physical environments and objects are very rich in multisensory information, with a high density of perceptions, realistic representations, and authentic contexts. Besides natural objects, the physical side of AR can also include artificial objects, for example buildings or physical models. Physical models representing natural objects are tangible but may not be as rich in multisensory information as their natural counterparts. On the virtual side in AR, virtual representations can be manifold and specifically designed by the instructors. Written text, 2D images, 3D models, spoken text, and sounds are some examples of representations that can be implemented. The elements can be dynamic or static and interaction with them can be designed exactly as desired. AR thus combines physical and virtual elements. Rau (2020) describes physical representations as tangible and thus manipulatable by hand, and virtual representations as presented on a screen manipulated through controllers or touchscreen input. Based on a review of the complementary advantages of physical and virtual representations, one suggested heuristic for instructional design describes that physical and virtual representations should be combined purposefully. Their respective potentials for the conveyance of different concepts should be taken into account as they offer complementary advantages through different learning mechanisms. The specific combination and design of physical and virtual representations in AR should thus be considered, which can be guided by theories on human information processing.

One theory of human information processing is the Cognitive Theory of Multimedia Learning (CTML) by Mayer (2020a), including three assumptions. The first assumption states that information is processed in two separate channels, one for visual-spatial and one for auditory-verbal information. This dual-channels assumption is based on an integration of Paivio’s dual-coding theory (e.g., Paivio, 1986) and Baddeley’s working memory model (e.g., Baddeley, 1999). Because AR can include both visual-spatial and auditory-verbal information in the form of images, written text, spoken text, and sounds, the two channels of information processing can be leveraged for the design of effective AR-based learning experiences, for example by taking into account different multimedia learning principles. The second assumption of CTML describes that the processing capacity of human working memory is limited (Mayer, 2020a). The three types of processing, extraneous processing (caused by bad instructional design, no goal-oriented learning), essential processing (caused by complexity of content, essential learning) and generative processing (motivated effort of learner, deeper understanding) should thus be distributed in working memory in a way as to support learning best, avoiding extraneous, managing essential, and facilitating generative processing. This can be connected to cognitive load theory, which will be described in more detail in Section 2.3. The third assumption of CTML describes that processing is always active because information has to be actively attended, selected, organised, and integrated (Mayer, 2020a). The active processing can then lead to the construction of a coherent mental model or knowledge structure. CTML can be applied to AR due to its potential to include various forms of representation. Many authors have suggested CTML as a relevant learning theory for AR (e.g., Buchner, Buntins, et al., 2022; Garzón et al., 2020; Sommerauer & Müller, 2018). Due to its partly digital nature, AR can include pictorial and textual representations, and information for different sensory modalities, like vision and audio, making AR a perfect environment for multimedia learning and multiple representations. CTML as a theory of information processing can thus inform the design of the combination of representations in AR-based learning experiences.

The integrated model of text and picture comprehension by Schnotz and Bannert (2003) describes a similar approach of information processing. The authors distinguish between descriptive representations that use symbols to describe objects with signs for relations, and depictive representations that use iconic signs and structural relations. In the model, these forms of representations are processed in different channels because they are based on complementing but different sign systems and principles of representations (Schnotz & Bannert, 2003). External descriptive representations are described to be processed into internal text surface representations, then a deeper level propositional representation which can then be constructed into a mental model, at which point a transition from a descriptive to a depictive representation takes place. External depictive representations are described to be processed into an internal visual perception or image and then a deeper level mental model, which can inspect propositional representations. A continuous construction-inspection interaction takes place between propositional representation and mental model. So, different than with CTML, no one-to-one relationship between external and internal representations is assumed. External descriptive representation leads to both internal descriptive and depictive representation, and the same is assumed for external depictive representation (Schnotz & Bannert, 2003). In AR, external depictive and descriptive representations can be displayed and combined, so that this model of human information processing can also be used to inform the design of AR-based learning experiences.

It is thus apparent that the combination of different forms of external representations needs to be taken into account when designing in AR but additional aspects of multiple external representations (MERs) beyond their form should also be considered. The DeFT framework (Ainsworth, 2006) describes an approach that does not only focus on the form of the representation, but also the design parameters unique to MERs, their pedagogical functions, and cognitive tasks necessary for interaction. Concerning the design parameters, Ainsworth describes that both the represented and the representing world need to be considered in external representations. For MERs she lists unique design dimensions, including the number, the information distribution, the form of the representations and the translation between them. Concerning the functions of MERs, Ainsworth (2006) lists complementary, constraining, and constructing functions. Concerning the cognitive tasks necessary to learn with MERs, Ainsworth

(2006) describes that learners need to understand the form of the representation, the relation between domain and representation, and choose how to select or construct an appropriate representation. In AR, the design, functions, and tasks need to be applied to a unique combination of physical and virtual elements. These can complement each other, distributing information between the two worlds, enriching wisely chosen physical objects through well-designed virtual elements that can take many different forms. It is especially important to support translation between physical and virtual elements to indicate their connection because of their different nature. Also, the authentic and often familiar nature of physical objects might help constrain the interpretation of virtual elements. Deeper understanding can be constructed in this combination, for example by supporting abstraction through multiple examples from the authentic real world, extending knowledge over different contexts, and relating virtual and physical objects to each other. When it comes to cognitive tasks that learners need to achieve, some might be more easily supported through AR. For example, the relation between a virtual representation and its domain may be easier if the experience takes place within an authentic context, which AR can supply. This indicates the unique possibilities of MERs in AR.

As described in the context of the DeFT framework, one design aspect that needs to be considered is the translation and thus relation of different representations. For AR this means the translation between physical and virtual elements, which can lead to a coherent mental model of the different worlds. Seufert and Brünken (2006) describe that coherence formation between different external representations can be difficult. They distinguish between local and global coherence formation. Local coherence formation describes the understanding and mental integration of information provided within a single representation, while global coherence formation describes integration of information provided in multiple representations (Seufert & Brünken, 2006). In AR, there can be multiple virtual and multiple physical representations, so that another distinction can be added, the coherence formation within one world, virtual or physical, and the coherence formation across the two worlds. Coherence formation across physical and virtual elements may hold new challenges.

Common representations used in AR are 3D objects, text and digital media like sound and video (Mystakidis et al., 2021). Chang and colleagues (2022) take a closer look at the AR affordances that are leveraged in their meta-analysis of 134 studies, finding that 3D visualisation is exploited most often in 78% of the studies, 70% of studies exploit immersion, 60% exploit contextualisation and 25% leverage collaboration. 3D visualisations are thus an important kind of visualisation in AR (see Section *1.1.1. Augmented reality visualizations* in Paper 3, Krüger et al., 2022 for more details). An AR-based learning experience that incorporates many different kinds of representations has been presented by Shaghaghian and colleagues (2022). In their AR application, the learning objective is to understand the spatial transformation of 3D objects, including a translation into different kinds of representations based on a dynamic link. A physical 3D model is included on the physical side, and on the virtual side a virtual version of the 3D model that can be rotated based on virtual parameter input is shown as anchored to the underlying surface. In addition to overlays of transformation matrices and mathematical functions based on transformations of the physical and virtual model, graphical information about the transformations is shown integrated into the physical space. Very different symbol systems and representations are thus included in this application, helping with translations between these. Another AR-based learning experiences that shows a lot of virtual and physical representations and leverages the combination of the two worlds in an elaborate way, is the application on electromagnetism by Radu and Schneider (2022). Here, the physical part is a model that replicates an audio speaker system. Through an HMD, virtual 3D visualisations of otherwise invisible activities within the speaker are shown, including magnetism, electricity, and sound waves. Also, 2D image-based representations of electromagnetism that are also shown on a physical poster are added virtually in a combined view, so that multiple very different representations are shown and can be mentally integrated. This shows the potential range of combining representations in AR. While this offers many opportunities, there may be even more challenges when it comes to the purposeful implementation of these and thus the design of AR. Design guidelines can help instructors and may be transferable from other areas.

As stated above, CTML can be applied in the design of AR-based learning experiences. The theory includes various multimedia design principles directed at the reduction of processes that are not directly relevant to learning (i.e. extraneous processing), some directed at managing processes that deal with the content of the learning material (i.e. essential processing) and some directed at fostering processes that lead to deeper learning and schema construction (i.e. generative processing; Mayer, 2020f). Sommerauer and Müller (2014) argue and show that multiple design principles can be implemented through AR, applying the multimedia principle, the spatial and temporal contiguity principles, the signalling principle, and the modality principle in a mathematics exhibition. Similarly, another design of a mobile AR application for English language education applied the signalling principle, spatial and temporal contiguity principle, modality principle, and segmenting principle (H.-Y. Lin & Tsai, 2021). One of the principles for reducing extraneous processing is the spatial contiguity principle, which describes the advantage of visually-spatially integrated multimedia representations of related images and texts compared to separate representations (Mayer, 2020d). This principle is based on the reduction of visual search processes and on freeing capacities in working memory for in-depth processing. The split-attention effect in CLT describes the same phenomenon (Ayres & Sweller, 2014) and many empirical studies have found a positive effect on learning outcomes when the principle is followed (Schroeder & Cenkci, 2018). With AR, this principle is applicable to the combination of virtual and physical elements. These elements can be presented in the same visual view using AR technologies, being integrated at the actual location of the physical object in the three-dimensional space of the physical world (Altmeyer et al., 2020). Thees and colleagues (2020), for example, found that ECL was lower with an integrated presentation of superimposed information using AR glasses than with a presentation shown separately on a screen, even though no effects on learning outcome were found. In another study with a similar setup, they however found no lower ECL for the AR glasses condition and higher learning gains for the separate display group (Thees et al., 2022). The idea of applying multimedia learning design principles is further explored in Paper 5 including Study 6 and Study 7 in the current dissertation (Krüger & Bodemer, 2022a).

In summary, AR offers the possibility of combining various MERs, including physical and virtual representations (see Rau, 2020), verbal, pictorial, auditory, and visual representations (see CTML; Mayer, 2020a), and depictive and descriptive representations (see integrated model of text and picture comprehension; Schnotz & Bannert, 2003). Their combination, including their forms, tasks, and functions (see DeFT framework; Ainsworth, 2006), and local and global integration (see Seufert & Brünken, 2006), should be considered carefully in the design of educational applications. The current dissertation will further analyse and evaluate the specific characteristics of AR as a form of visualising information through a combination of different forms of physical and virtual representations. The resulting framework including the characteristics contextuality, interactivity, and spatiality will be described in Section 3.

2.2 Learning Achievement

Learning has been defined as a change in the learners' knowledge caused by an experience in a learning environment (Mayer, 2020a). Mayer defines different kinds of knowledge that can be achieved from an educational experience, including facts, concepts, procedures, strategies, and beliefs. He further defines five kinds of knowledge structures that are built when learning, namely process, comparison, enumeration, classification, and generalisation. Knowledge is said to be stored in long-term memory, which can hold large amounts of information for a long time with the need to be activated to become available for further processing. Learning can be assessed in different ways, including learning outcomes, characteristics, and processes (Mayer, 2020b). Mayer describes retention tests and transfer tests as ways to assess learning outcomes. In a meta-review by Xu and colleagues (2022), different kinds of learning achievement that have been shown to be measured in studies on AR include achievement test scores, lab skills test scores, knowledge test scores, and spatial test scores.

Concerning AR, it has overall been suggested that it can support learning achievement. In a meta-analysis, Garzón and Acevedo (2019) found a summarised medium to large effect of using AR on learning gains in 64 studies (Cohen's $d = 0.68$) and in 2020, Garzón and colleagues found a similar summarised effect of AR on learning outcomes (Cohen's $d = 0.72$). In another meta-analysis of 38 studies, a large effect size (Hedge's $g = 0.92$) of a summary of all kinds of learning outcomes (e.g., performance, cognitive load, and emotion) of AR in K-12 education was found (Zhang et al., 2022). Through mixing the different kinds of variables, this cannot be attributed to learning achievement, but shows in general a difference between AR and non-AR when it comes to learning. In the field of science learning, Xu and colleagues (2022) found a total medium to large effect size of Hedge's $g = 0.74$ from 35 studies that measured academic achievement, with 22 of the comparisons showing a positive effect, 16 comparisons without a significant effect, and one comparison with a significant negative effect. In their meta-analysis of (quasi-)experimental studies, H.-Y. Chang and colleagues (2022) analysed the summarised effect found in 168 studies comparing AR to non-AR for different learning outcomes. They found a medium effect concerning subjective self-ratings of learners through questionnaires and surveys (Hedge's g = 0.49), a medium to large effect concerning assessed knowledge and skills (Hedge's g = 0.65), and a medium to large effect concerning assessed performance in authentic tasks (Hedge's g = 0.74). While this shows in general a positive effect of AR, the effect sizes have a very high variability, including a high number of studies in which non-AR leads to better results than AR, especially in the knowledge and skills category.

In summary, a generally positive effect of AR on learning outcomes is observed in empirical research. In the current dissertation, learning achievement is investigated in all seven studies (Krüger et al., 2019, 2022; Krüger & Bodemer, 2020, 2022a, subm.), including different learning topics and objectives (see Section 4 for an overview of the seven studies). While the generally positive outlook for implementing AR in education shows its potential for instructional settings, most research on AR does not examine more closely how and when exactly AR supports learning, due to the high number of media comparison studies (see Section 1.2). In the following, I will introduce multiple constructs that may be relevant to explain how learning is supported by AR, specifying constructs concerned with learning processes. Further looking into the effects of AR on these constructs through intra-medium (Surry & Ensminger, 2001) and value-added studies (Buchner & Kerres, 2023) can provide insights into the question of how AR supports learning, as described in Section 1.2.

2.3 Cognitive Load and Workload

One cluster of constructs that is relevant for learning with MERs and thus with AR is cognitive load and related constructs like workload, mental load, or mental demand. A central theory in the field of multimedia learning is the cognitive load theory (CLT), describing different effects of instructional design on learning explained by three different types of cognitive load. Although it is one of the most influential theories in instructional design, its tripartition of the types of cognitive load has only been applied in one of 64 studies in a systematic map of studies on cognitive load in AR-based learning environments until October 2019 (Buchner et al., 2021). Another central construct for AR-based learning experiences is subjective workload of tasks as defined in the NASA task load index (TLX) by Hart and Staveland (1988). This conceptualisation was used in 36 of 64 studies in the systematic map of AR studies (Buchner et al., 2021) and is thus the conceptualisation that is used most often in research on AR. In the following, CLT and NASA TLX will be introduced and research on AR in education including these concepts is explored.

CLT is one of the most commonly adopted theories for the design of multimedia learning material. It is based on the assumption that working memory capacity is limited and that learning material should be designed in a way as to not cause cognitive overload, specifically by reducing load caused by irrelevant activities that are not in accordance with the learning goals (Chandler & Sweller, 1991). This type of load is called extraneous cognitive load (ECL) and is assumed to fill up working memory capacities together with intrinsic cognitive load (ICL) and germane cognitive load (GCL) (Sweller et al., 1998). While ECL is thus defined as unnecessary load elicited by processing of poorly designed instructional materials, ICL is defined as cognitive load that is elicited by processing the content of the learning material in accordance with the learning objective, and GCL as cognitive load that is elicited by elements of the instructional design that evoke deeper processing and schema construction (Sweller et al., 1998). ECL and GCL can be influenced by the design of the learning material, which can lead to increased ECL when not designed well and can support GCL when it activates deeper processing. ICL is influenced by the content in interaction with learners' relevant prior knowledge. In Section 2.1, the counterparts of these three types of cognitive load in CTML were described: extraneous, essential, and generative processing, respectively. In an updated conceptualisation of CLT, GCL is not described as an individual part of cognitive load anymore, but as "the working memory resources that are devoted to dealing with ICL rather than ECL" (Sweller et al., 2019, p. 264). Based on this, one questionnaire that had been developed by Leppink and colleagues (2013) to measure the three types of cognitive load separately has later been shortened to two scales measuring ECL and ICL (Leppink et al., 2015). However, ICL has been shown to correlate with a passive component of load that is influenced by the learning material, while GCL correlated with an active component describing the effort that learners themselves put into the learning task (Klepsch & Seufert, 2021). This shows that a differentiation between these types of load might still be relevant for the instructional design and measurement of cognitive load. In this dissertation, I thus keep using the conceptual tripartition of the types of cognitive load. Klepsch and colleagues (2017) developed and validated a questionnaire with three subscales, measuring the complexity in content (ICL) and design (ECL), and the additionally invested cognitive processes (GCL).

The conception of the NASA Task Load Index (NASA TLX) by Hart and Staveland (1988) is another form of differentiating subconstructs of load. Six global constructs of workload are described in this model. Mental demand, physical demand, and temporal demand are three task-based constructs included here (Hart & Staveland, 1988). These types of demand are elicited by the difficulty of the task execution. Perceived performance and effort are behaviour-based constructs included in the NASA TLX conceptualisation, describing learners' subjective experience of their performance and how much effort they had to invest to achieve this performance. Frustration as the sixth component is person-based, describing a psychological impact of task demands and behaviour. In order to assess people's workload, the task load index is a scale that asks users to rate these concepts (Hart & Staveland, 1988) which has been applied in many contexts already (Hart, 2006). Using the NASA TLX conceptualisation to define task load in AR-based learning can deliver a more differentiated picture in addition to the CLT conceptualisation. When compared to CLT, the focus of the concepts includes more task-based constructs, which are considered in interaction with learner characteristics and behaviour. The CLT types are focused more on instructional design than on task execution, as the NASA TLX comes from a more general human-machine interaction background and not an educational background. Especially physical demand, which can be important in interactive learning settings, brings a new component in comparison to CLT which focuses on cognitive processes. Contrary to CLT, no differentiation is made between learning irrelevant and relevant mental demand elicited by the task in the NASA TLX. When comparing the behaviour-based constructs of effort and performance in the NASA TLX to CLT, the differentiation between passive and active components of cognitive processing can be helpful (Klepsch & Seufert, 2021). The active component describes the effort learners actively invest into cognitive processing. When more effort is invested, GCL can be increased, showing a connection between the two conceptualisations.

In a systematic review on studies measuring cognitive load in AR-based learning, more than half (56%) media comparison studies showed lower or equal cognitive load and higher performance for AR in comparison to non-AR (Buchner, Buntins, et al., 2022). Other studies also found lower cognitive load without measuring performance (6%) or found no differences in cognitive load and performance (17%). There were also a few studies with higher cognitive load or worse performance in AR. This shows that the results are mixed concerning empirical insights on cognitive load in AR. As already mentioned above, the tripartition of cognitive load based on CLT has not yet been researched comprehensively in AR-based learning environments (Buchner et al., 2021), so that these studies often do not differentiate between relevant and irrelevant cognitive load. Because AR is believed to and has been repeatedly shown to improve learning outcomes (see Section 2.2), it should be assumed that it does not only offer the possibility to decrease cognitive processing irrelevant for learning (i.e. ECL), but also to increase cognitive processing relevant for learning (i.e. GCL). One of the limitations of AR has been described as creating cognitive overload in learners as a result of the complexity of activities, challenges of integrating AR in educational settings, but also the technological limitations of AR at that point in time (Dunleavy & Dede, 2014). This shows potentially increased extraneous cognitive processing and load through AR. On the other hand, AR has been described to lead to better learning achievements, as described in Section 2.2, which suggests an increase in germane cognitive processing and load.

As already described above, many studies on educational AR examine cognitive load or task load, using various conceptualisations. The results of the studies are mixed concerning a positive or negative impact of AR. In a study by Wang and colleagues (2022), for example, a traditional biology lesson was compared with an AR implementation of a 3D model of the human respiratory system. Cognitive load was decreased in the group using AR in comparison to the non-AR group. However, the type of cognitive load measured in this study was not further defined, which shows one of the issues with research on cognitive load in AR. As described above, in AR it can be important to distinguish between cognitive load that supports learning, and cognitive load that can be detrimental. A study implementing the NASA TLX conceptualisation of workload showed that although no general difference in task load was found between a group learning about manual material handling in engineering education through HMD-based AR and a group learning in an in-class environment, different subconstructs of task load differed significantly (Guo & Kim, 2020). The authors found increased mental demand, effort, and frustration, and decreased perceived performance for the AR group. This shows the importance of examining different subconstructs of workload, and not just an overall score. Another distinction is often made between mental effort and mental load. When comparing a traditional physical, a virtual 3D, and a combined virtual-physical AR implementation of a lesson on magnetic fields, learners using the AR implementation had lower scores for mental effort and mental load than both of the other groups (Liu et al., 2021). The above-mentioned distinction between ECL, GCL, and ICL based on cognitive load theory is also increasingly used in research on AR. In a study comparing learning about the human brain with cross-sections or learning with an 3D AR implementation, both ECL and GCL were higher in the cross-section group (Henssen et al., 2020). In contrast to the above proposed positive effects of AR on cognitive load, Zumbach and colleagues (2022) expected ECL to be increased when learning with AR in comparison to paper-based instructions due to more potential distractions. They further expected the AR experience to be perceived as a game, decreasing deeper information processing and thus GCL. While they did not find significant difference for GCL, ECL was indeed significantly increased in the AR group. In another study comparing an AR-supported lesson on stereochemistry including 3D models of molecules to a lesson including only 2D drawings, cognitive load measured as ECL and ICL did not differ significantly (Elford et al., 2022). This shows that an implementation of AR does not necessarily lead to better results for cognitive load, so that it is necessary to take a closer look at how particular AR designs may have an influence.

When comparing different implementations or designs of AR in terms of cognitive load instead of applying media comparisons, more insights into effective design can be gained. In a study comparing a monoscopic and a stereoscopic presentation of a virtual anatomical 3D model of the lower extremities within an AR HMD, no influence on cognitive load measured through a summarised NASA TLX score was found (Bogomolova et al., 2023). This shows that not all design decisions necessarily have an influence on cognitive load. Concerning cognitive load influenced by the application of multimedia design principles in AR as mentioned in Section 2.1, an integrated design in accordance with the spatial contiguity principle has been found to significantly decrease ECL with a small effect but no effect on conceptual knowledge (Thees et al., 2020). In contrast, the same comparison in a similar setting was found to only descriptively but not significantly decrease ECL with a small effect size while, opposite to expectations, also significantly decreasing knowledge (Thees et al., 2022). Again, this shows the complexity and inconclusiveness of research on cognitive load in AR-based learning.

In summary, cognitive load and workload are important concepts when it comes to learning with AR. The research is very diverse, including a lot of different concepts and sometimes not all information about how cognitive load is operationalised. The exact mechanisms of how cognitive and workload are increased or decreased in AR are not fully known, yet. In the current dissertation, cognitive load including a split into the three types ECL, ICL, and GCL, is investigated in all studies, further filling the research gap on this specific conceptualisation of cognitive load. Some of the included studies provide a focus in their specific research questions and hypotheses. Study 1 focuses on ECL and ICL, Study 2 focuses on ECL (both Paper 1, Krüger et al., 2019). Study 3 (Paper 2, Krüger & Bodemer, 2020) and Study 6 and 7 (Paper 5, Krüger & Bodemer, 2022a) focus on ECL and GCL, Study 5 focuses on GCL (Paper 4, Krüger & Bodemer, subm.), and Study 4 (Paper 3, Krüger et al., 2022) includes all three types of cognitive load. In addition to this conceptualisation of cognitive load and for the possibility of comparing this research with other AR-based research, the NASA TLX conceptualisation of task load has also been applied in three studies in the current dissertation, namely Study 3 (Paper 2, Krüger & Bodemer, 2020), and Study 6 and 7 (Paper 5, Krüger & Bodemer, 2022a). See Section 4 for an overview of the seven studies.

2.4 Immersion

In addition to cognitive load, another important construct that needs to be considered in AR-based learning is immersion. Immersion has been described from a technological perspective and from a psychological perspective. From a technological perspective, Slater and Wilbur (1997) describe immersion as the degree of how well a technology can create an illusion of reality. For a high degree, the illusion should be inclusive, extensive, surrounding, and vivid. While Slater and Wilbur (1997) describe the psychological experience of this technological immersion as presence, Witmer & Singer (1998) use the term immersion to describe the psychological feeling of being enveloped by a certain environment. These are thus different definitions of the same word which currently coexist in the literature. A review of various definitions of immersion clusters these definitions into the three dimensions system immersion, narrative immersion, and challenge immersion (Nilsson et al., 2016). System immersion defines the term based on the system's properties, narrative immersion defines the term based on a person's response to a narrative, and challenge-based immersion defines it based on the response to a challenge. Due to the many different definitions, it is thus important to define how the term immersion is being used. In the current dissertation, immersion will be used in its psychological definition, as the subjective sense of being inside a certain environment. Still, immersive properties of technologies that can support the feeling of (system) immersion are also acknowledged. As to not exclude research in the literature review in this area just because different terms are used, research examining constructs described as (sense of) presence, engagement, and involvement, which have been described to be used interchangeably with the term immersion (Nilsson et al., 2016), will also be included. I will use the terms as they are in the cited research, under the assumption that these describe very similar or the same feelings as immersion.

VR, MR, XR, and AR as immersive technologies can embed learners in a technology-enriched context. Immersive learning is concerned with learning through the usage of artificial experiences that learners perceive as non-mediated (Dengel, 2022). The instructional design perspective views these experiences as tools to enrich learning, while the perspective on learners' internal processes describes the construction of internal models or knowledge. In models like the Cognitive Affective Model of Immersive Learning (CAMIL; Makransky & Petersen, 2021) and the Educational Framework for

Immersive Learning (EFiL; Dengel & Mägdefrau, 2018) psychological immersion or sense of presence are described as key aspects of the immersive learning experience. In the CAMIL, immersion is defined as a technological factor that has an influence on learners' presence as the feeling of being there (Makransky & Petersen, 2021). Presence in turn is described to have an influence on many different affective and cognitive factors, including cognitive load and motivation, which then are assumed to influence learning outcomes. In the EfiL, immersion is also defined as a description of the technology, which is part of the instructional affordances, has an impact on perception through presence, and influences the immersive learning potential, including motivational, cognitive, and emotional factors (Dengel & Mägdefrau, 2018). Presence in turn influences learning activities, while also having a two-sided relationship with the immersive learning potential.

As also described in these models, research on immersive learning environments often shows positive effects of technological system immersion on presence and thus the feeling of immersion, enjoyment, and motivation (Makransky, 2021). Makransky describes two paths through which immersion can influence learning, the affective and the cognitive path. The affective path describes how immersive experiences can increase enjoyment, which can help learners focus on the task and may increase intrinsic motivation and generative processing. The cognitive path describes the potential of immersion to either decrease ECL through the removal of distractions from the environment, but also the potential to increase ECL due to the potential complexity of the environment distracting from the learning task (Makransky, 2021). Empirical results on learning outcomes are mixed showing both positive and negative relations between immersion and learning, especially when it comes to declarative knowledge instead of procedural knowledge, spatial knowledge, or behavioural transfer. Based on these insights, the immersion principle has been introduced recently as a multimedia design principle in the CTML. The principle describes that immersive environments do not necessarily promote learning but can lead to distractions and cognitive overload due to the many displayed elements that are not directly related to the content of the learning material (Mayer, 2020e). While the described models CAMIL and EfiL and the immersion principle focus on immersive VR, they might be at least partly transferable to AR as another immersive medium, although there are AR-specific aspects to consider.

To define immersion more specifically for AR, Kim (2013) developed a framework concerning context immersion in mobile AR environments. The concept of context immersion is described as being immersed and thus enveloped through awareness of the surrounding context information. It includes several features, for example, an autonomous experience, location tracking, and an embodied space. The context can be a time and location-based context, object-based context, and user-based context. For the case of location-based AR, Georgiou and Kyza (2017b) describe three levels of immersion that build upon each other: 1) engagement, including the learners' interest for the activity, time investment in the activity, and usability of the application; 2) engrossment, including learners' emotional attachment to and focus of attention during the activity; and 3) total immersion, including presence, which is described as the feeling of being surrounded by the environment, and flow, the learners' full absorption into the activity. In this conceptualisation, (sense of) presence is thus a part of the experience of total immersion, together with flow. Based on these three levels of immersion, the augmented reality immersion (ARI) questionnaire has been developed and evaluated, including six sub-scales for the different sub-components of immersion (Georgiou & Kyza, 2017c). Salar et al. (2020) tested a comprehensive model to show the relations between the six sub-constructs. They found an influence of usability on interest, an influence of emotional attachment on interest and focus of attention, an influence of focus of attention on flow, and an influence of presence on flow and focus of attention in a cohort of university students using AR in science learning.

Research on immersion in AR-based education shows that learners' immersion can have a positive effect on learning. For example, learners with higher immersive profiles including the six ARI subconstructs had better learning outcomes concerning basic chemistry concepts than learners with lower immersive profiles (Uriarte-Portillo et al., 2022). Furthermore, in location-based AR environments, learners with high levels of immersion displayed different learning behaviours (Georgiou & Kyza, 2017b) and had higher learning outcomes (Georgiou & Kyza, 2017a, 2018). Y.-H. Chen and Wang (2018) distinguish between the effect of sense of presence on learning achievement in low-presence and high-presence learners. The results suggest that for low-presence learners, presence has an influence on learning achievement, but not for high-presence learners.

Supporting an assembly task with a handbook, a screen-based AR visualisation, or a HMD-based AR visualisation, different levels of immersion of learners were found (Generosi et al., 2022). Immersion was perceived lower with the handbook than the two types of AR visualisations. In some studies, the influence of AR on specific sub-constructs of immersion has been examined. In a study focusing on the engagement level of the ARI subconstructs (i.e. interest and usability), AR on handheld devices, MR on HMDs, VR on HMDs, and paper-printed learning material was compared (Zhao et al., 2023). Descriptively, the AR and MR groups had the highest usability scores, and the MR group had the highest interest score by far. This shows that AR and MR, here distinguished based on the device used, may provide a solid basis through engagement as the lowest level of immersion. One of the subfactors of total immersion in the ARI conceptualisation is flow. In the study by Wang and colleagues (2022) comparing a traditional biology lesson to an AR implementation of a 3D model of the human respiratory system that was also mentioned in Section 2.3 on cognitive load, flow experience was increased in the AR group in comparison to the non-AR group. In their study also mentioned in regard to cognitive load, Zumbach and colleagues (2022) found higher experienced immersion of learners learning with AR in comparison to paper-based instructions, but no differences in experience of flow.

In summary, immersion is an important concept when it comes to learning with AR. There is still a lot of research necessary when looking at the exact mechanisms of how the perception of immersion is influenced by different designs of AR. In the current dissertation, immersion is only investigated in one study, namely in Study 5 (Paper 4, Krüger & Bodemer, subm.) in its conceptualisation including the subconstructs interest, usability, emotional attachment, focus of attention, presence, and flow.

2.5 Motivation

Motivation has been defined as the drive to execute specific actions, based on people's beliefs, values, and goals (Eccles & Wigfield, 2002). There are a lot of different frameworks defining different forms of motivation and motivationally relevant variables for learning and in general in educational settings. Eccles and Wigfield (2002) have clustered and defined theories that focus on people's expectancies concerning their competence and control, theories that focus on the specific reasons for executing tasks, theories that integrate people's expectancy and value, and theories that integrate motivation and cognition. Wigfield and Eccles (2000) describe their own expectancy-value theory of achievement motivation, in which ability beliefs, task expectancies, and subjective values are related to each other. Ability beliefs describe people's perceived competence to execute an activity, task expectancies describe their expected success at a task, and subjective values describe perceived usefulness, importance, and interest of the task execution. Based on these concepts, they also developed a questionnaire with three subscales (Wigfield, 1994; Wigfield & Eccles, 2000).

A common distinction of motivation is made between intrinsic and extrinsic motivation. While intrinsic motivation is described as motivation that is innate to the person and depends on their interest in the activity itself, extrinsic motivation comes from external rewards and punishments (Deci & Ryan, 1985). These are further differentiated through levels on a continuum of self-determination, reaching from intrinsic motivation, to identified regulation, to external regulation, to amotivation. These levels have also been picked up by Guay and colleagues (2001) to develop the situational motivation scale (SIMS). This scale has subscales to measure intrinsic motivation, identified regulation, external regulation, and amotivation separately. Another questionnaire to measure intrinsic motivation is the intrinsic motivation inventory (IMI) first used in studies by Ryan and colleagues (Plant & Ryan, 1985; Ryan, 1982; Ryan et al., 1983) which measures constructs that are relevant for motivation, including subscales on interest/enjoyment (which is defined as the subscale to measure intrinsic motivation) and the related constructs perceived competence, perceived choice, effort, and pressure/tension.

Because AR-based educational settings are designed by instructors with a specific purpose in mind, another relevant framework that can be applied in AR-based research is the ARCS framework, which describes attention, relevance, confidence, and satisfaction as four levels of motivation that can and should be influenced by the design of the learning material (Keller, 2010). Attention needs to be evoked before learning can start. It includes the three sub-components perceptual arousal, inquiry arousal, and variability. Relevance is the next level that needs to be evoked to maintain motivation. Again, three sub-components are described: goal orientation, motive matching, and familiarity. Confidence is the next level in the ARCS model with the three sub-components learning requirements, success opportunities, and personal control. The final level in the model is learners' satisfaction. It includes the sub-components intrinsic reinforcement, extrinsic rewards, and equity (Keller, 2010). The goal of the ARCS model is to give instructional designers a quick overview of major dimensions of learning motivation and support them in stimulating and sustaining motivation through different strategies. Based on the model, a questionnaire with the four subscales attention, relevance, confidence, and satisfaction has been developed to examine learners' motivational experiences (Loorbach et al., 2015). The ARCS model can be and has been applied to AR-based learning experiences. For example, Chia-Chen and colleagues (2022) used it to design the AR-based application Cosmos Planet Go, and Wei and colleagues (2015) designed a complete ARCS-based teaching scheme for a course in which students learned to design an AR environment themselves.

In the AR-related area of multimedia learning, a motivational component has been added to CTML as part of the cognitive-affective theory of learning with media (CATLM) by Moreno (2006). The assumption of affective mediation added to the model proposes that motivational factors regulate cognitive processing and affect, and thus mediate learning. This includes the assumption that motivation determines how much of learners' cognitive resources are actually invested in a learning task (Moreno, 2010). In her discussion of CLT from this cognitive-affective perspective, Moreno states that it is important to consider a relation between load, affect, and motivation because it is not the cognitive capacity but the actually invested cognitive resources that affect learning. Paas and colleagues (2005) also argue that motivation or learner involvement predicts mental effort, which in turn predicts performance, and Mayer (2014) similarly states that affective instructional design features can influence how much learners engage in cognitive processing. In a different conceptualisation, motivation has been described as an outcome of cognitive load instead of a parallel process (Feldon et al., 2019). Here, ECL is described as a cost that has an influence on motivational beliefs. R. E. Clark and colleagues (2006) further describe that there are specific motivational challenges in complex learning environments. These challenges include an automatic attention shift to irrelevant stimuli and increased negative emotions as a reaction to cognitive overload, and a decrease in mental effort when self-efficacy is inappropriately high or low. The connection between motivation, effort, and learning outcome describes the importance of considering motivation in learning environments, especially when the environments are complex and could discourage learners. AR has the potential to be leveraged for the design of complex environments, which shows how important motivational aspects might be for AR-based learning environments.

Motivation plays a role in both the CAMIL (Makransky & Petersen, 2021) and the EFiL (Dengel & Mägdefrau, 2018) models on immersive learning also mentioned in regard to immersion in Section 2.4. In the CAMIL, intrinsic motivation is being described as a factor that is influenced by the immersive affordances presence and agency and in turn influences different types of knowledge (Makransky & Petersen, 2021). In the EfiL, motivational factors are described as immersive learning potential, influenced by instruction affordances and influencing learning activities which in turn influence learning outcomes, amongst others (Dengel & Mägdefrau, 2018). When looking at literature on AR in education in particular, motivation appears often as a variable in research. In many systematic reviews, an increase in motivation is described as one of the advantages of AR (e.g., Akçayır & Akçayır, 2017; Bacca et al., 2014; P. Chen et al., 2017; Garzón et al., 2019; Herpich et al., 2019; Radu, 2012, 2014). In a review focusing specifically on motivation and performance in AR in secondary education, 83% of the 13 included investigations showed a positive effect, specifically on the four levels of motivation described in the ARCS model (Amores-Valencia et al., 2022).

Many studies on AR in education have focused on the design of motivational applications. As already described above, the ARCS model is often referenced in this regard. The application Cosmos Planet Go by Chia-Chen and colleagues (2022), for example, was designed on the basis of the ARCS model. The authors identified relevant teaching strategies that could be implemented for each of the factors attention, relevance, confidence, and satisfaction. They compared the strategies applied in a control group without AR, and an experimental group using an AR-based implementation of the lesson plan, and evaluated those with the IMMS questionnaire mentioned above. The authors found a higher score for all four factors of the ARCS model in the AR-based lesson in comparison to the traditional lesson and better learning outcomes on a post-test. This shows that the implementation of motivational design features in AR can improve motivational experience and learning outcomes. However, because design features for all four categories of motivational aspects are implemented in the AR version, it is not completely clear through which mechanisms motivation was increased. In another study the focus was not on implementing motivational features but on comparing an AR-based and a conventional multimedia science learning course (A.-F. Lai et al., 2019). Learners were found to be more motivated in AR, showing higher scores in attention, relevance, confidence, and satisfaction. This shows that even when no specific design features for motivation are implemented, AR can improve motivation compared to a traditional approach.

Besides the ARCS model, other constructs of learning motivation have been examined in educational AR. In the study by Wang and colleagues (2022) on a traditional biology lesson in comparison to an AR implementation of a 3D model of the human respiratory system already mentioned in regards to cognitive load outcomes and flow experience in earlier sections, learning motivation was increased in the AR group in comparison to the non-AR group. Although a questionnaire was used that defines different subconstructs of motivation, the measurement is not further defined, so that the exact motivational effects cannot be distinguished. Self-efficacy, as a motivational construct, did not differ significantly between the two groups in the study. In a study comparing an AR application and a simulation for learning about insects in a natural science course, learning approaches, including strategies and motives, were assessed (Yang & Tsai, 2020). The authors found higher deep motivation in the AR group, which describes that learners had higher intention to make sense of and understand the materials when learning with AR than with a simulation. This shows that AR cannot only increase motivation in comparison to traditional lessons, but also in comparison to other multimedia material.

In summary, motivation is an important concept when it comes to learning with AR. The research points into the direction that learning motivation can in general be increased through AR, although the exact mechanisms are not completely clear. In the current dissertation, the motivational impact of AR has been inspected in different study designs. Motivational aspects of participants before the learning phase of the study, following an expectancy-value conceptualisation, are investigated and used for sample descriptions in Studies 3 (Paper 2, Krüger & Bodemer, 2020), Study 4 (Paper 3, Krüger et al., 2022), Study 5 (Paper 4, Krüger & Bodemer, subm.), and Study 6 and 7 (Paper 5, Krüger & Bodemer, 2022a). Motivational aspects in association with using the respective AR application are investigated in Study 1 (Paper 1, Krüger et al., 2019) and Study 5 (Paper 4, Krüger & Bodemer, subm.), although with different conceptualisations. In Study 1, the situational intrinsic motivation conceptualisation is used, including subfactors intrinsic motivation, identified regulation, external regulation, and amotivation. In Study 5, the ARCS conceptualisation is used, including the subfactors attention, relevance, confidence, and satisfaction. See Section 4 for an overview of the seven studies.

2.6 Spatial Abilities

Another concept that is relevant in the context of AR-based learning experience is the learners' spatial abilities, especially in the context of 3D visualisations. The alternative forms of studies suggested by Surry & Ensminger (2001) and Buchner and Kerres (2023) described in Section 1.2 include aptitude-treatment-interaction or respectively learner-treatment interaction studies. Spatial abilities are one category of learner characteristics that can play a role when learning with AR and examining these in relation to learning processes and outcomes may help answer the question of when using AR can be effective. Cheng and Tsai (2013) also describe the necessity to investigate learner characteristics like spatial abilities for research on AR in education, as it might interfere with learners' experiences, learning processes or outcomes.

Spatial abilities have been classified in the category of abilities of visual perception, including abilities that focus on how individuals deal with spatial presentation and orientation (Carroll, 1993). Further, dimensionality has been identified as an inherently relevant aspect in this, but the spatial distribution of elements in the physical or imaginal space has been identified as a main concern. Carroll (1993) finally defines spatial abilities as abilities that have to do with searching the visual field and mentally representing the perceived objects including their spatial characteristics (e.g., forms, shapes, and positions), and with mentally manipulating these mental representations. Two factors have been identified as defining clusters of spatial abilities, the spatial relations and the spatial visualisation factor (Pellegrino et al., 1984). While the spatial relations factor describes abilities for rapid mental rotation and transformation processes, the spatial visualisation factor describes abilities for more complex mental visualisation and manipulation within the parts of an object without the goal of a quick solution. Tests for spatial relations abilities thus often have a time limit but are less complex, while tests for spatial visualisation abilities are more complex but have no time limit.

3D mental rotation abilities are one category of spatial abilities within the spatial relations factor. Individuals with high mental rotation abilities are quick in determining if a 3D object is a rotated version of another 3D object or a completely different object (Pellegrino et al., 1984). Shepard and Metzler (1971) found that the time to determine this increased when the angular difference in the orientation of the 3D objects is higher, which suggests that individuals mentally rotate the object until it has the same orientation. This was also supported by what the participants of the study told in interviews about how they proceeded. One form of measuring mental rotation abilities is the Mental Rotation Test (MRT) by Vandenberg & Kuse (1978). In this test, the 2D drawings of four 3D figures must be compared to a drawing of a reference figure. Two of the four options are rotated versions of the reference figure, which must be identified as quickly as possible because there is a time limit for answering all items. A redrawn version of this MRT has been published by Peters and colleagues (1995).

When learning with 3D representations, learners' 3D spatial abilities including mental rotation abilities play an important role. In the literature, two types of hypotheses have been proposed, an ability-as-compensator and an ability-as-enhancer hypothesis. The ability-as-compensator hypothesis, as suggested by Höffler (2010), describes that learners with lower spatial abilities profit from 3D representations while learners with higher spatial abilities do not need those representations, because they can handle the mental transformation from 2D representations to 3D. The ability-as-enhancer hypothesis, as suggested by Huk (2006), describes that learners with higher spatial abilities profit from 3D representations while learners with lower spatial abilities do not have the skills to handle those to their advantage. As described in Section 2.1, AR has the potential to display 3D representations of virtual objects, so that spatial abilities need to be considered in this case. However, it is not fully clear if the ability-as-compensator or the ability-as-enhancer hypothesis is appropriate in this case.

In a study by Ho and colleagues (2022), an AR HMD application displaying a 3D model of the inner regions of the human brain was compared to a physical model of the same. Mental rotation abilities were found to be positively associated with test performance, but only when using the AR HMD application and not when using the physical model. Mental rotation abilities thus may play a role when learning with AR models but not with physical models. In the study comparing a monoscopic and a stereoscopic presentation within an AR HMD displaying a virtual anatomical 3D model of the lower extremities which was already mentioned in regards to cognitive load, no influence of mental rotation abilities was found on the relation between the form of presentation and cognitive load or knowledge, although the abilities were in general positively correlated with knowledge about structure names and functions tested at plastinated specimen (Bogomolova et al., 2023). The additional depth cues provided by the stereoscopic presentation did not interact with learners' mental rotation abilities, but in general the mental rotation abilities seem to have predicted if learners could transfer their knowledge onto authentic physical 3D representations.

In summary, spatial abilities, including mental rotation abilities, can be influential when it comes to learning with AR, especially with spatial learning as an objective. However, the literature is not yet fully clear on the exact role and mechanisms of how learners' spatial abilities influence learning in AR, so that more research is necessary. In the current dissertation, mental rotation abilities of participants before the learning phase of the study are investigated and included as a moderator variable in Study 4 (Paper 3, Krüger et al., 2022).

An addition to spatial abilities as learner characteristics with a potential influence on learning with AR-based representations, AR has also been used for training of spatial abilities. In a review of 32 studies from 2010 to 2019, most studies show an enhancement of spatial abilities through AR-based training (Papakostas et al., 2021). The authors further described that the Purdue Spatial Visualisation Test for Rotation (PSVT:R), the Differential Aptitude Test for space relations (DAT:SR) and the Mental Rotation Test (MRT) were most often used to test spatial skills. Due to the nature of the studies presented in the current dissertation, in which skills training is not the focus, spatial abilities training is not discussed further. In Study 1 (Paper 1, Krüger et al., 2019), mental rotation abilities have been investigated as an outcome variable after the learning phase, but treated as a by-product of learning with 3D models and without a specific focus on skills training.

2.7 Summary

In total, there are many constructs that play a role when learning with AR. Here, I presented learning with MERs, learning achievement as an outcome variable, cognitive load, workload, immersion, and motivation as process-related variables, and spatial abilities as learner characteristics.

The described constructs often cannot be considered separate from each other. When looking at the CTML by Mayer (2020a) as a theory on MERs, for example, cognitive load considerations are a central part of the theory. The CATLM by Moreno (2006) as an extension of the theory further includes motivational aspects and individual differences (e.g., spatial abilities) as moderator variables. Motivation is here described as a factor that mainly has an influence on the actually invested mental effort, and thus the active component of cognitive load (see Klepsch & Seufert, 2021). Looking at the models described concerning immersion and motivation in immersive learning, CAMIL (Makransky & Petersen, 2021) and EFiL (Dengel & Mägdefrau, 2018), it can further be seen, that the different outcome and process variables are related to each other. The CAMIL proposes that the feeling of presence or immersion has an influence on affective and cognitive factors, including motivation and cognitive load, which in turn are described to influence learning outcomes like different types of knowledge. In the EFiL, motivational, cognitive, and emotional factors as part of the immersive learning potential are influence by instruction affordances and in turn have an influence on learning behaviours and outcomes. The affective path of immersion described by Makransky (2021) is described to work through an increase in enjoyment, which may increase intrinsic motivation, focus learners’ attention, and lead to more generative processing, which can in turn lead to better learning outcomes. This shows that the constructs might need to be considered in their interplay as well as their individual mechanisms.

Although there has been research concerning the mentioned learning processes, learning outcomes, and learner characteristics, the particular mechanisms through which they are influenced by or related to specific design decisions in AR-based educational environments are often not yet fully known. For this, more systematic research is necessary, which leads to the necessity of first defining attributes that play a specific role in AR, as suggested by Surry and Ensminger (2001; see Section 1.2).

3 The ARcis Characteristics

It is obvious that AR has some characteristics that other technologies do not have, which is mainly due to the combination of virtual and physical elements. While the virtual elements in AR can be purposefully designed and placed by instructors based on their specific learning goals, physical elements in the real world are by definition authentic, realistic, detailed, and rich in information but not as manipulatable. When learning takes place inside the physical world and includes physical elements in the learning material, more senses than just learners' visual sense are automatically involved because there are surrounding sounds, the possibility to touch and feel, and even smells and tastes. In many different (meta-)reviews of literature on AR in education, various affordances of AR have been defined and examples from six papers over the last ten years summarising multiple aspects can be seen in Table 2. Affordances that are mentioned multiple times are, for example, visualisation of the invisible (Cheng & Tsai, 2013; MacCallum & Jamieson, 2017; Parsons & MacCallum, 2021; Wu et al., 2013), learning with and about spatial representations and concepts (Bower et al., 2014; MacCallum & Jamieson, 2017; Wu et al., 2013), situated or contextualised learning (Bower et al., 2014; Cheng & Tsai, 2013; Dunleavy & Dede, 2014; MacCallum & Jamieson, 2017; Parsons & MacCallum, 2021) and practical skill development (Cheng & Tsai, 2013; Parsons & MacCallum, 2021). Although this list of affordances provides a good overview of what AR can be used for, they are formulated on different levels, including (technological) features of AR software (e.g., visualising the invisible) and hardware (e.g., portability of devices), pedagogical or didactical concepts (e.g., situated learning, collaborative learning), and proposed learning outcomes (e.g., conceptual understanding, practical skills). Furthermore, not all of these affordances can be solely attributed to AR-based learning environments, although the specific way in which AR affords most of them is probably unique.

To work towards the achievement of the theoretical Subgoal 1 of the current dissertation, which aims at theoretically defining characteristics of learning with AR and analysing how specific mechanisms may have an impact on learning, I will present three characteristics of AR. These characteristics are supposed to summarise the currently proposed affordances of AR on a more abstract level and from a psychological, learner-focused perspective based on the system characteristics proposed by Azuma (1997). For this purpose, we developed the ARcis framework including the three characteristics contextuality (c), interactivity (i), and spatiality (s) in Paper 1 (Krüger et al., 2019). The goal of the definition of these AR-specific characteristics is to provide a basis for systematic research on and design of purposeful AR-based learning settings. In the following sections, the characteristics and proposed relations with learning processes and outcomes will be described in more detail.

Table 2

Examples of Different AR Affordances Described in the Literature

Paper	AR affordances / characteristics
Cheng and Tsai (2013)	▪ enhance spatial abilities ▪ enrich practical skills ▪ support conceptual understanding and change ▪ enable collaborative inquiry-based activities within physical environments
Wu and colleagues (2013)	▪ provide 3D-object based, interactive learning ▪ enable ubiquitous, collaborative and situated learning with virtual elements in real environments ▪ offer sense of presence, immediacy, and immersion ▪ visualise invisible concepts or events ▪ bridge gap between informal and formal learning
Bower and colleagues (2014)	▪ rescale virtual objects with clear representation of spatial concepts and contextualisation in real-world environment ▪ overlay contextually relevant information as "perfectly situated scaffolding"
Dunleavy and Dede (2014)	▪ present multiple perspectives ▪ leverage physical space ▪ access external resources ▪ support student motivation ▪ allow situated learning and transfer ▪ leverage context sensitivity
MacCallum and Jamieson (2017)	▪ visualise 3D and the invisible ▪ contextualise information ▪ provide portability of devices to interact with location ▪ enable social and shared engagement
Parsons and MacCallum (2021)	▪ reduce negative impact like risks and costs ▪ visualise the invisible ▪ support development of practical skills in a spatial context ▪ provide portability of devices across locations ▪ enable situated learning in context

In addition to Subgoal 1, the following sections will also provide first insights towards the achievement of Subgoal 3 of the current dissertation. The practical Subgoal 3 is focused on practically applying the theoretical and empirical insights into the characteristics and mechanisms for the design of AR-based learning experiences. For the purpose of applying the theoretical insights of the ARcis framework, the three educational AR applications created for and used in the main Studies 3, 4, and 5 will be analysed concerning the implementation of these characteristics, showing a range of possible applications. The application "powAR" used in Study 3 (see Section *IV.C. Materials and Apparatus* in Paper 2, Krüger & Bodemer, 2020 for a detailed description of the application) includes pattern-based AR markers on paper cards that can be scanned to display virtual, animated 3D models of components of combined-cycle power plants (see Figure 1). Concerning the physical interaction with the learning material, learners can either receive already constructed marker clusters of different complete power plants or must construct them themselves. Furthermore, concerning the mental interaction with the learning material learners either must figure out for themselves which power plant compositions to compare to answer hypotheses about power output and efficiency or this information is provided as part of the instructions. The application “heARt” used in Study 4 (see Section *2.3. Material and apparatus* in Paper 3, Krüger et al., 2022 for a detailed description of the application), includes an image-based marker as part of a worksheet with textual and pictorial information about the human heart. This marker can be scanned with the application, either showing a virtual version of the 2D graphic with labels for the names of the components of the human heart or showing a labelled virtual 3D model of a human heart (see Figure 2). In the application “ARbor” used in Study 5 (see Section *2.2. Materials* in Paper 4, Krüger & Bodemer, subm. for a detailed description of the application), image-based markers can be scanned to show additional information about plants. Learners can either receive these markers and thus the information directly at the respective plants or further away but still in the general vicinity of the plants (see Figure 3). In the following sections, the description of the three characteristics will form a basis for the analysis of these three applications.

Figure 1

Usage of powAR Application in Study 3: Scanning Markers To View Power Plant

Figure 2

Usage of heARt Application in Study 4: 3D and 2D Visualisation of Human Heart

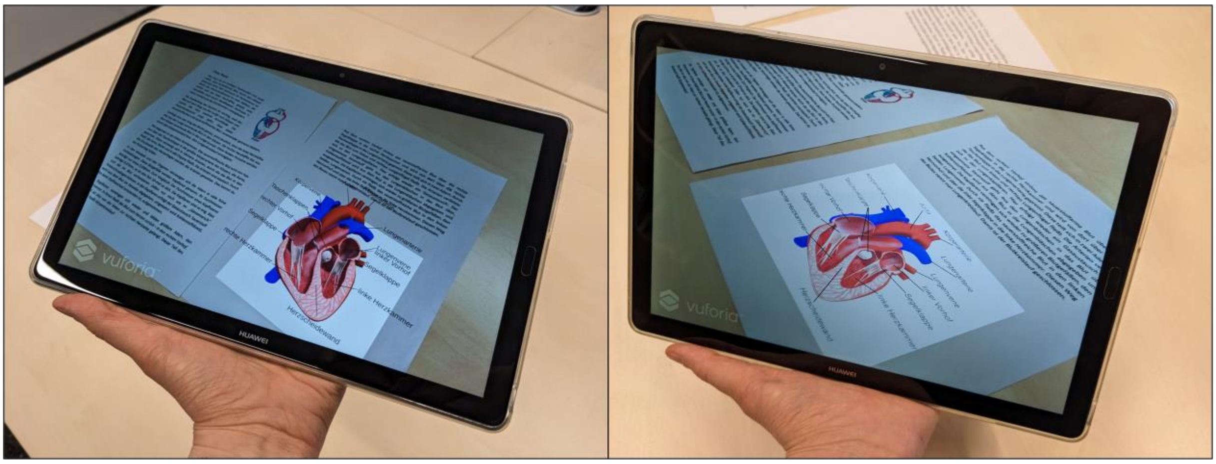
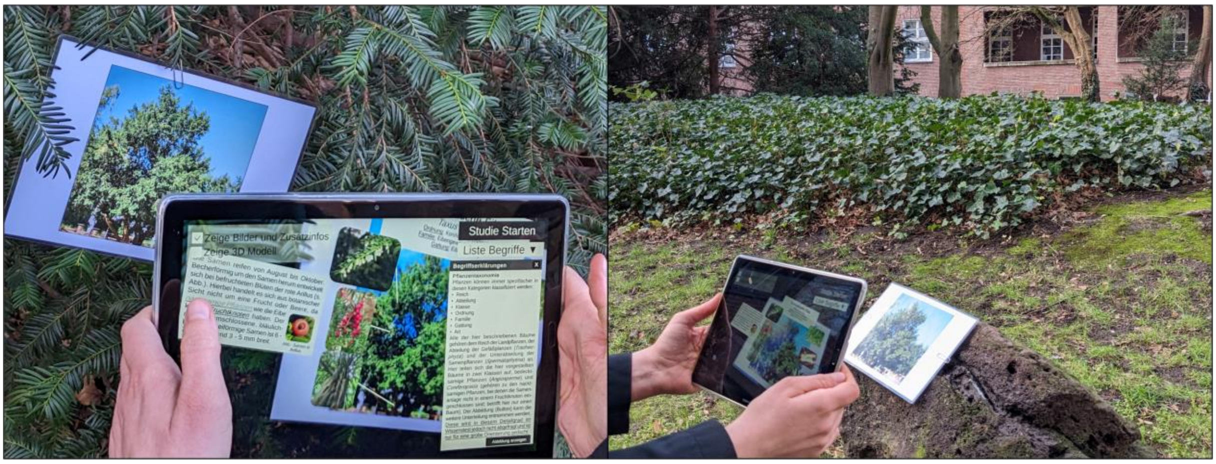


Figure 3

Usage of ARbor Application in Study 5: Accessing Information About Near or Far Plants



3.1 Contextuality

Contextuality describes the possibility of a combined presentation of virtual and physical elements in AR, which enables learners to perceive representations from both worlds in an integrated form (see Section *2.1 Contextuality* in Paper 1, Krüger et al., 2019). This characteristic is based on the first characteristic of AR systems described by Azuma (1997): combination of real and virtual elements. Since the virtual elements do not cover the context (as in VR, for example), the context can be perceived in parallel. The realities are also not separated from each other (as is the case, for example, with normal tablet use), but can be displayed in an integrated manner. For educational settings, this includes the possibility to place virtual, instructor-designed elements within a physical world setting, so that relevant instructional information can be placed exactly where it is needed, and learning can be situated within a relevant environment. In the above-mentioned affordances, this aspect is best reflected in the contextualisation of information (Bower et al., 2014; MacCallum & Jamieson, 2017), provision of situated learning opportunities (Dunleavy & Dede, 2014; Parsons & MacCallum, 2021; Wu et al., 2013) and the "perfectly situated scaffolding" (Bower et al., 2014, p. 6). Thus, the focus here is on features of the physical world in which AR-based learning takes place and its thematic connection with virtual elements.

3.1.1 Contextuality in AR-based Learning

Contextuality can be leveraged in the design of AR-based educational experiences that support learning (see Section *2.1 Contextuality* in Paper 1, Krüger et al., 2019, for basic and Paper 4, Krüger & Bodemer, subm., for more detailed information on this). Through the parallel perception of virtual elements and the physical context, the placement of instructional information and the connection of virtual and physical elements play an important role in AR. Through this and the use of mobile technologies, learning can be embedded within relevant environments in the physical world, which is in accordance with one condition to improve learning based on constructivist learning theory (Dunleavy & Dede, 2014). Through this mobility and the possibility of location-based learning, the feeling of authenticity can be supported within learners, learners can be contextualised in a location and grounded in reality (Wu et al., 2013). AR applications have been characterised as context-aware or context-independent (Wen & Looi, 2019) and place-dependent or place-independent (Dunleavy & Dede, 2014), describing different levels of relevance of the physical context for the learning experience. Reid and colleagues (2005) describe three levels of meaningfulness of the surrounding place in environments where virtual media files are placed at physical locations: (1) arbitrary linkage, where any place with enough space can be used; (2) physicality, where the atmosphere and general characteristics of the environment are part of the experience, and (3) particular location, where the exact location with its elements is important. Wetzel and colleagues (2011) similarly classify mobile AR games based on their semantical location context on a spectrum from (1) independent, to (2) loosely coupled, to (3) dependent. The differentiation of these levels allows for a specific description and implementation of AR-based learning experiences.

Bower and colleagues (2014) describe that AR can bring the real world into an instructional setting like a classroom and that instructional elements can be embedded in physical environments, leading to contextualised and authentic learning characterised within the theory of situated learning. Situated learning has been highlighted as a theoretical framework relevant to AR in many papers (e.g., Bower et al., 2014; Cheng & Tsai, 2013; Dunleavy & Dede, 2014; Garzón et al., 2020; Sommerauer & Müller, 2018; Wu et al., 2013). Situated learning describes that learning is inherently linked to the context in which it takes place, making learning in an authentic context desirable (Brown et al., 1989). Young (1993) further describes that learning should be contextualised within real environments. By embedding learning in a relevant physical context, AR-based contextuality can support the mental connection and integration of virtual elements, physical objects, and the physical environment, thus promoting learning that is characterised by its combination of physical and virtual elements.

Using AR can immerse learners within a physical real-world context, which has been described as context immersion (Kim, 2013). Georgiou and Kyza (2017) define immersion in location-based AR on different levels, with the highest levels including flow and presence (see Section 2.4). Immersion has furthermore been connected to learners' motivation to learn. Makransky and Petersen (2021), for example, predict in their Cognitive Affective Model of Immersive Learning (CAMIL) that sense of presence, which describes the feeling of immersion, has an influence on motivational factors and in turn on learning outcomes. Additionally, the placement of virtual information within a corresponding physical environment can show the relevance of this information, which Keller (2010) describes as an important aspect for increasing motivation in the ARCS model (see Section 2.5). For a more detailed discussion of the connection of contextuality to immersion and motivation, see Section *1.1 Immersion and motivation in contextualized AR learning environments* in Paper 4, Krüger and Bodemer (subm.). Placing virtual information within a corresponding physical environment might thus have positive effects on feelings of immersion and motivation.

3.1.2 Recent Research on Contextuality

In recent research in the field of AR-based learning experiences, it has been shown that situating the experience in a relevant context can have positive effects on learning processes and outcomes. In a recent study, AR was used to situate cultural heritage education at a specific cultural location, the Basilica of Saint Catherine of Alexandria in Galatina (De Paolis et al., 2022). The authors describe the capability of AR to provide contextually relevant information at specific locations, which they implemented in an application including a storytelling-based audio guide and visual AR elements. However, they only examined learners' perceived usability, user experience, and task load without a control group, so advantages of the contextualisation and effects on learning outcomes are not clear.

In a study that focused on the feelings of immersion and learning outcomes, Georgiou and Kyza (2021) examined the use of strong semantic coupling between a narrative about a mysterious disease outbreak and the surrounding physical environment in AR in a problem-based learning setting. This was achieved by placing QR codes for retrieving information and additional physical cues at places of interest instead of at random places. Students in the strongly coupled experience reported higher presence and total immersion and had better learning results concerning their reasoning about relevant topics, but not their factual knowledge. Furthermore, students' learning scores were positively related to their immersive experiences when they learned with the strongly coupled set-up, showing a relation between these constructs. Contextual coupling thus seems to have an influence on feelings of immersion and reasoning as a specific type of learning, which furthermore seem to be related.

In another recent study, contextualised vocabulary learning was implemented in AR, displaying keywords for physical objects embedded into the physical space (Weerasinghe et al., 2022). A head-mounted display was used for an integrated real-time visualisation and compared to a tablet-based implementation with a photo instead of a real-time view of the environment as the physical context. Immediate but not delayed recall was improved by the contextualised AR visualisation, task load was decreased, and motivation increased. This shows the potential positive influences on motivational and cognitive processes when leveraging the characteristic of contextuality in AR.

Comparing the use of physical objects and AR marker cards, another study was executed in the context of education on behaviour during a fire outbreak (Huang et al., 2022). Learners either looked for and scanned relevant physical objects (e.g., fire extinguisher) or marker cards to receive information and instructions about a fireground situation. Learning with the physical objects led to a bigger increase in knowledge and lower mental load than learning with the marker cards. This shows the positive influence of using authentic physical objects on cognitive processes. A study that has been specifically designed with the contextuality characteristic in mind, focusing on the closeness of positioning virtual information in relation to corresponding physical objects, is Study 5 in this dissertation (Paper 4, Krüger & Bodemer, subm.). It will be described in more detail in Section 4.4.1.

3.1.3 Application of Contextuality

When looking at contextuality in AR more closely, two levels of physical elements can be distinguished (see Section *1 Spatial integration in AR learning environments* in Paper 4, Krüger & Bodemer, subm.). On one hand, AR-based learning is contextualised in the physical environment in general. On the other hand, the virtual elements are anchored to specific physical objects (e.g., surfaces, AR markers, or natural objects). Contextuality in AR can differ in the thematic relevance of the physical environment or physical anchors for the virtual elements. It can also differ in the perceivability of the context, where virtual elements might overlay and thus cover physical elements in a way that they are not visible anymore. To clarify the potential differences, I will in the following analyse the three introduced AR applications used in studies included in this dissertation concerning these differences in thematic relevance and context visibility.

In the powAR application used in Study 3 (Paper 2, Krüger & Bodemer, 2020) and shown in Figure 1, different factors influence its contextuality. The physical anchor elements are pattern-based AR markers that show black and white patterns and the names of the power plant components. They only have a small thematic relevance through the displayed names, but otherwise only a technological purpose for anchoring the virtual models. The physical environment in the study was a table inside a lab room, which also has no thematic relevance for the virtual elements and is not part of the learning content but just a surface to place the paper markers on. In general, the application can be used in any environment with enough space. The environment could be changed by moving the learners and the material to another location. This shows the potential of AR to move material to locations with higher thematical relevance, for example when showing this material inside a real combined-cycle power plant. Concerning the context visibility, the physical markers are fully covered by the virtual floor plates below the virtual models, so that information transferred through the markers would not be perceivable. However, the only relevant information from the markers, i.e. the name of the component, is also shown as part of the virtual element, so that no information that is relevant for the learning content is lost. The environment is not covered further than the area that is already covered through the paper markers, so that if it was thematically relevant, it could be viewed through the integrated on-screen visualisation.

In the heARt application used in Study 4 (Paper 3, Krüger et al., 2022) and shown in Figure 2, contextuality is shown in different ways. The physical anchor in the form of a paper-based marker is image-based, encompassing an unlabelled version of the graphic of the human heart as part of a worksheet on the topic. This way, the thematic connection between the physical anchor and the virtual object is clear through the image and the contextualising text. The general environment was again a non-relevant table in a lab room, although movement to a more relevant location like a cardiologist office would also be possible here due to device portability. Concerning the visibility of physical anchor and environment, the virtual view fully covers the physical marker with a plane, although in the 2D version, this includes a full recreation of the marker image with textual labels, so that no information is lost by covering the image. In the 3D version, the 2D image is fully covered by the white plane, so that the labelled 3D model that is inserted above is not perceivable in the same view as the 2D image.

In the ARbor application used in Study 5 (Paper 4, Krüger & Bodemer, subm.) and shown in Figure 3, contextuality manifests itself in different ways. Both the physical anchor in the form of an image-based marker and the surrounding environment are relevant for the learning material. Scanning the image of one of the plants shows information about that plant, providing additional information about the plant's appearance. The general environment is relevant in both versions of the application, as the learners are standing in an outside area with the respective plants surrounding them. The version where the markers are placed directly at the plants adds another level of relevance, where the application itself does not change, but the specific location of the marker achieved through the portability plays a role in linking the virtual elements to physical objects. The information from the application could also be accessed at different locations just like with the other two applications described above. In the study, the application is used outside, which has an influence on the contextual relevance. To keep this contextual relevance, the application must be used at least in the vicinity of these specific plants, or directly in front of the specific plant. Using the application in an outside environment in which no plants grow would not offer the same contextual relevance. Concerning the visibility, the AR application displays the virtual information in such a way that the image marker is not fully covered but that both kinds of elements can be perceived simultaneously. The general environment including the plants is not covered at all and thus fully visible.

In summary, the three applications show three different ways in which the relevance and visibility of the physical context may be designed. An overview can be seen in Table 3. The physical anchor elements and general environment may have no thematical relevance at all (powAR), the physical anchor elements may have a small thematical relevance although without offering that much additional integrated information (heARt), or the physical context may have full thematical relevance on both the anchor and the general environment level (ARbor). Also, the physical anchor elements may be fully covered by the virtual elements it retrieves (powAR, heARt) or it may be designed to not be completely covered because the physical anchor has thematic relevance for the learning material (ARbor). Fully covering the general environment is a less usual implementation in AR, because by definition physical and virtual elements are perceivable simultaneously, with the physical environment serving as the basis for the experience.

Looking more closely at potential implementations of contextuality in AR, there are established design principles that should be considered. Through the combined display of virtual and physical elements in AR, it is in general necessary to consider potential distractions of the physical elements and environment. While the virtual components can be designed as needed for a specific learning objective, the physical environment that is part of the experience often cannot and should not be changed, as it should display the physical world authentically. This means that also information that is not relevant for the learning objective is apparent in the physical world, which may distract from the actual learning objective. This is also described in the immersion principle which states that immersive 3D VR environments include so many perceptual details that extraneous processing and load may be evoked, distracting from the actual learning goal (Mayer, 2020e). The principle focuses on VR environments but can also be applied to AR environments, which also include details of the surrounding environment. This describes a more specific version of the coherence principle, which states that extraneous information, which have nothing to do with the learning objective, should be excluded in order to not disrupt learning processes (Mayer, 2020c). In order to decrease cognitive load (see Section 2.3 for more information on cognitive load), representations used in learning experiences should thus be coherent for the learning objective and external information should be avoided. More information on this and a potential application of the coherence principle with a combination of auditory and visual representations in AR can be found in Study 7 (Paper 5, Krüger & Bodemer, 2022a), which will be described in more detail in Section 4.5.2.

Table 3

Different Implementations of Contextuality in Three AR Applications

Study	Contextuality
Study 3: *powAR*	General environment: *lab room with table* ▪ no thematical relevance ▪ not covered, full visibility Specific physical anchor: *pattern-based AR markers with component names* ▪ close to no thematical relevance ▪ fully covered by virtual 3D models, no visibility
Study 4: *heARt*	General environment: *lab room with table* ▪ no thematical relevance ▪ not covered, full visibility Specific physical anchor: *image-based AR marker with graphic of heart on work sheet* ▪ thematical relevance of picture and text context, although no authentic/applied context ▪ fully covered by virtual plane, no visibility
Study 5: *ARbor*	General environment: *outside lawn with surrounding plants* ▪ full thematical relevance of surrounding plants within the authentic real world, in version with markers attached to plants also relevance of more specific plants ▪ not covered, full visibility Specific physical anchor: *image-based AR markers with photos of plants* ▪ full thematical relevance of photo for learning application ▪ only very small part covered by virtual elements, nearly full visibility

3.2 Interactivity

Interactivity describes the various possibilities of material manipulation in AR, which enable learners to interact with virtual and physical elements, including the manipulation of virtual elements through the manipulation of physical elements (see Section *2.2 Interactivity* in Paper 1, Krüger et al., 2019). This characteristic is based on the second characteristic of AR systems by Azuma (1997): interactivity in real time. While interaction with virtual elements in AR is possible through controllers or touch-based interaction as in fully virtual environments, additionally hand-based or full body interaction with physical objects and motion sensors is possible in AR. For educational settings, this includes the possibility to embed meaningful interaction into an AR-based learning scenario, leveraging mediated physical-virtual interaction to support cognitive processes, also enabling self-directed perspective changing. In the above-mentioned affordances, this aspect is mentioned as part of the provision of 3D-based interactive learning (Wu et al., 2013) and by MacCallum and Jamieson (2017) as part of the portability of devices to interact with a location. Also, the development of practical skills mentioned by Cheng and Tsai (2013) and Parsons and MacCallum (2021) may indirectly be connected to the possibilities for interaction.

3.2.1 Interactivity in AR-based Learning

Interactivity can be leveraged in the design of AR-based educational experiences that support learning (see Section *2.2 Interactivity* in Paper 1, Krüger et al., 2019, for basic and Paper 2, Krüger & Bodemer, 2020, for more detailed information on this). Interaction with learning materials has been defined and considered from many different perspectives. One definition of interactivity in technology-enhanced learning in multimedia settings describes it as a reciprocal activity in which learners react to a system and the system reacts to learners (Domagk et al., 2010). When looking at virtual and physical manipulation, it could be shown that a well-designed combination of the two enhanced learners' conceptual understanding (Olympiou & Zacharia, 2012), which is relevant for AR because of its capability to combine virtual and physical interaction. Physical interaction comes with demands for the learners, but virtual representations offer the possibility to constrain interactivity to interactions that are relevant for the task and learning objects, which can lead to less overload and better learning outcomes (Barrett et al., 2015). In AR, the possibilities of viewing and manipulating virtual objects with constraints to their interactivity can be leveraged while staying in an authentic, physical environment.

The ICAP (interactive - constructive - active - passive) framework focuses on the positive sides of interaction with learning material and describes that overt, active learning behaviour can indicate and support cognitive processing of the learning material if the behaviour is relevant to the learning task (Chi & Wylie, 2014). Similarly, the theory of embodied cognition describes that physical interaction with learning material and the environment are an important part of a learning situation and support cognitive activity (Wilson, 2002). Embodied interaction with a combined physical and virtual environment has also been described by Kim (2013) as an important aspect of context immersion in AR (see Section 2.4 on immersion). Furthermore, Holmes and colleagues (2018) describe that perspective taking by movement around objects leads to more successful learning than rotating the object, and that for both versions of interaction it is better when the movement is self-directed rather than passive. In a systematic approach towards interactive learning environments, Domagk and colleagues (2010) developed the INTERACT model, which shows the relations between different learning processes and outcomes in interactive learning environments. According to this model, the learning environment, e.g., an AR-based learning experience, has an impact on learners' behavioural activity, cognitive activity, and motivation. Two features of interaction mentioned in the model are guidance and learner control (Domagk et al., 2010). For guidance, the learning environment is designed in such a way that it guides learners' (meta-)cognitive activity which in turn can influence their behaviour. For learner control, the learning environment needs to offer the potential for behavioural control, which in turn can influence learners' (meta-)cognitive activity. Learner control as a form of interaction has been described to influence learners' motivation (Scheiter, 2021), so that this can be considered a relevant variable when it comes to interactivity in AR (see also Section 2.5 on motivation).

In a critical consideration of physical engagement with learning material, research in the field of multimedia learning describes that mental interaction is more important for learning than physical interaction (R. C. Clark & Mayer, 2016). The authors describe that behavioural engagement may cause ECL that can impede learning, especially when learners lack the ability to execute the behavioural task (see Section 2.3 on cognitive load). If physical interaction only leads to further load but not to germane processing, it can be detrimental to learning instead of promoting learning. This shows that not every physical interaction leads to better learning outcomes, so that it is important to stimulate relevant mental engagement with physical interaction. R. C. Clark and Mayer (2016) mention six design decisions that may support generative learning, including following the multimedia principle and thus adding graphics to text, and adding questions for learners. The paradigm of active integration has been introduced as a way to purposefully use interaction to support mental integration processes that do not take place automatically when learning with MERs (e.g., Bodemer & Faust, 2006; Bodemer et al., 2004, 2005). External, physical integration is leveraged in a way that learners themselves integrate separated representations with the goal to support cognitive coherence formation (see Section 2.1 for the literature review on MERs). It is assumed that this leads to more elaborate processes of mental integration, fostering generative processing instead of just decreasing external processing. In relation to the many possibilities of designing interactive AR, including interactions that can lead to generative processing or to overload, physical demands and cognitive processing of the content need to be considered, especially in the distinction between ECL and GCL based on cognitive load theory (see Section 2.3).

3.2.2 Recent Research on Interactivity

Recent research in the field of AR-based learning experiences describes positive effects of different kinds of interaction on learning processes and outcomes. In a study on interaction in the form of learner control in a museum, four different types of AR learner control tools were implemented in an exhibit on herbarium specimen (W. Lin et al., 2022). Three groups learned with low control, medium control, and high control, depending on how many of the interactive tools were implemented. While learning outcomes and flow experiences did not differ across the groups, the high control group engaged with the exhibit for a longer time and especially used the low control tools more often than the other groups.

They also reported more motivation in interviews. This shows that motivation and effort to engage with material can be increased through the interaction with AR-based learning materials.

In a study in the educational field of electric engineering, different forms of interaction with AR-based learning materials were compared (Dutta et al., 2022). In a marker-based version of the application learners could move multiple markers and in a keypad-based version learners pressed physical buttons to solve tasks based on a problem statement. Usability measured through manipulability and comprehensibility was perceived as higher in the keypad-based than the marker-based group, although it needs to be mentioned that the applications differed in more than just the interaction method, so that the effects cannot be fully attributed to that difference. The results however suggest that different forms of interaction in AR may have different impact on learning.

In another study, a physical user interface in the form of an interactive timeline was implemented for embodied learning about history (Gogou & Kasvikis, 2022). Learners were asked to walk up and down the timeline going through the different centuries while completing a mission given to them by narrations. Their understanding of the timeline was improved after using this application, although it is not clear whether the embodied interaction or other aspects of the application induced learning. This shows that whole-body interaction may be appropriate for certain learning objectives, especially when the movements are connected to the learning objective.

In another study that describes the application of the above-mentioned active integration paradigm in AR, we examined the effects of externally integrating physical and virtual elements (Krüger et al., 2023). The learners either externally integrated physical text cards with virtual 3D models shown in an AR-based tablet application or received a pre-integrated version of the material. While the data did not support the hypothesis that the external integration would lead to increased GCL and a better learning outcome, we found differences in learners' behaviour of scanning the AR markers, which show a potentially increased effort when handling the tablet during physical integration, which may have negated the potentially positive effects of the external integration process. This shows that in the implementation of meaningful physical interaction for the elicitation of mental interaction specific factors need to be considered in AR. A study that has been specifically designed with the interactivity characteristic in mind, focusing on the elaborateness of physical and mental interaction with AR-based material, is Study 3 in the current dissertation (Paper 2, Krüger & Bodemer, 2020). It will be described in more detail in Section 4.2.1.

3.2.3 Application of Interactivity

In a closer look at interactivity in AR, three levels of interaction can be defined (see Section *III.B. Interaction in AR* in Paper 2, Krüger & Bodemer, 2020): Learners can 1) interact purely with virtual elements, 2) interact purely with physical elements, or 3) use mediated interaction to manipulate virtual elements by manipulating physical elements which are connected to the virtual elements (see also tangible interface metaphor, Billinghurst & Dünser, 2012). The interaction on all three levels can be more or less elaborate, enabling actions from a simple circling or moving of a virtual object up to the full creation of a new object. Furthermore, the necessary movement may be bigger or smaller, including finger-based up to whole-body interaction. Also, based on the assumption that mental interaction but not physical interaction on its own leads to relevant learning processes, the embedded physical interaction may be more or less relevant for the mental interaction with the learning content. To clarify the potential differences, I will in the following analyse the three introduced AR applications used in studies included in this dissertation concerning these differences in elaborateness of interaction, size of movement, and relevance for mental engagement on the three levels of purely physical, purely virtual, and mediated physical-virtual interaction.

In the powAR application used in Study 3 (Paper 2, Krüger & Bodemer, 2020) and shown in Figure 1, different types of interaction are possible. On the purely virtual level, learners can interact with the virtual app interface through buttons in order to access additional information that depends on the current set-up of the built power plant. This interaction is only necessary for information retrieval, so it is not very elaborate, and it is a small, finger-based movement. The button-based interaction does not play a role for the mental interaction with the learning content, although revealing the additional information on the application interface is necessary to achieve the learning goal. Concerning the purely physical interaction, there are no manipulations worth mentioning as the AR markers themselves transport no information, so that manipulating them without the virtual level is meaningless. The movement of the pattern-based markers is relevant for the mediated interaction, where movement of the physical paper markers leads to movement of the virtual 3D models of the power plant components. While in one version of the application the AR markers are already clustered, in the other version the learners need to assemble them to functioning power plants. The components that are meant to be linked and are placed close enough to each other connect and react to each other, displaying animations. This interaction is quite elaborate as it leads to the creation of new artefacts, both through the connection of pairs of components and the stepwise building of the complete power plant. The hand-based movement is bigger than a finger-based touch interaction, and relevant mental engagement may be elicited by moving the components to their respective places, supporting the construction of a mental model of the whole power plant. Another mediated interaction in this application includes the whole-body interaction of walking around and changing perspectives around the virtual models by moving the tablet.

In the heARt application used in Study 4 (Paper 3, Krüger et al., 2022) and shown in Figure 2, interactivity is only basic. There is no purely virtual interaction in the application, as there is no interface integrated. Purely physical interaction only includes the possibility to move around and turn the worksheets, which may be relevant for reading the text from a better perspective. The mediated physical-virtual interaction in the application includes the possibility to move and turn the physical sheet of paper including the AR marker in order to view the model from different perspectives. These are very simple hand-based movements that may be relevant in supporting the building of a mental model of the 3D object from different perspectives. Furthermore, whole-body movement is also possible to execute the same form of mediated interaction, namely moving around with the tablet to look at the model from different perspectives and from closer or further away.

In the ARbor application used in Study 5 (Paper 4, Krüger & Bodemer, subm.) and shown in Figure 3, different kinds of interaction are possible. Concerning purely virtual interaction, touch-based interaction with the interface can be executed, which includes clicking on different virtual elements in the application, such as the picture panels, the checkmark to change the view to the 3D model, and the words in the texts or the glossary list. These are small interactions leading to interface changes for information retrieval. The information retrieval is necessary to achieve the learning goal, even though the interaction itself does not necessarily support mental processing. Clicking on words in the text to receive more information may support the mental connection between the text panel in which the word was clicked, the word, and the explanation. Concerning purely physical interaction, it is possible to move the marker images, for example to get a closer look at the pictures of the plants. However, in the set-up of the experiment the images were deliberately fixed to specific spots so that interaction with them should not have been the learners' focus. A purely physical whole-body interaction is walking from one marker to the other, which was necessary here due to the markers being spread across the space. This borders in mediated physical-virtual interaction because this is the only way to access the virtual information about the different plants. In the version in which the markers are placed at the respective plants, walking around offers not only a form of transport but also spatial information, offering support to process the placement and distinction of the different plants. Due to the AR markers being fixed to specific positions, moving the markers to move the virtual information is not the focus in the mediated interaction, but learners moved in relation to the markers. For example, the tablet could be moved closer and further away to zoom in and out of virtual information, and in the view of the 3D model of the plant perspective changing was possible through tablet movement. This simple interaction can support information retrieval like looking at more details in the pictures when moving closer.

In summary, the three applications show different manifestations of interactivity with AR-based elements, considering the elaborateness of the interaction, the size of the movement, and how relevant it is for cognitive processes. An overview can be seen in Table 4. The purely virtual interaction in the three applications reaches from touch-based interaction with buttons or other virtual elements (powAR, ARbor) to no virtual interaction. Most of this interaction only has the goal to retrieve additional information from the application. The purely physical interaction in the three applications also differs from no meaningful interaction (powAR) to interaction with the working sheet that is part of the learning material (heARt) and walking between different locations in which the markers are placed (ARbor). The mediated virtual-physical interaction also differs, although in all three applications movement around and of the AR markers enables changes in perspective and zooming in and out of the virtual elements. This interaction can, however, also be more elaborate, including the construction of new virtual artefacts when assembling physical paper cards in a correct way (powAR). Furthermore, when it comes to virtual 3D objects, perspective changing around the object can support mental model construction (heARt).

Table 4

Different Implementations of Interactivity in Three AR Applications

Study	Interactivity
Study 3: *powAR*	Purely virtual interaction: *touch-based interaction in virtual interface* ▪ not very elaborate, only retrieving additional information ▪ small, finger-based interaction with virtual buttons ▪ information relevant, but interaction itself not relevant for mental processes Purely physical interaction: *movement of paper cards* ▪ not elaborate ▪ bigger, hand-based interaction ▪ no meaningful interaction Mediated physical-virtual interaction: *movement of paper cards to move/rotate virtual models and moving/walking around virtual models* ▪ not so elaborate when only viewing models from different perspectives, but quite elaborate when creating new virtual artefacts when assembling paper cards ▪ bigger, hand-based interaction when moving paper cards; big, whole-body interaction when walking around to view models from different perspectives ▪ relevant mental engagement may be elicited supporting the construction of a mental model of the power plant when externally constructing it
Study 4: *heARt*	Purely virtual interaction: *n.a.* Purely physical interaction: *movement of worksheet* ▪ not very elaborate, maybe necessary to retrieve information due to easier reading ▪ bigger, hand-based interaction ▪ relevant to retrieve information through easier reading, but interaction itself not relevant for mental processes Mediated physical-virtual interaction: *movement of worksheet to move/rotate virtual element and moving/walking around virtual element* ▪ not so elaborate, only viewing models from different perspectives ▪ bigger, hand-based interaction when moving sheet, big whole-body interaction when walking around the marker to view model from different perspectives ▪ relevant mental engagement may be elicited supporting the construction of a mental model of 3D model from different perspectives
Study 5: *ARbor*	Purely virtual interaction: *touch-based interaction in virtual interface* ▪ not very elaborate, only retrieving additional information ▪ small, finger-based interaction with virtual elements ▪ information relevant for learning objective, interaction with words in texts may support mental connection of respective text panels, words, and explanations Purely physical interaction: *walking to the different images* ▪ not elaborate, only walking ▪ big, whole-body interaction ▪ mental processing of different placement of information possible Mediated physical-virtual interaction: *moving closer to AR markers and thus virtual elements, walking to the different AR markers with virtual information* ▪ not so elaborate, only walking towards different information and viewing information from closer or further away ▪ big, whole-body interaction when moving the tablet or walking around ▪ information accessed is relevant for learning objective, but interaction itself not relevant for mental processes

Potential implementations of interactivity in AR should consider established design principles, including multimedia design principles. Through the different levels of interaction, it is important to discuss which interactions evoke relevant mental processes. One potential implementation of interaction in a multimedia-based learning setting, is learner control (Scheiter, 2021). Potential learner control can include sequencing, selection, content control, pacing, and representation control. Different types of learner control can lead to different cognitive processes, and the learner control principle states that learning is increased when instructional support and levels of prior knowledge are high enough to cope with the control options. If learners do not have much prior knowledge, a guided activity may be more appropriate, providing them with a pedagogical agent to guide their learning processes (Moreno & Mayer, 2007). For AR this also means that learner control should not only support relevant mental processes but should also be appropriate and not overloading for the learners who execute it.

3.3 Spatiality

Spatiality describes the potential of AR to add unique spatial properties to virtual elements when placing them within the physical world, which enables learners to perceive the spatial expansion of virtual elements and to observe virtual elements at fixed points within the physical space (see Section *2.2 Spatiality* in Paper 1, Krüger et al., 2019). This characteristic is partly based on the third characteristics of AR systems described by Azuma (1997): registering 3D elements in the real world, mainly due to the possibility to scan and track surfaces. Compared to screen-based applications without AR capabilities, the virtual object can have more depth and compared to physical objects it can be displayed in the most convenient size. Furthermore, it can be attached to specific physical locations and objects. For educational settings, this includes the possibility to spatially integrate physical and virtual objects so that mental integration may be simplified. Moreover, the use of 3D representations may better support learners in the construction of complete mental models of 3D objects. In the above-mentioned affordances, this aspect is best reflected in the aspects of providing 3D-based visualisation and learning (MacCallum & Jamieson, 2017; Wu et al., 2013) and supporting representation of spatial concepts (Bower et al., 2014).

3.3.1 Spatiality in AR-based Learning

Spatiality can be leveraged in the design of AR-based educational experiences that support learning (see Section *2.3 Spatiality* in Paper 1, Krüger et al., 2019, for basic and Paper 3, Krüger et al., 2022, for more detailed information on this). Through adding virtual elements into three-dimensional physical space, they obtain spatial properties that virtual elements bound to screens do not have. Image-based AR technology has been used in multiple studies to help learners in comprehending the 3D structure of objects (Cheng & Tsai, 2013). The learning of spatial structures in spatial domains has been described as one of the main learning benefits of AR (Radu, 2014). In AR, 3D objects can be presented more realistically and with more depth cues (e.g., motion-based, see Craig, 2013a), which enables more authentic learning with virtual elements. 3D representations of spatial objects can play an important role in supporting learning (see Section *1.1. Dimensionality of representations in education* in Paper 3, Krüger et al., 2022 for more details). Wu and Shah (2004) describe that learners have difficulties with translating external 2D representations into internal 3D representations or mental models because they have difficulties identifying depth cues in 2D visualisations and form 3D mental images from 2D structures. The use of 3D models has been suggested to lead to the construction of more advanced mental models than the use of 2D illustrations (S.-C. Chen et al., 2015), as they do not first need to be mentally converted from a 2D to a 3D representation. This shows that using 3D representations in AR may be especially relevant when the content of the learning material is spatial, potentially decreasing ECL elicited when converting 2D to 3D representations and increasing GCL when supporting the construction of 3D mental models (see Section 2.3 on cognitive load). When it comes to 3D visualisations, learners' spatial abilities play an important role (see also Section *1.1.3. Spatial abilities* in Paper 3, Krüger et al., 2022). Here, an ability-as-compensator and an ability-as-enhancer hypothesis can be applied, suggesting that either learners with lower spatial abilities or learners with higher spatial abilities could profit from 3D representations (see Section 2.6 on spatial abilities).

Concerning the spatial anchoring of virtual elements inside the physical world, one important component includes their placement in relation to physical objects. Virtual elements can be used for textual annotations and labelling (Sugiura et al., 2019) and pictorial overlay over physical objects (Ferdous et al., 2019) due to their spatial anchoring and placement in the physical world. Spatially integrated annotation of the physical world can support task execution (Volmer et al., 2018). This spatial closeness of corresponding information can support its mental integration and decrease ECL (Ayres & Sweller, 2014). In relation to the different ways of using spatiality in AR, cognitive processing and load are thus relevant from the learner's perspective (see Section 2.3 on cognitive load).

3.3.2 Recent Research on Spatiality

Recent research in the field of AR-based learning experiences describes positive effects of spatial representations for learning processes and outcomes. In one study, for example, Shaghaghian and colleagues (2022) describe AR applications for learning about spatial transformations in math education. One application includes a physical 3D model, mathematical information about transformations of the model in a virtual overlay, spatially anchored virtual graphics on transformation, and a virtual version of the model that can be rotated based on parameter input. When looking at the pictures in the article it becomes clear that a 2D graphic of this would be very hard to understand, and in a user study an improvement of the knowledge about the topic could be found. This shows the advantages that the implementation of 3D visualisations in AR could bring for spatial understanding and learning.

In another study, 3D AR visualisations of weather phenomena for the education of pilots were designed (Meister et al., 2022). The application showed the development of thunderstorms in 3D. In a small evaluation an increase in knowledge from pre- to post-test and relatively low workload were found for learners. Although these results cannot be fully attributed to the 3D presentation because there was no control group for comparison, it can be assumed that it may have played a role here.

In a study specifically focusing on the comparison of 2D and AR-based 3D learner output, learners worked on a construction sequence for wood-framed elements (McCord et al., 2022). Both groups received the same 2D drawing of the wood frames but used either a 2D presentation of the construction on a worksheet or a 3D presentation of the pieces of wood in an AR HMD to define a construction sequence based on the drawings. While the 2D worksheet was less mentally demanding than the 3D AR implementation, it also led to more uncorrected mistakes. In AR, the placement in physical space, the size of the wood frames, and the dimensionality of presentation potentially all played a role in supporting learners in noticing their mistakes, leveraging multiple potentials of spatiality in AR. However, due to these many features it cannot be further concluded through which mechanisms the learning process was impacted.

In another study that we executed, a moderating influence of different types of spatial abilities on the impact of an AR version of a visualisation of the International Space Station on a spatial learning task execution and outcomes was found (Krüger & Bodemer, 2021). On one hand, learners with low 3D spatial visualisation abilities profited from an AR in comparison to a non-AR visualisation in the execution of the learning task. On the other hand, learners with high 2D spatial memory abilities profited from an AR in comparison to a non-AR visualisation in the spatial knowledge test. This shows that it is important to take learners' spatial abilities into account when 3D objects are used in an AR-based learning experience. A study that has been specifically designed with the spatiality characteristic in mind, focusing on the dimensionality of representation of a virtual object in AR and learners' spatial abilities, is Study 4 in the current dissertation (Paper 3, Krüger et al., 2022). It will be described in more detail in Section 4.3.1.

3.3.3 Application of Spatiality

When looking more closely at spatiality in AR, two types of factors can be distinguished (see Section *1. Introduction* in Paper 3, Krüger et al., 2022): the anchoring of virtual elements at specific points in the physical space, and the spatial properties of the virtual elements, including spatial expansion and thus size and dimensionality of the virtual elements themselves. In the spatial placement of virtual elements in the physical world, two levels of technological placement can be distinguished: image-based and location-based AR described by Cheng and Tsai (2013) or accordingly vision-based and location-aware AR described by Dunleavy and Dede (2014). To describe the different experiences without the technologically informed terms, I choose and introduce the terms "element level" and "world level". On the smaller scale element level, there is a direct link between the physical anchor element (e.g., AR marker or physical object) and the virtual element, which enables mediated interaction and spatial integration of virtual and physical elements. The bigger scale world level is the spatial placement of virtual information at specific locations and within specific environments in the physical world. Furthermore, in AR, virtual elements can copy the spatial expansion of physical elements in a three-dimensional physical space. Dimensionality of the virtual object plays an important role here and both

2D images and 3D objects can be recreated. Moreover, the usage of 3D objects can be more or less relevant for the learning objective. To clarify the potential differences, I will in the following analyse the three introduced AR applications used in studies included in this dissertation concerning these differences in spatial anchoring to the physical world on the small-scale element level and the large-scale world level, and the spatial properties of the virtual elements, including size, dimensionality, interconnectedness of components, and relevance for the learning material.

In the powAR application used in Study 3 (Paper 2, Krüger & Bodemer, 2020) and shown in Figure 1, spatiality plays different roles. Concerning the spatial connection, on the big scale world level the virtual elements are not spatially linked to a specific location but placed inside a lab for the study. On a small, element level, the virtual models are spatially connected to the respective pattern-based markers with each virtual object being linked to its own physical paper marker. Concerning the spatial properties of the virtual objects it can be said that all objects are 3D models that are approximately hand-sized and thus miniature versions of the different components of power plants, but otherwise authentic representations. Spatial linking of the virtual objects that can be connected to each other takes place when they are close to each other, so that the spatial structure and interconnection between components is apparent. This interlinking between virtual objects can thus support an understanding of the spatial structure of the power plant, which is relevant for the learning objective as it helps with the understanding of the processes taking place.

In the heARt application used in Study 4 (Paper 3, Krüger et al., 2022) and shown in Figure 2, dimensionality as part of the characteristic spatiality is in focus. Concerning the spatial connection on a world level scale, there is again no fixed link but a placement inside a lab for the study. On a small, element level, the virtual graphic or model is spatially linked to the image-based marker, in the 2D condition overlaying the image on the physical paper exactly with a virtual version with added labels. In the 3D version, the model appears to be floating over the worksheet. Concerning the dimensionality of the virtual object, the 2D version includes only 2D and no 3D elements, but the 3D version includes the 3D model of the human heart, which has approximately the size of a real human heart and appears to take up space in the physical room. The structure of the components of the heart is apparent within the model, but no connection to other virtual elements takes place. Although the visualisation is based on a cross-section, which can also be displayed as a 2D image, the depth of the 3D model can help with the understanding of the spatial structure of the human heart and is thus relevant for the learning goal.

In the ARbor application used in Study 5 (Paper 4, Krüger & Bodemer, subm.) and shown in Figure 3, spatiality appears in different ways. On the world level scale, the AR experience is placed at a specific lawn surrounded by plants that are part of the learning material. This placement is due to the placement of the AR markers, which are placed either directly at the respective plants that they offer information about, or at stones in front of the lawn. On an element scale level, the virtual information shown in the application is spatially linked to the specific marker image that they show additional information about, with the lines to the pictures connecting the specific parts of the plant like a line from the leaves to the picture of the leaves. Concerning the dimensionality, it can be said that nearly all information in the application is in 2D, although the pictures shown when scanning the marker are protruding a little. The virtual pictures and texts that belong to each other are spatially linked by being placed close to each other. However, for each plant there is one 3D model included, showing a minimised virtual representation of the respective plant, with size relations of the different plants approximately fixed for size comparisons.

In summary, the three applications show different manifestations of spatial linking and dimensionality that are possible in AR applications. An overview can be found in Table 5. On a world scale level, the location in which information is placed can be inside (powAR, heARt) or outside (ARbor). On an element scale level, all three applications include a spatial connection between a physical anchor (i.e. AR markers) and respective virtual elements. Concerning the dimensionality, the applications all included some form of 3D models, except for the 2D version of the heARt application, although the ARbor application mainly relies on 2D information in the form of texts and picture to transport information. Concerning size, the 3D models were miniature versions of objects that cannot be shown in their real size (powAR, ARbor) or shown approximately in their real size (heARt).

Table 5

Different Implementations of Spatiality in Three AR Applications

Study	**Spatiality**
Study 3: *powAR*	<u>Anchoring in physical world:</u> *marker-based placement of virtual elements* ▪ markers placed on a table in a lab room at world level, portable ▪ virtual elements standing upon their respective markers at element level <u>Spatial properties of virtual object:</u> *multiple virtual models, one per marker* ▪ 3D models of power plant components ▪ hand-sized model, size very decreased in comparison to real objects ▪ objects linked to each other when spatially close to each other, building full model
Study 4: *heARt*	<u>Anchoring in physical world:</u> *marker-based placement of one virtual element* ▪ marker placed on a table in a lab room at world level, portable ▪ virtual element on top of (2D version) or floating above (3D version) marker at element level <u>Spatial properties of virtual object:</u> *one virtual model* ▪ 3D models of human heart in 3D version; 2D image in 2D version ▪ palm-sized model, size similar to real object size ▪ no spatial linking of different virtual objects
Study 5: *ARbor*	<u>Anchoring in physical world:</u> *marker-based placement of individual virtual elements* ▪ markers placed at different locations in the physical world with walking distances between them, (partly) fixed ▪ virtual elements on top of their respective markers at element level <u>Spatial properties of virtual object:</u> *multiple virtual models, one per marker* ▪ 2D images and texts; 3D models of plants ▪ models of plants hand-sized, size decreased in comparison to real objects, size relations of different objects approximated ▪ spatial linking of pictorial and corresponding textual information within one marker, placed next to each other

Looking more closely at potential implementations of spatiality in AR, there are established design principles that should be considered. Through the possibility of AR to include multiple representations, multimedia design principles concerned with the integration of different forms of representation might need to be considered (see also Section 2.1). The spatial contiguity principle describes that corresponding textual and pictorial representations should be presented close to each other (Mayer, 2020d). In AR, this spatial integration of representations is not only possible with words and images, but also with virtual and physical elements. This can be used to support the mental integration of elements that belong together. More information on this and a potential application of the spatial contiguity principle with a combination of virtual and physical elements in AR can be found in Study 6 (Paper 5, Krüger & Bodemer, 2022a), which will be described in more detail in Section 4.5.1.

3.4 Interplay of the Three Characteristics

In the previous sections, I defined the three ARcis characteristics individually. A unique thing about AR is, however, the combination of these characteristics and thus the potentials of their interplay (see also Section *2.4 Interplay of the three characteristics* in Paper 1, Krüger et al., 2019). When comparing AR and non-AR, multiple characteristics may offer meaningful differences for learning. In Study 1 (Paper 1, Krüger et al., 2019), two tablet-based applications are compared, one AR and one not. Here, both spatiality and interactivity were defined to have meaningful differences (see Section 4.1.1 for more detail on this differentiation). This shows that AR and non-AR can differ in more than one aspect, which entails that media comparisons do not necessarily provide an insight into specific mechanisms due to the potential of confounding variables. When both spatiality and interactivity differ, it cannot be determined by which of the two a potentially positive effect is elicited.

When it comes to the placement of virtual information, spatiality and contextuality of AR are closely connected and thus sometimes hard to separate. The thematic proximity and relevance as part of contextuality is directly linked to the spatial anchoring of virtual elements within the physical world as part of spatiality including the spatial proximity of corresponding physical and virtual elements. In the ARbor application (used in Study 5, Paper 4, Krüger & Bodemer, subm.; see Figure 3), for example, the spatial placement of the AR markers and thus the virtual elements attached to them is based on the thematical relevance of the physical objects to which the markers are attached. Contextuality can thus profit from the potential of spatial placement within the physical world. The meaningful difference for the learning objective is the closeness to relevant physical objects, which is achieved through the spatial placement. Just displaying virtual elements close to physical elements may not influence learning processes when the virtual and physical elements are not also thematically connected.

Another close connection exists between the interactivity and spatiality in the case of mediated interaction using a tangible interface metaphor in AR. When learners move physical objects (e.g., AR markers) in order to move virtual elements, this is only possible through the spatial linkage of the virtual to the physical elements. In the powAR application (used in Study 3, Paper 2, Krüger & Bodemer, 2020;

see Figure 1), for example, animated virtual 3D models of power plant components are spatially linked to AR markers in the form of paper cards. Through moving the paper cards, the virtual 3D models can be moved and can then interact in the virtual space when placed next to each other. Interactivity thus also profits from the potential of spatial linkage with physical objects.

In another connection between interactivity and spatiality, the spatial perception of virtual 3D objects is supported by interactivity. When viewing a virtual 3D object in AR, learners can change their perspective by walking around or by rotating a physical anchor object, if applicable. This way, they can view the object from different perspectives, which supports motion-based depth cues and 3D perception. In the heARt application (used in Study 4, Paper 3, Krüger et al., 2022; see Figure 2), the spatial perception of the virtual model of the human heart is possible because learners can walk around the AR marker while holding the tablet and looking at the virtual visualisation. Spatiality thus profits from the potential of walking and moving objects in the physical space.

Although for the design of experimental studies aiming at comparing very specific mechanisms and thus requiring very small manipulations of purposeful design elements the separation of the three characteristics can be very helpful, in designing AR applications for the field it can be hard and even disadvantageous to separate the three characteristics. For the design of most effective and efficient AR-based learning applications, it is further important to look at the specific learning objective and evaluate how the three characteristics and their interplay can be leveraged to support the achievement of the specified goals.

Besides looking at the interplay from a design-perspective, it is also important to examine the learners' perspective and how they perceive and work with different kinds of applications. Even when a design is intended to only change one of the three characteristics, it may also have an impact on learners' behaviour and perception concerning the other characteristics. In Study 4 (Paper 3, Krüger et al., 2022), we implemented a first version of the ARcis questionnaire, which was designed in an attempt to quantify learners' experience when learning in AR based on the three characteristics contextuality, interactivity, and spatiality (see Krüger & Bodemer, 2022b for a first evaluation of the questionnaire). In the results it could be seen that not only spatiality, which we wanted to manipulate in the study through the dimensionality of the virtual object, but also both contextuality and interactivity were perceived as higher when a 3D instead of a 2D virtual visualisation was used. This shows how hard a distinction between the three characteristics can be not only in the design but also in learners' perception. A 3D object may increase the affordance of walking around to view it from different perspectives, increasing at least its perceived interactivity (see Section *4.5. Limitations and future studies* in Paper 3, Krüger et al., 2022 for a more detailed discussion of these results).

In addition to research that looks at specific aspects of the individual ARcis characteristics, there should thus also be research looking at their interplay, for example through study designs looking at interactions of different factors.

3.5 Summary

In order to achieve insights for the first subgoal of this dissertation, I defined specific characteristics of AR-based learning and analysed which mechanisms might be relevant to support learning based on the literature on AR. The three characteristics contextuality, interactivity, and spatiality have been defined as respectively the combination and integration of matching contexts and virtual elements, the possibilities of manipulation of physical and virtual objects, and the placement and dimensionality of virtual elements in physical space. All of these have been proposed to have a potential influence on learning when applied in an AR-supported instructional setting, including learning that is situated in a relevant context, learning in which virtual elements are manipulated through the manipulation of physical elements, and learning about spatial structures. I propose that a separate consideration of these characteristics can help with structured research and design of AR-based learning experiences.

A summary of factors within the ARcis characteristics contextuality, interactivity, and spatiality that may be relevant for educational experiences and were described above can be found in Table 6. One important observation is that for each characteristic one factor has to do with its meaningfulness for the learning process: how meaningful is it for the targeted mental processes and learning objectives to place the information in this context, to include this form of interaction, and to display 3D objects? Instructors should keep the learning objectives and mental processes that should be achieved in mind when designing or implementing an AR-based learning experience. Based on this, the characteristics of AR can then be leveraged purposefully.

Looking more specifically at the mechanisms that may be involved in learning with AR based on the three characteristics and their influence on learning processes and outcomes, different points can be summarised from the previous paragraphs. Due to the combination of virtual and physical information, a main asset of AR-based learning lies in the support of building an integrated mental model of these two sources of information. Through the combination of corresponding but complementary information (contextuality), learners' attention can be moved to this integration. Furthermore, through the spatially close placement (spatiality), this focus may be further increased, with the possibility to use an external, physical integration (interactivity) to support the mental integration process. This shows that mechanisms included in all three characteristics can have an influence on cognitive processes and thus cognitive load. Another mechanism has to do with increasing the perceived relevance and thus motivation of learners when situating tasks and information in their corresponding context (contextuality). Enjoyment and motivation can also be increased through the possibility of whole-body interaction with learning material (interactivity). Feelings of immersion can be increased by situating learners in authentic contexts (contextuality) and evoking embodied interaction with the environment (interactivity) can increase context immersion. Further mechanisms include mental model construction of an object based on a true 3D representation (spatiality) that can be intuitively observed from different perspectives (interactivity), thus removing the necessity to transform a 2D representation into 3D and decreasing unnecessary cognitive load that may be especially damaging for learners with lower spatial abilities, while increasing load concerned with building the mental model in 3D. In conclusion, these mechanisms describe that contextuality and interactivity can have an impact on immersion, motivation, and cognitive load, while spatiality can mainly have an impact on cognitive load with learners' spatial abilities as a potentially moderating variable. The learning process-related constructs immersion, motivation, and cognitive load are in turn expected to have an influence on learning outcomes. A summary of these constructs that may be relevant when learning with AR and their relations to the ARcis characteristics can be found in Figure 4.

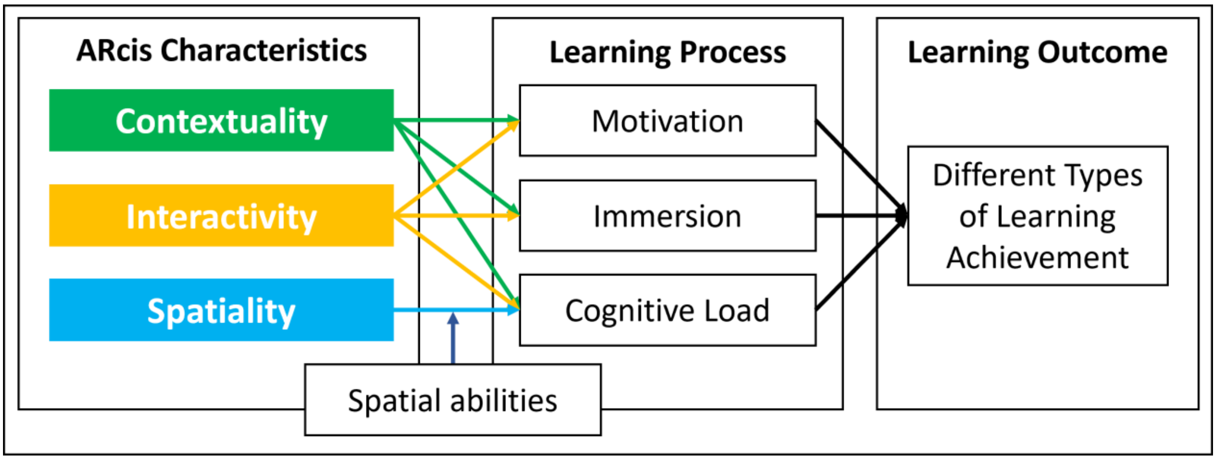
Table 6

Relevant Factors within the Three ARcis Characteristics when Learning with AR

Contextuality	Interactivity	Spatiality
Combination of virtual and physical elements in AR	*Potentials of material manipulation and interaction in AR*	*Spatial properties gained by virtual objects in AR*
▸ Two levels of connection: ▪ Contextualised in physical environment in general ▪ Anchored to specific physical elements	▸ Three levels: ▪ Purely physical interaction ▪ Purely virtual interaction ▪ Mediated physical-virtual interaction	▸ Two levels of anchoring virtual elements in physical world ▪ Small-scale element level ▪ Large-scale world level
▸ Factors on both levels: ▪ Relevance of context for learning material ▪ Visibility of context when scanning	▸ Factors on all three levels: ▪ Elaborateness of interaction ▪ Size / amount of movement ▪ Relevance of interaction for cognitive processes	▸ Spatial properties of virtual object: ▪ Size of / place taken up by object ▪ Dimensionality of models ▪ Interconnectedness of components within ▪ Relevance of dimensionality

Figure 4

Relevant Variables when Learning with AR



4 Empirical Research on the ARcis Characteristics

In the five papers included in the current dissertation, seven studies have been executed. The current section describes the work towards the achievement of empirical Subgoal 2, which aims at empirically examining how the characteristics of learning with AR and their specific mechanisms influence learning. The empirical studies were based on the three ARcis characteristics and potential mechanisms defined in Section 3: contextuality, interactivity, and spatiality. Based on these, three general, broad research questions were formulated to guide the study design to examine AR-based learning (RQ1 - RQ3). Furthermore, bridging empirical Subgoal 2 and practical Subgoal 3 of the current dissertation, empirical studies with a focus on specific design implementations were executed. As AR is predestined for the implementation of multimedia material, which is also described in connection to the three ARcis characteristics in Section 3, RQ4 was formulated accordingly. The four general research questions are:

RQ1: How does contextuality influence learning in AR?

RQ2: How does interactivity influence learning in AR?

RQ3: How does spatiality influence learning in AR?

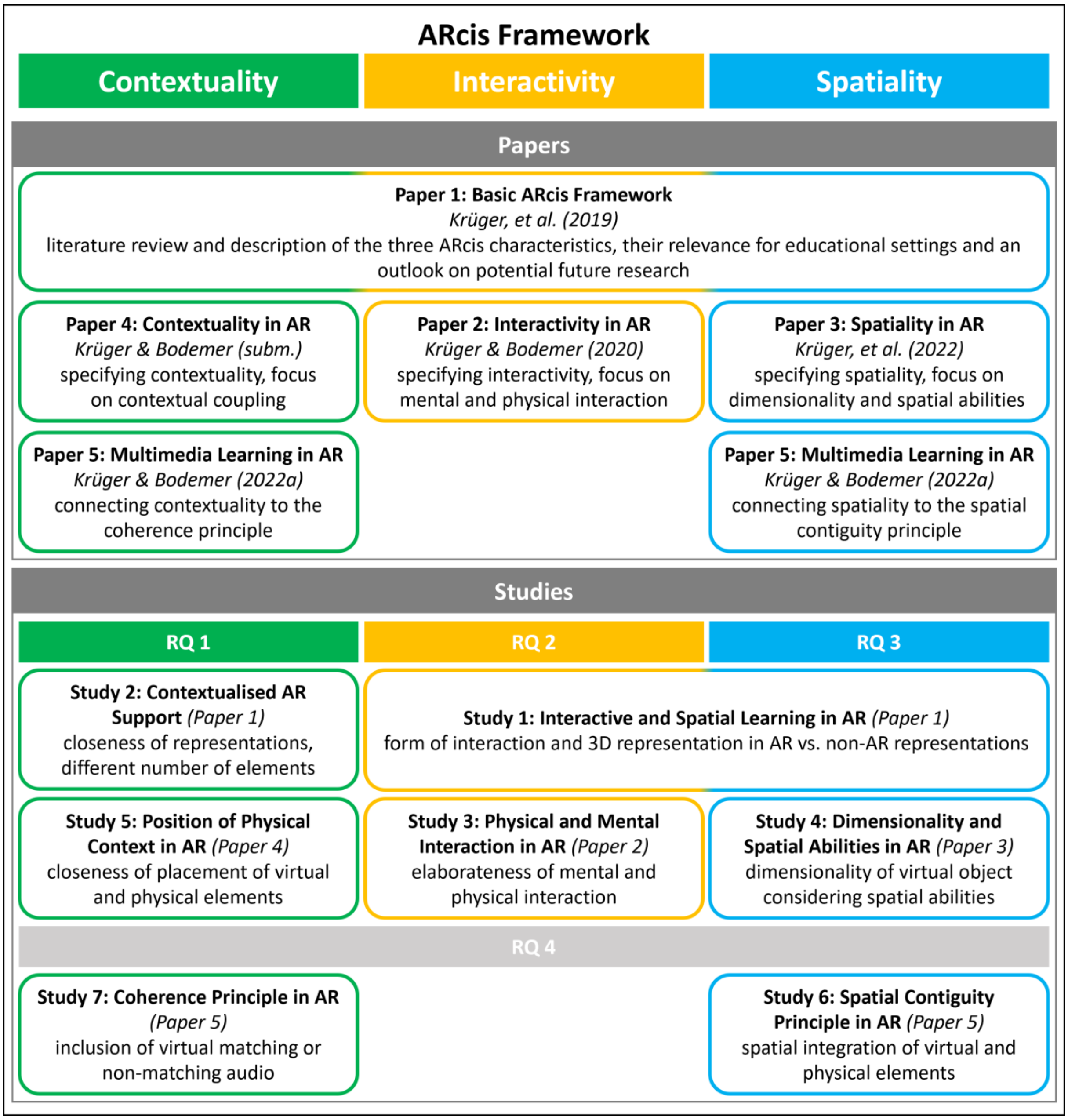
RQ4: How does the implementation of multimedia design principles in AR influence learning?

These general research questions were further specified to focus on specific aspects and mechanisms of the characteristics. For RQ1, the influence of the placement of virtual information in relation to corresponding physical objects in AR on learning processes and outcomes was focused. The question was investigated in Study 2 (Paper 1, Krüger et al., 2019) and Study 5 (Paper 4, Krüger & Bodemer, subm.). While Study 2 focuses on the closeness of representations of group awareness information and pictures of people in an AR mock-up, Study 5 examines the closeness of virtual information about plants to corresponding physical plants in an AR-based learning setting in nature. For RQ2, the role of mental and physical interaction in AR for learning processes and outcomes was focused. This question was investigated partly in Study 1 (Paper 1, Krüger et al., 2019) and fully in Study 3 (Paper 2, Krüger & Bodemer, 2020). While Study 1 compares touch-based interaction to mediated physical-virtual manipulation of animated 3D models of power plant components through paper-based AR markers, Study 3 examines the interaction of mental and physical interaction with those virtual 3D models through paper-based markers. For RQ3, the influence of the dimensionality of the virtual representation in AR on learning processes and outcomes was focused. This question was investigated partly in Study 1 (Paper 1, Krüger et al., 2019) and fully in Study 4 (Paper 3, Krüger et al., 2022). While Study 1 compares non-AR pseudo-spatial representations to AR-based spatial representations of virtual animated 3D models of power plant components, Study 4 examines the role of the dimensionality of a virtual representation of a human heart, also considering spatial abilities. For RQ4, two design principles defined as relevant for AR (see Sections 3.1.3 and 3.3.3) were chosen, focusing this research question on AR-based applications of the coherence and the spatial contiguity principle and their influence on learning processes and outcomes. This question was investigated in Study 6 and Study 7 (Paper 5, Krüger & Bodemer, 2022a). While Study 6 examines the application of the spatial contiguity principle (related to spatiality and RQ3) concerning the spatial integration of virtual texts and physical plants in a botanical garden, Study 7 examines the application of the coherence principle (related to contextuality and RQ1) concerning the inclusion of virtual bird or other sounds in material about birds.

First insights for answering these four general research questions were gained through the seven studies. In Figure 5, an overview of the papers and studies grouped in accordance with the ARcis characteristic and research question is provided. In the following, all papers will be summarised, with a description of the theoretical development of the ARcis framework and the respective study or studies. In Table 7, the specific theoretical and empirical goals per paper and study are summarised.

Figure 5

Overview Over the Five Papers and Seven Studies Included in This Dissertation



4.1 Paper 1: Basic ARcis Framework

Title: "Augmented Reality in Education: Three Unique Characteristics from a User's Perspective" (Paper 1, Krüger et al., 2019)

In Paper 1, we introduce the three AR-specific characteristics contextuality, interactivity, and spatiality. These are defined based on their system-focused counterparts in the definition by Azuma (1997) and related to their potentials for educational settings in a review of current literature. We formulate potential research questions for each characteristic, two questions for individual learning and two for group-based learning settings. Examples of potential research questions described in this publication are "Do people indeed learn better when they are in a relevant context than when they are not and which (cognitive, motivational, and emotional) factors play a role in this?" (see Section *2.1 Contextuality* in Paper 1, Krüger et al., 2019, pp. 414–415) for contextuality, "How must interaction with the material be designed to evoke higher order thinking processes?" (see Section *2.2 Interactivity* in Paper 1, Krüger et al., 2019, p. 415) for interactivity, and "Is using a three-dimensional AR object as beneficial for learning spatial structures as real objects are, in comparison to screen-based objects?" (see Section *2.3 Spatiality* in Paper 1, Krüger et al., 2019, p. 416). Further, the characteristics' potential interplay is described (see Section *2.4 Interplay of the three characteristics* in Paper 1, Krüger et al., 2019). For a full description of the characteristics see Section 3. We then describe two exemplary studies of how to define and research those characteristics in AR-based educational applications, which I will describe in the following two sections.

4.1.1 Study 1: Interactive and Spatial Learning in AR

In the first study, the goal was to compare a tablet-based AR application to a tablet-based non-AR application concerning their influence on learning processes and outcomes. The general research question can be formulated as "How does an AR in comparison to a non-AR version of a simulation influence knowledge, cognitive load, spatial abilities, and motivation?". Both applications included the exact same information with 3D models of components of a combined cycle power plant and the potential to generate hypotheses about the efficiency and energy output based on different power plant configurations. The learners had the task to build the power plants to test the self-generated hypotheses. The hypothesis generation is based on simulations and the inquiry cycle as described by de Jong and van Joolingen (1998) with the aim that learners explore the possibilities and thus construct their knowledge of causal relations in the configuration and workings of the power plant. In the AR application, the power plant components can be accessed by scanning paper cards that function as AR markers, so that learners move the virtual objects when moving the paper cards with their hands. In the non-AR application, the components can be dragged and dropped into focus and moved with touch-based interaction on the screen. This shows a difference between the interactive potentials and experience the two kinds of applications provide, clearly showing the potentials of AR-based applications described in the characteristic interactivity (see Section 3.2). Another difference can be seen in the spatial presentation of the virtual objects. The AR application displays virtual objects inside the physical 3D space, providing learners with a more spatial representation in the context of the physical world than the non-AR version in which the elements are placed in a virtual space. This shows a difference between the spatial presentations the two kinds of applications provide, clearly showing the potentials of AR-based applications described in the characteristic spatiality (see Section 3.3). Due to the similarities between the applications, equivalence was expected concerning conceptual knowledge and cognitive load, but a difference was expected for motivational factors and spatial abilities. The study with N = 56 participants shows equivalence concerning conceptual knowledge and ICL, but not concerning ECL. The two groups did not differ concerning motivational factors or spatial abilities after the learning phase. AR is thus not necessarily better than a traditional tablet-based simulation, so that it is important to look more closely at specific factors that could influence learning in AR.

4.1.2 Study 2: Contextualised AR Support

In the second study, the focus was on contextuality as an individual factor. The goal of the study was to examine the influence of closeness of representation of textual information to real-world elements on effectiveness and efficiency in learning. The general research question can be formulated as “How does the closeness of virtual text to corresponding real-world images influence learning outcomes, task execution and cognitive load and how does the number of displayed elements influence this?”. To answer this question, an AR setting was simulated through photos and textual information in order to control the study situation. Participants in the study were shown textual cognitive group awareness information (see Bodemer et al., 2018) about different people on photos and had the task to decide which people they wanted as learning partners for specific learning objectives. The information were presented close to the photo or further away from it, considering the characteristic of contextuality (see Section 3.1) in the way that relevant textual information were either displayed as clearly belonging to their real-world context, or as clearly separate. Furthermore, there were different levels of difficulty in determining the learning partners, with differing numbers of people shown in the tasks. We hypothesised that closer information placement would lead to less ECL, lower time on task, and better retention of the information, especially when the task complexity was higher with more people being displayed. The study with N = 38 participants showed no effect of proximity of the information on self-reported ECL and information recall. There was, however, a difference in reaction time to a secondary task, measuring cognitive load during the task, and in time on task, with faster times for closer presentation. Number of people displayed did have a significant main effect on secondary task reaction time and time on task, but the pattern did not show a straight-forward increase with an increasing number. No interaction effects were found, thus not fully supporting the hypothesis. Task execution could thus be supported with closer information placement, but learning, which was not a goal in this task design, was probably not elicited. In order to learn more about specific learning processes and outcomes in AR, settings with a focus on learning need to be established based on the ARcis characteristics.

4.2 Paper 2: Interactivity in AR

Title: "Different Types of Interaction with Augmented Reality Learning Material" (Paper 2, Krüger & Bodemer, 2020)

The second paper extends the definition of interactivity as a characteristic of the ARcis framework. Specifically, potentials of interactivity in AR are described, including the elaborateness of the interaction with learning material (see Section *II. Three AR Characteristics* in Paper 2, Krüger & Bodemer, 2020). Further, differing perspectives on physical interaction in interactive learning environments are explored, describing the potential for a positive influence on learning or overload if not implemented in a purposeful way. We state that based on the literature interactive learning material should not only be physically interactive, but that this physical interaction should induce germane cognitive processes and thus mental interaction with the learning material (see Section *III.A. Interactive Learning* in Paper 2, Krüger & Bodemer, 2020). It is stated that with AR very elaborate interaction can be achieved, so that it might be especially important to examine how interaction should be designed purposefully. We define this elaborateness of interaction based on the three different levels of interaction in AR: purely physical interaction, purely virtual interaction, and interacting with virtual through physical elements. Also, it is described how this third form of interaction is made possible through the connection to physical AR markers (see Section *III.B. Interaction in AR* in Paper 2, Krüger & Bodemer, 2020). Based on this literature review, the research question and hypotheses on which Study 3 is based are formulated.

4.2.1 Study 3: Physical and Mental Interaction in AR

The third study is concerned with interactivity in AR-based learning experiences. Its goal was to examine the influence of physical and mental interaction on cognitive load and learning outcomes when learning in an AR-based experience. The research question for the study is "How do mental and physical interaction in AR learning material influence cognitive load, task load and learning outcomes?". The learners were asked to answer hypotheses about different power plants' energy output and efficiency. Mental interaction was manipulated through the instructions, which stated which power plants had to be compared to test the respective hypothesis (low mental interaction) or no information about that (high mental interaction). Physical interaction was manipulated through the set-up of the paper-based AR markers, which either had to be assembled into working power plants by the learners (high physical interaction) or were already assembled (low physical interaction). Based on the literature, we hypothesised that mental interaction supports learning as reflected by increased GCL and improved learning outcomes, but that physical interaction can have both a positive influence on GCL and learning outcomes, but also a negative influence on learning processes through ECL. The study with $N = 128$ participants shows no main effects of mental and physical interaction on knowledge but a significant interaction effect, showing higher knowledge in the groups where only either mental or physical interaction was high than in the groups where both were high, or both were low. Concerning both GCL no main or interaction effect was found and ECL did also not differ between groups. Moreover, concerning the exploratorily examined NASA TLX constructs no effects were found. However, the descriptive results suggest higher mental, physical, and temporal demand in the conditions with high mental and high physical interaction. From these results it can be concluded that both too little own activity due to much support and too much own activity due to little support may lead to less learning in an interactive AR environment. The results in terms of cognitive load were inconclusive in this study, but in observations of some of the learners it was seen that less support sometimes led to more errors and thus possibly misconceptions. In total, the results suggest that learners needed some support when learning with the interactive AR environments, as the combination of high mental and high physical interaction led to worse learning outcomes than when one of the interaction types was low. Thus, as is the case with simulations, it should be considered that learners need support, so that not all possibilities of interaction should necessarily be offered, but restrictions and guidance may be necessary.

4.3 Paper 3: Spatiality in AR

Title: "Learning with augmented reality: Impact of dimensionality and spatial abilities" (Paper 3, Krüger et al., 2022)

In the third paper, spatiality as a characteristic of the ARcis framework is extended. Specifically, factors of spatiality with a focus on the dimensionality of representations are described. In this context, the potential role of spatial abilities is also defined. The potentials of the dimensionality of representations in educational settings in general are examined in this paper, proposing that this has an influence on the construction of 3D mental representations in learners (see Section *1.1 Dimensionality of representations in education* in Paper 3, Krüger et al., 2022). We describe the specific case of 3D visualisations in AR, which are special due to increased depth cues in comparison to usual screen-based visualisations, even without stereoscopic AR headsets. The potentials of AR to support learning about spatial content are discussed (see Section *1.1.1 Augmented reality visualizations* in Paper 3, Krüger et al., 2022). Due to the potential influence of the dimensionality of the representation on cognitive load, findings about cognitive load when learning with 3D representations are presented, proposing that ECL can be decreased and GCL can be increased when learning with 3D representations in AR (see Section *1.1.2. Cognitive load* in Paper 3, Krüger et al., 2022). We further propose that spatial abilities need to be taken into account when learning with 3D representations, describing the ability-as-compensator and the ability-as-enhancer hypothesis as two potential, opposed approaches (see Section *1.1.3. Spatial abilities* in Paper 3, Krüger et al., 2022). Based on this literature review, the research questions and hypotheses on which Study 4 is based are formulated.

4.3.1 Study 4: Dimensionality and Spatial Abilities in AR

The fourth study is concerned with spatiality in AR-based learning experiences. Its goal was to examine the influence of the dimensionality of a virtual model and the influence of spatial abilities in learning with 3D objects. The research questions for this study are "How does the dimensionality of the visualisation of a 3D object in AR influence cognitive load and learning outcomes, and which role do spatial abilities play in this relationship?". The learners were asked to read a text about the human heart and look for mentioned components that were labelled in an AR-based graphical representation. Dimensionality was manipulated through this graphical representation, which was either a 3D model of a cross-section of the heart or a 2D image of that same cross-section. Learners' mental rotation abilities were measured before the learning phase. Based on the literature, we proposed that a 3D representation can support the learning of spatial aspects of a 3D object better than a 2D representation, particularly in individuals with low spatial abilities. We hypothesised that in general the 3D representation would lead to higher GCL and spatial knowledge and lower ECL than the 2D representation. We furthermore assumed that learners with lower spatial abilities would benefit more from the 3D representation than learners with higher spatial abilities. The study with N = 150 participants firstly shows a big effect of dimensionality on learners' perceived spatiality, but also medium effects on perceived contextuality and interactivity as part of a manipulation check aiming at the three ARcis characteristics. Concerning the hypotheses, scores for knowledge of the spatial relationships of the components of the heart and GCL were higher when learning with the 3D representation. Knowledge concerning spatial positions of the components and ECL did not differ significantly, but descriptively pointed in the expected direction. Knowledge concerning general aspects did not differ but was also not equivalent. ICL was equivalent in the two groups, as expected. Neither ECL nor GCL were found to be mediators for the relationship between dimensionality and knowledge. Concerning the role of mental rotation abilities, we found that, opposite to our expectations, learners with high abilities benefited from the 3D representation for their knowledge gain, while learners with low abilities did not. Two moderated mediation models showed no moderation effects of the relationship of dimensionality on ECL or GCL as part of their mediation of the relationship between dimensionality and knowledge. However, the mediation of this relationship through GCL was significant for learners with high but not average or low mental rotation abilities. In general, the study supports the assumption that spatial learning is better supported by 3D representations, but with the restriction that this is only the case for people with high spatial abilities. When using 3D representations in AR-based learning experiences, the spatial abilities of the learners should thus be taken into account, as not all learners may be able to learn with or handle 3D representations easily. A certain degree of spatial abilities might be necessary to process 3D representations in AR in an effective way to lead to learning, while 2D representations might not automatically lead to mental transformations into 3D representations.

4.4 Paper 4: Contextuality in AR

Original title: "Positioning augmented reality information for learning in nature: An exploratory pilot study" (Paper 4, Krüger & Bodemer, subm.)

In the fourth paper, contextuality as a characteristic of the ARcis framework is extended. Specifically, the linking of corresponding virtual and physical information is explored (see Section *1 Spatial* *integration in AR learning environments* in Paper 4, Krüger & Bodemer, subm.). We describe that on the small scale level, the physical anchor can be relevant for the learning material, and on the big scale level, the physical general environment can be relevant for the learning material (see Section 3.1.3). We make the connection of AR and situated learning as a relevant learning theory, describing how virtual elements can be situated within a relevant physical context through AR. The placement of the virtual information in AR is further connected to the construct of immersion, describing how people may be more immersed when the context is relevant to the learning material. Also, we describe how the placement of virtual information in a relevant context may increase learners' motivation, based on different motivational learning theories (see Section *1.1 Immersion and motivation in contextualized AR learning environments* in Paper 4, Krüger & Bodemer, subm.). Based on this literature review, the research questions on which Study 5 is based are formulated.

4.4.1 Study 5: Position of Physical Context in AR

The fifth study is concerned with contextuality in AR-based learning experiences. The goal is to examine effects of the position of virtual information to corresponding physical objects on cognitive and motivational learning processes and outcomes. The research question for the study is "How does the closeness of placement of thematically relevant learning material in a physical context in AR-based learning environments influence learning behaviour, processes and outcomes?". The participants received a tablet-based application with which they were asked to walk around a lawn outside. Here, images of different plants were placed and could be scanned with the tablet to receive information about the respective plant. They were asked to focus especially on differences and similarities between the plants to prepare for the knowledge test. Closeness of the relevant virtual elements to the physical surroundings was manipulated through the placement of the marker images. Those were placed either directly at the respective plants that were growing around the lawn (near condition), or they were placed in some distance on top of stones that lay in front of the lawn (far condition). Based on the literature, we hypothesised that the presentation of virtual information directly anchored to the corresponding physical object has a positive influence on learning processes and outcomes, especially with regard to motivation, immersion, and engagement, and that learning behaviour would differ. In the study with N = 19 participants various patterns of interaction with the application are shown. Their interview responses suggest that they (would have) focused more on the physical plants, (would have) felt more surrounded by the environment and were or would have been more motivated when receiving the material close to the physical plants instead of further away. Participants in the near condition stated that they compared and connected virtual and physical elements, while participants in the far condition paid mostly no attention to the physical plants. These results suggest that the closeness to the physical plants implicitly guided learners' attention to the plants and their connection to the virtual material, without actually instructing them to shift their focus. Experienced motivation and immersion also seem to be higher. When learning in a natural environment with a learning objective focusing on a connection between physical objects and virtual elements it may thus be important to place the virtual information close to the corresponding physical elements.

4.5 Paper 5: Multimedia Learning in AR

Original title: "Application and Investigation of Multimedia Design Principles in Augmented Reality Learning Environments" (Paper 5, Krüger & Bodemer, 2022a)

The fifth paper deals with an initial consideration of the transfer of the multimedia principles in Mayer's CTML (Mayer, 2020f) to AR environments. Specifically, it describes the potential integration of multimedia design principles into AR-based learning settings, connecting the spatial contiguity principle to the ARcis characteristic spatiality and the coherence principle to the ARcis characteristic contextuality. In the paper, CTML is defined as a relevant theory for AR-based learning experience due to its often multimedia-based nature (see Section *1.1. Multimedia Learning* in Paper 5, Krüger & Bodemer, 2022a). Furthermore, multiple sensory modalities can be used in AR-based learning experiences (see Section *1.2. Sensory Modalities* in Paper 5, Krüger & Bodemer, 2022a). Two prominent multimedia design principles are chosen that may be specifically relevant to AR due to its unique nature: spatial contiguity principle and coherence principle (see Section *1.3. Multimedia Design Principles* in Paper 5, Krüger & Bodemer, 2022a). The relevance of CTML for AR is further explained and a connection of the ARcis characteristics to the two multimedia principles is made (see Section *1.4. AR Characteristics* in Paper 5, Krüger & Bodemer, 2022a). Based on the two chosen multimedia principles, two learning experiences are designed and the effect of the adherence to these principles is examined in two studies, which are described in the following two sections.

4.5.1 Study 6: Spatial Contiguity Principle in AR

The sixth study approached the spatial contiguity principle as a multimedia design principle to incorporate into AR-based learning experiences based on the ARcis characteristic spatiality. Its goal was to examine the influence of adherence to the principle in AR on cognitive load, task load, and learning outcomes. The general research question can be formulated as "How does the spatial integration of virtual information into the view of the physical world (simulated AR) influence cognitive load, task load, and learning outcomes?". Because the study was applied in a controlled lab-based setting instead of on location, a simulated AR setting with a video of a walk through a botanical garden as the physical world and overlayed virtual information was shown to the participants. They were asked to imagine that they were really walking through that environment to look at the virtual information about the different plants focused on in the video. Spatial contiguity was manipulated in the video by placing virtual information about the plants either as an integrated overlay in front of the plants or in a separated view on a tablet screen next to the plants. Based on the literature, we hypothesised that the integrated representation would lead to decreased ECL and task load, and increased GCL and knowledge than the separated representation. The study with N = 80 participants shows no significant differences in ECL, GCL, knowledge, and most subconstructs of task load. Temporal demand, however, was lower, and perceived performance was higher for the integrated than the separated presentation. The patterns on all variables were descriptively as expected, but some differences were very small. From these results it can be concluded that, although most differences were not significant, the data suggest the expected pattern in which the integrated presentation is superior to the separated presentation. The differences in the designs seem to not have been substantial enough to have a meaningful influence on learning. The spatial contiguity principle seems to be partly transferable to a combined physical-virtual learning experience, although this should be examined again in a design with more complex learning material, more relevance of integrating virtual and physical elements mentally, and a real AR application.

4.5.2 Study 7: Coherence Principle in AR

The seventh study approached the coherence principle as a multimedia design principle to incorporate into AR-based learning experiences based on the ARcis characteristic contextuality. Its goal was to examine the influence of adherence to the principle in AR on cognitive load, task load, and learning outcomes. The general research question can be formulated as "How does the inclusion of matching or non-matching virtual audio information in a learning experience with virtual and physical visual elements influence cognitive load, task load, and learning?". Because the study was also applied in a controlled lab-based setting instead of the real world, a simulated AR setting with a video of a walk through a forest as the physical world and overlayed virtual pictures of and text about birds was shown to the participants. They were asked to imagine that they were really walking through that environment to look at the different birds and information in the video. Coherence was manipulated in the video by excluding or including virtual audio when the birds were shown, where included sounds were either the birds' tweeting sounds (matching the topic) or other sounds (not matching the topic). Based on the literature, we hypothesised that when no virtual sounds were added, ECL and task load would be decreased while GCL and knowledge would be increased in comparison to when sounds were added. Concerning the comparison between matching and non-matching sounds, we hypothesised that matching sounds would lead to increased GCL and knowledge, and decreased frustration as part of task load. The study with $N = 130$ participants shows no significant effects. The descriptive patterns of ECL, mental demand, temporal demand, and effort support the proposed hypotheses, with the lowest scores for the version without sound. The other variables had different descriptive pattern against the hypotheses. These results are in total inconclusive, with some of the descriptive patterns suggesting support for the assumed effects, some with opposite patterns, but none with significant differences. Again, the differences in the designs through the addition of small virtual sounds seem to not have been substantial enough to have a meaningful influence on learning. It needs to be further examined if the coherence principle is transferable onto combined physical-virtual learning experiences, for example when the sounds are more disruptive, the learning material is more complex, and a real AR application is used in the physical world. In Table 7, a summary of the content of all papers and studies is shown.

Table 7

Summary of Papers, Components, and Goals in This Dissertation

Papers	Component and Subgoal	Goal / Specific Research Question
Paper 1 *Basic Arcis Framework* (Krüger et al., 2019)	Framework: Basis of Arcis Framework *[Dissertation Subgoal 1]*	Literature review on research on AR in education to formulate and describe the three ARcis characteristics (**contextuality**, **interactivity** and **spatiality**), their role in educational settings and potential influence on learning, and questions to address in future research.
	Study 1: Interactive and Spatial Learning in AR *[Dissertation Subgoal 2, RQs 2 and 3]*	Examination of a tablet-based AR simulation following an inquiry-based learning cycle in comparison to a tablet-based non-AR version of the same simulation. Specific RQ: How does an AR in comparison to a non-AR version of a simulation influence knowledge, cognitive load, spatial abilities, and motivation?
	Study 2: Contextualised AR Support *[Dissertation Subgoal 2, RQ 1]*	Examination of the effectiveness and efficiency of the representation of virtual information close to real-world representations (AR mock-up) and with differing amounts of information for task execution. Specific RQ: How does the closeness of virtual text to corresponding real-world images influence learning outcomes, task execution and cognitive load and how does the number of displayed elements influence this?
Paper 2 *Interactivity in AR* (Krüger & Bodemer, 2020)	Framework: Extension of ARcis Framework concerning Interactivity *[Dissertation Subgoal 1]*	Further definition of **interactivity** as a characteristic of AR-based learning experiences with a focus on differing perspectives on physical interaction in learning environments, which is suggested to either have a generally positive influence on learning, or to generally lead to overload if not implemented in a purposeful way.
	Study 3: Physical and Mental Interaction in AR *[Dissertation Subgoal 2, RQ 2]*	Examination of the role that physical and mental interaction plays when learning with an AR-based experience. Specific RQ: How do mental and physical interaction in AR learning material influence cognitive load, task load and learning outcomes?
Paper 3 *Spatiality in AR* (Krüger et al., 2022)	Framework: Extension of ARcis Framework concerning Spatiality *[Dissertation Subgoal 1]*	Further definition of **spatiality** as a characteristic of AR-based learning experiences with a focus on the dimensionality of representations, which is suggested to have an influence on the construction of 3D mental representations. In this context, the potential role of spatial abilities is also defined.
	Study 4: Dimensionality and Spatial Abilities in AR *[Dissertation Subgoal 2, RQ 3]*	Examination of the influence of the dimensionality of a representation of a 3D object in AR on learning processes and outcomes while considering the effects of spatial abilities. Specific RQ: How does the dimensionality of the visualisation of a 3D object in AR influence cognitive load and learning outcomes, and which role do spatial abilities play in this relationship?
Paper 4 *Contextuality in AR* (Krüger & Bodemer, subm.)	Framework: Extension of ARcis Framework concerning Contextuality *[Dissertation Subgoal 1]*	Further definition of **contextuality** as a characteristic of AR-based learning experiences with a focus on placing relevant virtual elements into physical environments in AR-based learning settings, which is suggested to have an influence on immersion, motivation, effort, and load.

Papers	Component and Subgoal	Goal / Specific Research Question
	Study 5: Position of Physical Context in AR *[Dissertation Subgoal 2, RQ 1]*	Examination of the influence of the closeness of virtual information to relevant physical objects in an AR-based real-world experience. Specific RQ: How does the closeness of placement of thematically relevant learning material in a physical context in AR-based learning environments influence learning behaviour, processes and outcomes?
Paper 5 *Multimedia Learning in AR* (Krüger & Bodemer, 2022a)	Framework: Extension of ARcis Framework concerning application of Multimedia Principles *[Dissertation Subgoals 1 and 3]*	Description of potential integration of multimedia design principles from the CTML into AR-based learning settings, connecting the spatial contiguity principle to **spatiality** and the coherence principle to **contextuality**.
	Study 6: Spatial Contiguity Principle in AR *[Dissertation Subgoals 2 and 3, RQ 4]*	Examination of the influence of the spatial integration of virtual and physical elements in combined physical-virtual environments following the spatial contiguity principle. Specific RQ: How does the spatial integration of virtual information into the view of the physical world (simulated AR) influence cognitive load, task load, and learning outcomes?
	Study 7: Coherence Principle in AR *[Dissertation Subgoals 2 and 3, RQ 4]*	Examination of the influence of excluding potentially incoherent elements in combined physical-virtual environments following the coherence principle. Specific RQ: How does the inclusion of matching or non-matching virtual audio information in a learning experience with virtual and physical visual elements influence cognitive load, task load, and learning?

4.6 Overview of Methods in the Empirical Studies

In the following, I will provide an overview of the methods used in the seven studies. After an overview of the different samples, I will give an overview of the design, including the manipulated variables and measured variables in the studies, the general procedure, material, and data analyses used. For further details on the methods, the specific papers can be consulted.

4.6.1 Samples

In the seven studies, different samples were used. An overview over the sample characteristics can be found in Table 8 and Table 9. For more details on differences between the conditions, the respective papers can be consulted. In all studies, a convenience sampling method was chosen, mainly advertising the study to students who were enrolled in the study programmes "Applied Cognitive and Media Science" and "Psychology" at the University of Duisburg-Essen. They received participant hours when taking part in the studies, which was the compensation in all studies. Only in Study 3, a money-based compensation was a possible alternative. Not all sample characteristics were reported or collected in all studies, but if they were collected without being reported, they were added here for completeness.

The number of participants in the studies differed greatly, reaching from $N = 19$ (Study 5) to $N = 150$ (Study 4) with a total of 601 participants over all seven studies. For Study 5, the sample characteristics data of only 8 of the 19 participants are available, due to data lost in a cyberattack. Age and gender of the participants can be seen in Table 8. The age of the participants was 22.5 years on average, with a range between 17 and 61 as some convenience sampling took place in private circles of the investigators. In all studies, more female than male identifying people took part. This is mainly due to the nature of the study programmes in which the sampling took place, in which more female than male students are enrolled. The table shows that on average 95% of participants were students, with at least 86% in all studies. Most of them came from the study programmes mentioned above.

Table 8

Sample Characteristics: Age, Gender, Job

Study	*N*	age		gender		students
		min - max	*M* (*SD*)	m	f	%
1	56	18 - 34	22.13 (3.27)	16	40	100%
2	38	18 - 32	21.95 (3.38)	5	33	100%
3	128	18 - 40	22.55 (3.90)	39	89	96%
4	150	17 - 31	21.81 (2.98)	41	109	96%
5	19[a]	18 - 33	22.62 (4.52)	3	16	100%
6	80	17 - 33	22.21 (3.14)	20	60	95%
7	130	18 - 61	23.72 (7.95)	34	96	86%
Total	601	17 - 61	22.50 (4.81)	158	443	95%

Note. In some studies, single variables were not reported but have been added here for completeness.
[a] data for age available only for 13 of 19 participants

In addition to the general demographics described in the previous paragraph, in nearly all studies additional sample characteristics were collected (see Table 9). On one hand, the amount of experience with the technology was assessed. This included questions about how often participants had used general mobile applications (“Mobile”), mobile learning applications (“M learn.”), general mobile AR applications (“AR”), and mobile AR learning applications (“AR learn.”) on tablets or smartphones in the past. These were answered in a five-point response format: “never” (1), “rarely” (2), “now and then” (3), “often” (4), “regularly” (5). In total, mobile devices in general were used regularly by the participants in the different samples, while mobile learning application were used between now and then and often. Experience with AR applications was low, with usage from never to rarely. AR learning applications had been used even less often. This shows that the samples were generally less used to handling AR. Besides these technology-focused experiences, the self-reported knowledge beliefs, task expectancy, and value were collected prior to the learning tasks. These were also answered in a five-point response format from 1 (low) to 5 (high). Knowledge belief was below the middle of the scale for all studies, while expectancy and value were around the middle of the scale for most studies. This shows that in general the participants in the samples did not believe that they were very knowledgeable in the topics and on average did not bring full motivation for the learning task.

Table 9

Sample Characteristics: Technology Experience and Expectancy-Value Questionnaire

Study	Experience with technology				Expectancy-value questionnaire		
	Mobile ***M (SD)***	**M learn.** ***M (SD)***	**AR** ***M (SD)***	**AR learn.** ***M (SD)***	**Belief** ***M (SD)***	**Expectancy** ***M (SD)***	**Value** ***M (SD)***
1	4.66 (0.55)	2.75 (1.07)	1.80 (0.90)	1.21 (0.56)	-	-	-
2	-	-	-	-	-	-	-
3	4.70 (0.61)	2.48 (1.07)	1.58 (0.84)	1.09 (0.31)	1.21 (0.33)	2.37 (0.60)	2.03 (0.60)
4	4.65 (0.86)	2.37 (1.11)	1.77 (0.75)	1.21 (0.53)	1.95 (0.63)	2.87 (0.70)	3.22 (0.72)
5 [a]	5.00 (0.00)	3.50 (1.20)	1.75 (0.71)	1.63 (0.74)	1.46 (0.43)	2.44 (0.62)	2.65 (0.51)
6	4.79 (0.50)	3.08 (1.35)	2.23 (0.98)	1.69 (0.96)	1.91 (0.73)	2.78 (0.70)	2.69 (0.86)
7	4.87 (0.42)	2.96 (1.15)	1.81 (0.81)	1.31 (0.68)	1.59 (0.50)	2.69 (0.71)	2.40 (0.65)
Total	4.74 (0.63)	2.69 (1.18)	1.80 (0.85)	1.28 (0.64)	1.65 (0.70)	2.67 (0.70)	2.60 (0.83)

Note. In some studies, single variables were not reported but have been added here for completeness. The dashes show when data have not been collected in that study.
[a] data available only for 8 of 19 participants

4.6.2 Design and Variables

The designs of the studies in general included a 2x9 mixed design in Study 2, a 2x2 between-subjects design in Study 3, a three-groups between-subjects design in Study 7, and a between-subjects designs with two groups in all other studies. An overview of the manipulated variables in the different studies can be found in Table 10. In Study 1, factors on interactivity and spatiality were manipulated in a comparison of an AR and a non-AR tablet-based application. In Study 2, Study 5, and Study 7, factors concerning contextuality were manipulated, with foci on the closeness of virtual and physical or real-world elements and the inclusion or exclusion of virtual sounds. In Study 3, factors concerning interactivity were manipulated, with a focus on mental interaction and physical interaction. In Study 4 and Study 6, factors concerning spatiality were manipulated, with foci on the dimensionality of representation and the spatial integration of physical and virtual elements.

Different variables were measured in the seven studies (see Table 11 for an overview). Variables measured before the learning phase and reported as sample characteristics were knowledge belief, task expectancy, and value (Studies 3, 4, 5, 6, and 7), and pre-knowledge (Studies 3 and 5). Mental rotation abilities were also measured before the learning phase and examined as a moderating learner characteristic in Study 4. Variables on cognitive load were measured and used as dependent variables in all studies, although with different types in focus: Study 1 focused on ECL and ICL, Study 2 on ECL, Study 5 on GCL, Study 3 on ECL and GCL, and Studies 4, 6, and 7 analysed all three types of cognitive load. Task load was also analysed in Studies 3, 6, and 7. In all four studies, all six subconstructs were assessed (mental demand, physical demand, temporal demand, performance, effort, and frustration). Motivation was assessed in Study 1 and 5, with a focus on intrinsic motivation in Study 1 and design-based motivation in Study 5. Immersion with the subconstructs interest, usability, emotional attachment, focus of attention, presence, and flow was measured in Study 5. The ARcis experience with perceived contextuality, interactivity, and spatiality as subconstructs was measured and used as a manipulation check in Study 4. Mental rotation abilities, which were analysed as a moderator in Study 4, were measured and examined as a dependent variable in Study 1. A task behaviour variable of time on task was collected in Study 2, and behaviour during a learning task was analysed in Study 5. All studies included learning outcome measures in the form of knowledge tests with different types of knowledge in focus. In Figure 6, all independent variables clustered by ARcis characteristic, the dependent learning process variables which were hypothesised to be influenced in the different studies, and the learning outcomes are shown. For all of the learning process variables, it is expected that they ultimately have an influence on the resulting learning outcomes in the forms of, for example, conceptual or spatial knowledge, although this mediating relation is not always formulated into hypotheses and tested.

Table 10

Independent Variables Manipulated in the Studies in This Dissertation

Study	ARcis characteristic	Independent variable and manipulation
Study 1	Interactivity & Spatiality	*Form of representation of virtual animated models, 2 groups design:* a) hand-interaction-based 3D AR representation b) touch-interaction-based pseudo-3D non-AR representation
Study 2	Contextuality	*Representation of textual information about real-world elements, 2x9 design:* 1. closeness of representation: a) close representation b) separate representation 2. number of elements (within): 2 – 10 elements
Study 3	Interactivity	*Elaborateness of interaction with AR-based materials, 2x2 design:* 1. mental interaction: a) high interaction b) low interaction 2. physical interaction: a) high interaction b) low interaction
Study 4	Spatiality	*Dimensionality of representation of a virtual object in AR and relation to spatial abilities, 2 groups design:* a) three-dimensional representation b) two-dimensional representation
Study 5	Contextuality	*Closeness of placement of virtual information to relevant physical objects, 2 groups design:* a) near placement b) far placement
Study 6	Spatiality	*Spatial integration of representations of virtual and physical elements, 2 groups design:* a) spatially integrated representation b) spatially isolated representation
Study 7	Contextuality	*Inclusion of virtual sounds in an AR-based setting, 3 groups design:* a) exclusion of sounds b) inclusion of matching sounds c) inclusion of non-matching virtual sounds

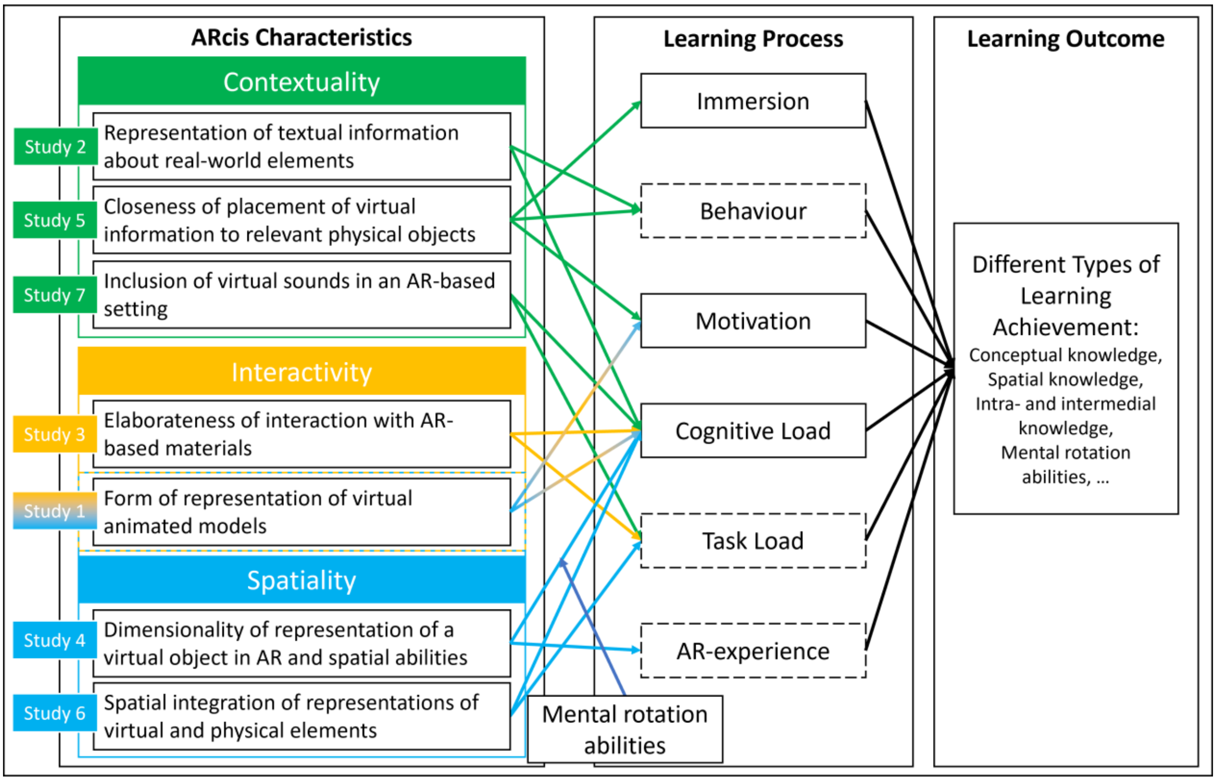
Table 11

Measured Variables Reported in the Studies in This Dissertation

Variable	Study 1	2	3	4	5	6	7
Pre							
Belief, Expectancy, Value			×	×	×	×	×
Pre-Knowledge			×		×		
Mental Rotation Abilities				×			
Dependent Variables							
Learning Task Behaviour		×			×		
Cognitive Load	×	×	×	×	×	×	×
→ ECL	×	×	×	×		×	×
→ ICL	×			×			
→ GCL			×	×	×	×	×
Task Load			×			×	×
Motivation	×				×		
→ Intrinsic	×						
→ Design-based					×		
Immersion					×		
ARcis Experience				×			
Mental Rotation Abilities	×						
Learning Outcomes	×	×	×	×	×	×	×

Figure 6

ARcis Characteristics, Factors, and Variables per Study

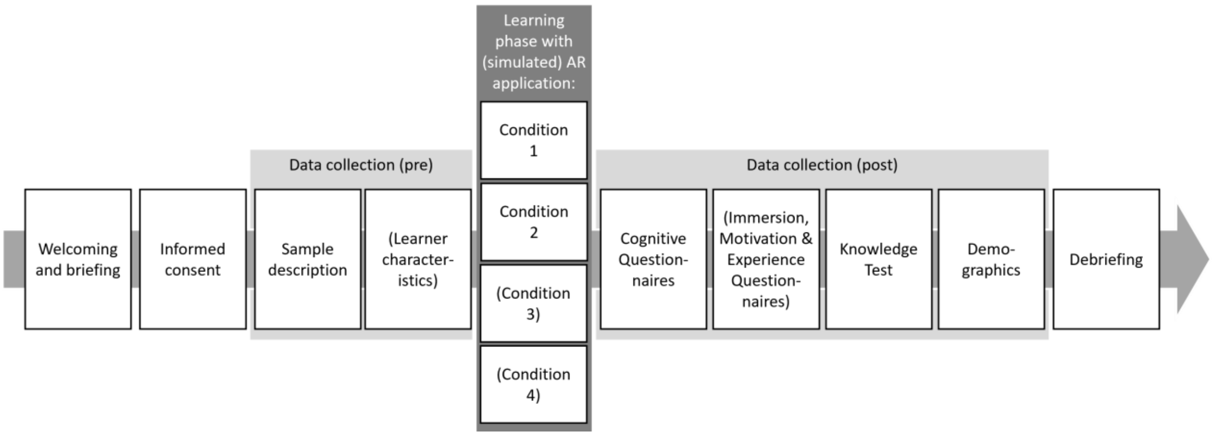


4.6.3 Procedure

In general, the procedures of the seven studies were very similar, with a summary shown in Figure 7. The studies all started with a welcoming and briefing by the researcher in the beginning, and informed consent was given in each of the studies. Then, pre-learning phase variables were measured for sample description (knowledge beliefs, task expectancy, value, and pre-knowledge test) or as learner characteristics (mental rotation test). Then, the learning phase with different AR-based applications took place, in which the different manipulations described in Table 10 were applied. In some studies, data were collected during this learning phase. Afterwards, data collection for the dependent variables took place. First, the cognitive and task load questionnaires were administered, then immersion, motivation, and AR experience questionnaires if applicable. After that, knowledge tests were applied. In the end, demographic data were collected, the participants were debriefed and dismissed. The specific procedures can be found in the respective papers.

Figure 7

Summarised Study Procedure



4.6.4 Measurement

In the seven studies, different questionnaires and tests were used to measure the sample characteristics and independent variables. Sample characteristics in the form of knowledge belief, task expectancy and value were measured with the expectancy-value questionnaire (Wigfield & Eccles, 2000), adapted to the respective learning subject and translated to German. The former "ability belief" scale was reformulated to instead ask for knowledge belief because of the nature of the learning objectives not being about abilities but about knowledge. Learners' pre-knowledge was measured in both Study 3 and Study 5 with self-developed tests for basic concepts of the respective topic. Spatial abilities in the form of mental rotation abilities were measured with a mental rotation test (Peters et al., 1995) in both cases.

In order to measure cognitive load with its subconstructs ECL, ICL, and GCL, three different cognitive load questionnaires were administered. In Study 1, the questionnaire by Leppink et al. (2015) with its ECL and ICL subscales, and in Study 2 the ECL subscale of the questionnaire by Leppink et al.

(2013) were used, both translated to German. In the other studies, the German cognitive load questionnaire by Klepsch et al., (2017) was used, either only the GCL and ECL subscales (Studies 3, 6, and 7), or all three subscales (Studies 4 and 5). Task load with its subconstructs mental demand, physical demand, temporal demand, performance, effort, and frustration was measured with the NASA TLX (Hart, 2006; Hart & Staveland, 1988) translated to German. Motivation was measured with the situational intrinsic motivation scale (SIMS; Guay et al., 2001) with the subscales intrinsic motivation, identified regulation, external regulation, and amotivation in Study 1. In Study 5, design-based motivation with the subconstructs attention, relevance, confidence, and satisfaction based on the ARCS-model (Keller, 2010) was measured with the reduced instructional materials motivation survey (RIMMS; Loorbach et al., 2015). Immersion with the subconstructs interest, usability, emotional attachment, focus of attention, presence, and flow was measured with the AR immersion questionnaire (ARI; Georgiou & Kyza, 2017c) in Study 5. The ARcis experience with perceived contextuality, interactivity, and spatiality as subconstructs was measured with a first version of the self-developed ARcis questionnaire (Krüger & Bodemer, 2022b). As task behaviours, time on task was measured in Study 2. Tracking of interaction with the AR application through event logging during the learning phase was used for behavioural measures in Study 5.

All studies included learning outcome measures in the form of knowledge tests with different types of knowledge in focus. They were all measured through knowledge tests that were self-developed systematically on the basis of the learning material. In Study 1, conceptual knowledge about the combined cycle power plants presented in the learning phase was tested with 20 multiple choice items. In Study 2, the previously presented cognitive information about people was asked in seven items spread between 16 tasks, asking questions about the people shown in pictures, six with multiple choice questions and one with an open text field. Study 3 included a total of 17 multiple choice questions, with five questions about the power plant components, five questions about power plant efficiency and energy output, and seven questions for knowledge transfer. In Study 4, a difference was made between spatial knowledge and general knowledge. Three different knowledge scores were measured through questions about the spatial placement of components of the human heart (spatial: components), the spatial relations of these components (spatial: relations), and general aspects without relation to spatial concepts (general). Study 5 included 30 questions in different formats to measure conceptual knowledge, including recall, recognition, and transfer questions, multiple choice, multiple response, and open questions, and questions including texts and pictures as questions and answer options. In Study 6, ten multiple choice questions were used to measure conceptual knowledge, with the questions further differentiated into five items including pictures of plants and text-based answer options (picture-text items), and five items including names of plants and text-based answer options (text-text items). The same was the case for Study 7, with a total of eight multiple choice questions, four picture-text and four text-text items. This shows the variability of measures of knowledge used in the different studies, including very different knowledge concepts based on the respective learning objectives.

4.6.5 Data Analyses

In the seven studies, mainly quantitative data analysis was applied. The most used inference statistical approaches were difference tests, including independent samples *t*-tests for testing of difference hypotheses in Studies 1, 2, 3, and 6, and Mann-Whitney *U* tests due to non-normal data distribution in Study 4. Further, analyses of variances (ANOVAs) were executed in Study 2 (factorial 2x9 mixed ANOVAs), Study 3 (factorial 2x2 between-subjects ANOVAs), and Study 7 (1x3 between-subjects ANOVAs with a-priori-defined Helmert contrasts). In addition to difference hypotheses, there were also some equivalence hypotheses, which were tested through two one-sided *t*-tests (TOSTs) with equivalence bounds set based on the smallest detectable effect for the respective sample size. In some studies, we further looked at correlations between variables, mostly exploratorily, using Pearson's *r* correlations in Study 2, and Kendall's τ correlations due to non-normal data distribution in Study 3. In order to explain potential interplays between variables, mediation analyses, moderation analyses, and moderated mediation analyses were executed in Study 4. In addition to quantitative analyses, we furthermore used qualitative data in Study 5. On one hand, statements from interviews were clustered and analysed. On the other hand, tracking data from the usage of the application were analysed qualitatively (e.g., order of events) and quantitatively (e.g., count of events).

4.7 Integration of Empirical Results

The seven studies described in Sections 4.1 to 4.5 provide interesting insights into learning with AR. In the following, I will integrate these results in accordance with the general research questions formulated in the introduction to Section 4, forming them into a complete picture within the ARcis framework based on the structure in Figure 6. This will provide a basis for the general discussion in Section 5.

Studies 2, 5, and 7 examined the ARcis characteristic contextuality, thus providing insights for RQ1 "*How does contextuality influence learning in AR?*". While Study 2 and 5 examine the closeness of representation of virtual and physical elements, Study 7 examines the implementation of the coherence principle in AR. The contextual closeness of group awareness information to people for a group formation task in Study 2 was shown to not have an influence on self-reported ECL and information recall, but to have a positive effect on reaction time on a secondary task and time on task. These results suggest that task execution may be supported by the closeness of the contextually corresponding information, although learning may not be directly supported. While Study 2 focused more on task execution and less on learning, Study 5 was executed with a learning-specific focus. In its exploratory approach, no definite conclusions can be drawn yet. However, the results show first indications that when information is embedded into the physical environment by placing virtual information directly at corresponding objects and thus leveraging AR-specific contextuality, the physical objects and their connection with the virtual elements are attended to more closely. Furthermore, immersion, motivation, and engagement with the physical part of the learning material are indicated to be increased by closer placement. This shows potential positive effects on cognitive and affective aspects of the learning situation when considering contextuality. In Study 7, the focus was on the potential pitfalls of learning in a contextually rich environment, which may violate the coherence principle. When adding virtual sounds that matched or did not match the material, however, no definite pattern of an increased cognitive load and task load was shown, so that the addition of small sounds may not have had that much of a negative influence on learning processes. Still, also not a positive influence of adding matching sounds was found concerning GCL, which was expected due to potential motivational effects. Concerning knowledge there were also no significant differences. In conclusion, the results concerning RQ1 on the influence of contextuality on learning in AR indicate that contextually relevant placement of information might impact task execution, cognitive, and affective learning factors positively, and that not all addition of corresponding virtual information leads to cognitive overload.

Studies 1 and 3 examined the ARcis characteristic interactivity, thus providing insights for RQ2 "*How does interactivity influence learning in AR?*". While Study 1 included the form of physical interaction as a variable, Study 3 specifically focused on the interaction of mental and physical interaction. In Study 1, an AR-based implementation of hand-interaction with 3D representations led to equivalent factual knowledge and ICL, and not equivalent but also not different ECL, than a non-AR implementation of touch-interaction with pseudo-3D representations. Spatial abilities and motivation did not differ after the learning phase. AR is thus not necessarily better than non-AR, and a closer look at specific mechanisms of interactivity is necessary. In Study 3, we took a closer look, focusing specifically on the distinction between and interaction of physical and mental interactivity of AR-based learning material. We found an unexpected interaction effect, showing higher knowledge when either mental or physical interaction was high, but not when neither or both were high. While this effect was not apparent for ECL, GCL, or the NASA TLX subconstructs, their descriptive results show a similar pattern. The results suggest that in AR it may be necessary to provide some interactivity to induce learning processes, but to limit the interactivity to prevent potential overload. In total, indications could be found that learner-controlled interactivity may thus play a role in supporting learning, while also having the potential to disrupt learning.

Studies 1, 4, and 6 examined the ARcis characteristic spatiality, thus providing insights for RQ3 "*How does spatiality influence learning in AR?*". While Study 1 included the form of virtual 3D presentation as a variable, Study 4 specifically examined the dimensionality of the virtual object also considering spatial abilities, and Study 6 examined an AR-based implementation of the spatial contiguity principle. In Study 1, not only the interactivity was different for the two implementations, but also the 3D representation, which was more spatial in the AR implementation and more pseudo-spatial in non-AR. The results showing equivalence for some variables and no difference for others, as described above, thus also necessitate a closer look at spatiality in AR. In Study 4, we took a closer look, comparing the impact of a virtual 3D in comparison to a virtual 2D representation of a spatial object in AR. Different types of knowledge were influenced differently, with only knowledge about spatial relationships of object components being impacted positively through the 3D representation but not knowledge about spatial positions and general knowledge. GCL was also higher with the 3D representation, but the relationship between dimensionality of representation and knowledge was not mediated by GCL. Also, ECL did not differ and did not mediate the relationship between dimensionality and knowledge. However, a moderation of this relationship through spatial abilities was found, showing that learners with higher spatial abilities profited from the 3D visualisation for their knowledge gain, while learners with lower spatial abilities did not. A moderated mediation analysis showed no significant moderation of the relationship of dimensionality and cognitive load, suggesting that the moderation of the relationship between dimensionality and knowledge was not attributable to differences in cognitive load for learners with different spatial abilities. In Study 6, the focus was on spatial integration of physical and virtual learning material. While there were no significant differences in ECL, GCL, most NASA TLX subconstructs, or knowledge, the descriptive pattern suggests a positive effect of the integrated in comparison to the separated representation. In total, indications could be found that spatiality in the form of 3D representations may help support learning about spatial relationships but not necessarily about other types of knowledge, although a certain level of spatial abilities seems to be necessary to profit from the representation. Furthermore, spatiality in the form of spatial integration suggests potentials for improving learning processes and outcomes.

Study 6 and 7 additionally provided insights for RQ4 "*How does the implementation of multimedia design principles in AR influence learning?*". While Study 6 examined the spatial contiguity principle, Study 7 examined an AR-based implementation of the coherence principle. The results of the implementations as described above were different for the two studies, showing a general tendency towards supporting the positive influence of following the spatial contiguity principle, while showing a less clear picture concerning the influence of following the coherence principle.

Additional outcomes that did not inform the research questions but provide interesting insights into different aspects concerning the dependent variables were gained. In Study 3 on mental and physical interaction we looked at the correlations between NASA TLX subfactors and cognitive load types. We found moderate positive correlations between GCL and mental demand, GCL and effort, and ECL and frustration. Furthermore, a moderate negative correlation was found for mental demand and knowledge test score. This shows potential connections between cognitive load, task load, and knowledge.

In general, the studies show that a systematic examination of specific characteristics and mechanisms of AR can supply insights about its effective and efficient use within learning environments. From insights in this area of research, recommendations for design and implementation can be derived, which can have a direct influence on educational practice. In the following section, the results of the studies will be discussed within the theoretical background and the ARcis framework.

5 Discussion

As stated in Section 1.3, the goal of this dissertation is to gain insight into specific characteristics of AR-based learning and how they can be leveraged to support learning processes and outcomes. Three subgoals to reach this overarching goal were formulated, a theoretical, an empirical, and a practical subgoal. In the following, I will summarise how these subgoals were reached in the current dissertation as a basis for further discussion of the results concerning these goals within the theoretical background.

For Subgoal 1, *"theoretically defining characteristics of learning with AR and analysing how specific mechanisms may have an impact on learning"*, I described the ARcis framework in Section 3, including the three characteristics contextuality, interactivity, and spatiality. Those characteristics describe the combination of virtual and physical elements, the potentials of material manipulation and interaction, and the spatial properties of objects as unique assets of AR-based educational environments that can be used to support specific learning objectives. Different mechanisms that might play a role for this support have been identified and described, summarised in Section 3.5.

For Subgoal 2, *"empirically examining how the characteristics of learning with AR and their specific mechanisms influence learning"*, I formulated four research questions in Section 4, including three questions based on the ARcis characteristics and one question on the evaluation of multimedia design principles in AR. The seven studies executed in the five papers included in the current dissertation provide first answers to these research questions, which have been summarised in Section 4.7.

For Subgoal 3, *"practically applying the theoretical and empirical insights into the characteristics and mechanisms for the design of AR-based learning experiences"*, I described the application of the three ARcis characteristics in different settings in Section 3 and the implementation and evaluation of the multimedia design principles in AR in Study 6 and 7 in Section 4. A more integrated discussion of this will in the following be presented in Section 5.3.

In the following sections, theoretical insights and empirical findings from the current dissertation will be integrated and their theoretical and practical implications will be discussed. Afterwards, limitations and an outlook on future research will be given.

5.1 Integrated Discussion of Theory and Empirical Results

In Section 2, I described different theories and constructs that are important to consider when examining learning with AR. In the following, I will integrate the insights from the current dissertation concerning the influence of AR focused on these areas: MERs, learning achievement, cognitive load and workload, immersion, motivation, and spatial abilities.

5.1.1 Multiple External Representations

Concerning MERs in AR as described in Section 2.1, the current dissertation in general supports the idea that one of the unique potentials of AR is the combination and integration of virtual and physical elements. The ARcis characteristic contextuality builds on this aspect of AR, describing how learning can be situated in an authentic environment and the relation of physical and virtual elements can be supported through an integrated representation. Virtual and physical elements may provide complementary information (see also Section *4. Discussion* in Paper 4, Krüger & Bodemer, subm.), which is one of the functions described in the DeFT framework (Ainsworth, 2006). This was also supported by the results in Study 5 (Paper 4, Krüger & Bodemer, subm.), showing that learners specifically describe that they attended more to the physical objects when the corresponding virtual elements were accessed directly at them. An integration of these could help learners construct a more complete mental model of the content, especially when the learning material is focused on the relevance of the information in the context of a real-world environment or activity.

The second ARcis characteristic, interactivity, highlights that representations in AR can be interactive in various ways, including manipulation of physical objects, manipulation of virtual elements, and mediated manipulation of virtual elements through the manipulation of physical objects. As described by Moreno and Mayer (2007), interactive learning environments can be relevant for knowledge construction, although this does not happen automatically and a distinction between behavioural and cognitive activity is necessary (see also Section *III. Interactivity in Learning Contexts* in Paper 2, Krüger & Bodemer, 2020). The authors further describe that guidance is necessary to support knowledge construction, which is also supported by the results in Study 3 (Paper 2, Krüger & Bodemer, 2020) which show that combining high mental and high physical interaction in AR-based learning material may not lead to the best learning outcomes. Interactive representations in AR-based learning environments thus need to be well-designed and might then support the construction of knowledge.

The third ARcis characteristic, spatiality, brings the attention to the possibility to incorporate true 3D visualisations in AR and anchor them in physical space. 3D representations can be classified as depictive (see integrated model of text and picture comprehension; Schnotz & Bannert, 2003) and pictorial (see CTML; Mayer, 2020a), and can specifically support learning about spatial objects or topics (see also Section *1.1. Dimensionality of representations in education* in Paper 3, Krüger et al., 2022). This is also supported by the results in Study 4 (Paper 3, Krüger et al., 2022), which show a positive effect of a 3D visualisation on learning of spatial relations of object components. Depicting 3D representations in AR can thus support learning, especially when the learning material focuses on spatial knowledge. In total, when extending the models and theories on MERs onto the specific case of MERs in AR, it is thus necessary to consider the specific potentials of combined virtual and physical representations that can be interactive and spatial.

5.1.2 Learning Achievement

Concerning learning achievement in AR as described in Section 2.2, a basic result found in Study 1 (Paper 1, Krüger et al., 2019) shows that AR is not necessarily better for learning outcomes than non-AR, supporting the current mixed results in the research area. This also supports the general criticism concerning media comparison studies and the necessity to identify specific characteristics and mechanisms that affect learning with AR. Furthermore, the insights in the current dissertation support the notion that learning objectives need to be considered for AR-based learning. Looking at the three characteristics, it is clear that they can be leveraged for different learning objectives.

Contextuality mainly includes the possibility to situate learning in an authentic environment, offering learners the opportunity to be physically and mentally involved in a real-world context. A learning objective that is particularly suited for this characteristic involves the mental integration of physical and virtual information, which can complement and enrich each other. Looking at the results from Study 5 (Paper 4, Krüger & Bodemer, subm.), participants reported paying more attention to physical objects and their connection to the virtual elements when they received the virtual information directly at the corresponding physical elements. Furthermore, they described complementary information and functions of the two types of elements, highlighting the authenticity and inherently multisensory nature of the physical objects and environments and on the other hand mentioning the potential to show unchanging texts and pictures virtually. This potential of integrating complementary physical and virtual information in AR can be promising for various areas of education that include the necessity to connect a theoretical input with a practical application or real-world situation.

Interactivity includes the possibility for active learning, including simulation-based or discovery learning, giving learners the opportunity to explore causalities or situations themselves. A learning objective that is particularly suited for this characteristic is learning of processes and complex causal relationships that can be supported by active learning opportunities. Looking at the results of Study 3 (Paper 2, Krüger & Bodemer, 2020), learners' knowledge was lower when there was only low mental and low physical interaction and also when high mental and high physical interaction were combined. In comparison, when only either mental or physical interaction was high, knowledge was increased. This shows the necessity to guide learners' interaction with AR-based learning materials, but also the potential to increase learning when some interaction is included. The interaction with learning material in AR can be promising for various areas of education in which processes and causal relationships are best discovered and experienced by learners themselves.

Spatiality includes the possibility for spatial learning of complex 3D structures, with suitable learning objectives covering the understanding of spatial structures, positions of components within an object, and spatial relations between those components. The results from Study 4 (Paper 3, Krüger et al., 2022), specifically showed that the dimensionality of a visualisation in AR had no influence on general knowledge and spatial component positions knowledge, but for spatial component relations knowledge learning with the 3D model led to better outcomes than learning with the 2D graphic. This shows that the type of knowledge needs to be considered and that especially spatial knowledge that is more complex and involves more elements, like knowledge about relations, may be supported when using 3D representations. This potential of displaying 3D objects in AR can be promising for various areas of education in which spatial objects and their structure are the focus of the learning content. In Study 6 (Paper 5, Krüger & Bodemer, 2022a), the spatial integration of physical and virtual objects as another part of spatiality showed no differences in learning if material was integrated or separated. In this context, we raised the issue of a consideration of local and global coherence construction based on Seufert and Brünken (2006), which in AR includes not only coherence within one form of or across forms of representation, but also coherence within one world (virtual or physical) or across worlds. Different types of coherence formation as a part of knowledge construction may be supported by different combinations of virtual and physical elements in AR. In total, it is important to consider the learning objective before considering how it can be best supported by AR. The separation into the three characteristics can help in structuring the potentials of AR-based learning experiences.

5.1.3 Cognitive Load and Workload

Concerning cognitive load and task load, as described in Section 2.3, different implications need to be considered based on the results from the current dissertation, including all three ARcis characteristics. Looking at contextuality, an important aspect that was mentioned in Section 3 includes the potentially distracting effects of placing instructional elements within a broader physical context that can include sounds, smells, and physical objects that are not part of or relevant for the learning material. In Study 5 (Paper 4, Krüger & Bodemer, subm.) learners indeed reported that the closeness to physical objects pulled their attention to those. In this case, the objects were interesting for the learning material, although they were not relevant for the knowledge test, so that they could have pulled attention away from more relevant aspects, which is in accordance with the immersion principle (Mayer, 2020e). This way, ECL might be induced, and attention may be pulled from more relevant processes. As described in the literature on seductive details, learning may be hindered when interesting but irrelevant information is added towards learning material (Sundararajan & Adesope, 2020), although affective components also need to be considered (Park et al., 2015). In Study 7 (Paper 5, Krüger & Bodemer, 2022a), which specifically tested the influence of auditive seductive details on cognitive and task load, we found no effects of adhering to or violating the coherence principle on those variables. The results were inconclusive, so that no definite conclusions can be drawn. This shows that the coherence principle may need to be reconsidered at least partly for AR-based learning settings. More studies are necessary, including a consideration of not only cognitive but also motivational effects of seductive details in AR.

Looking at interactivity as a factor, in Section 3 an important aspect that was mentioned concerning cognitive load focuses on the distinction between mental and physical interaction. While physical interaction can in general increase learning in AR, it is important to consider how these physical actions can elicit and support mental processes. Looking at the results of Study 3 (Paper 2, Krüger & Bodemer, 2020), although the learning outcomes suggest an underload when low mental and low physical interaction were combined, and an overload when high mental and high physical interaction were combined, this pattern was not fully found for the self-reported cognitive load and task load. Descriptively, the pattern was partly supported, but more research is necessary on how AR-specific learner control and guidance can be combined to support learning as effectively as possible.

Concerning the factor of spatiality, Section 3 introduces the potential of displaying 3D objects in three dimensions, relieving learners of the task to mentally transform 2D graphics into 3D mental models. It is suggested that unnecessary cognitive load from this transformation can be reduced and that cognitive load from deeper processes of 3D mental model construction can be increased. The results in Study 4 (Paper 3, Krüger et al., 2022) show that a 3D representation can indeed increase GCL, but no decrease in ECL was found. Together with the outcome that spatial relations knowledge was increased through the 3D presentation, one interpretation of these results suggests that the 3D presentation focused learners' attention on the spatial structure, while no transformation processes and thus no ECL necessarily took place when showing the 2D presentation. Another aspect of spatiality in AR, the spatially integrated presentation of virtual and physical elements, was examined in Study 6 (Paper 5, Krüger & Bodemer, 2022a). Descriptively, the expected positive effects of integration on cognitive load and task load were supported, and temporal demand was significantly lower for the integrated than the separated presentation, while perceived performance was significantly higher. This indicates a potential to decrease unnecessary cognitive load and increase relevant cognitive load when integrating virtual and physical representations, although more research is necessary to confirm this.

Additionally, we found correlations between cognitive load and task load subconstructs in Study 3 (Paper 2, Krüger & Bodemer, 2020). These showed positive relations of GCL and mental demand, GCL and effort, and ECL and frustration. They can be used to inform the connection between these two conceptualisations of workload.

5.1.4 Immersion

Concerning the construct of immersion as described in Section 2.4, from a theoretical point of view both contextuality and interactivity seem to be relevant in AR, as described in Section 3. Contextuality supports the potential of context immersion as described by Kim (2013). Learners can thus be immersed within a combined virtual-physical experience, including the possibility of contextual coupling. When looking at the results from Study 5 (Paper 4, Krüger & Bodemer, subm.), participants described that they felt more surrounded by the learning material when they were close to the corresponding physical objects. Concerning interactivity, the context immersion conceptualisation by Kim (2013) further included that embodied interaction has a positive impact on context immersion, thus describing the impact of interactivity on immersion. None of the studies included in the current dissertation empirically examined the effects of interactivity on immersion.

In the current dissertation, only one study examines immersion as an outcome variable. However, in another study that is currently in press we found that in an augmented 360° photo environment, all subconstructs of immersion (i.e. interest, usability, emotional attachment, focus of attention, presence, and flow) were increased when interactive learner control was implemented in comparison to when it was not (Krüger et al., in press), thus showing the impact of interactivity on immersion. Furthermore, interest and presence were increased in this study when the context of the learning material was visible in comparison to invisible, thus showing the impact of contextuality on some aspects of immersion. While the learning material in this study included no real AR environment, but 360° photos that were enriched with virtual information, the results might still be partly transferable to AR-based environments. In Section *4.1. Methodological Approach* in Paper 5 (Krüger & Bodemer, 2022a) we argue that the methodological approach of simulating an AR-based learning environment can help gather insights in a controlled experimental setting that can then be transferred to a more complex, real AR environment to be examined there.

5.1.5 Motivation

Concerning the construct of motivation as described in Section 2.5, both contextuality and interactivity seem to be relevant in AR, as described in Section 3. In Study 1 (Paper 1, Krüger et al., 2019), although expected, no significant difference was found in motivation for AR in comparison to non-AR, although descriptively a small advantage for AR was found. In this study, however, conclusions cannot be drawn for one specific factor, as described above.

Looking at contextuality as a characteristic, motivation might be influenced by aspects like feelings of relevance when learning in a matching physical environment, and sense of presence within an environment may have an influence on motivation. In Study 5 (Paper 4, Krüger & Bodemer, subm.) the participants described that they felt more motivated when receiving virtual information directly at the corresponding physical objects. This shows the potential of contextuality for learners' motivation, although more systematic research is necessary. Concerning interactivity, learner control has been described as a feature that can lead to increased motivation, although in the current dissertation no empirical examination of this took place so that more research on this is necessary.

In the current dissertation, motivation was only examined as an outcome variable in two studies. A study on augmented 360° photos that is currently in press and was already mentioned concerning the variable of immersion also examined learners' motivation based on the subconstructs of the ARCS model by Keller (2010). Visibility of context had a positive effect on attention, relevance, and satisfaction, while interactive learner control had a positive effect on attention, confidence, and satisfaction. We also found an interaction effect of the two factors, showing a smaller satisfaction when neither context was visible, nor learner control was possible in comparison to when either or both were applied. It thus seems that when implementing these aspects of contextuality and interactivity, motivation in the form of satisfaction profits from each, although it does not additionally profit from their combination. Again, while the learning material in this study included augmented 360° photos and no real AR environment, the results might still be partly transferable to AR-based environments, as also argued in Section *4.1. Methodological Approach* in Paper 5 (Krüger & Bodemer, 2022a).

5.1.6 Spatial Abilities

Concerning the construct of spatial abilities as described in Section 2.6, a connection has mainly been made to the factor of spatiality. In Section 3, the potential of learners' spatial abilities having an influence on learning processes especially when it comes to learning with spatial representations and about spatial constructs was described. In Study 4 (Paper 3, Krüger et al., 2022), we took an ability-as-compensator hypothesis perspective, expecting that especially learners with lower spatial abilities would profit from 3D visualisations. However, the results show support for an ability-as-enhancer hypothesis instead, showing that learners with high spatial abilities profited from 3D, while learners with low spatial abilities did not. For learners with high spatial abilities, the influence of the form of presentation on their spatial relations knowledge outcome was mediated by GCL, which was not the case for low and average spatial abilities learners. High spatial abilities may thus have supported germane cognitive processing of 3D visualisations, which in turn led to increased knowledge. In another perspective we looked at the possibilities of improving spatial abilities through AR in Study 1 (Paper 1, Krüger et al., 2019). No effect was found on spatial abilities when learning with AR in comparison to non-AR. However, we embedded no specific spatial abilities training and AR was only used for a short time, which may explain the results. In general, considering learners' spatial abilities when looking at the spatiality of AR-based learning settings thus seems important to gain a more complete picture. This was also supported by another study that we executed, showing a moderating influence of different types of spatial abilities on different types of learning task scores and learning outcomes (Krüger & Bodemer, 2021).

5.2 Theoretical Implications

Theoretical implications can be highlighted from the proposal of the ARcis framework, and the empirical results of the studies summarised in the previous section and described in more detail in the five included papers. In general, the current dissertation highlights the necessity to apply not only a media comparison approach when examining learning with AR. As stated in Section 1.2, alternative types of studies are intra-medium studies and aptitude-treatment-interaction studies. In the current dissertation, the application of intra-medium designs worked well for Study 3 (Paper 2, Krüger & Bodemer, 2020) on different types of interaction within AR, Study 4 (Paper 3, Krüger et al., 2022) on the dimensionality of representation within AR, and Study 5 (Paper 4, Krüger & Bodemer, subm.) on the positioning of corresponding information within AR. Study 4 in addition included an aptitude-treatment-interaction design for the learner characteristic spatial abilities. The studies offer unique insights that can be applied to extend the theory on learning with AR.

When it comes to MERs, the current dissertation mainly provides a focus on the three ARcis characteristics as definitions of functions of MERs in AR. Contextuality describes the combination of virtual and physical representations, interactivity describes the implementation of different interactive representations, and spatiality describes the possibility to implement 3D representations in physical space. The specific combination of these kinds of representations and its impact on learning processes and outcomes concerning different learning objectives still needs to be examined in further experimental studies. When looking at learning achievement in AR-based learning experiences, the current dissertation highlights that it is important to not only focus on general learning outcomes but to consider different types of learning objectives. These can be matched by the design of AR experiences based on the three ARcis characteristics. Contextuality may especially support a mental connection between physical and virtual elements, while for interactivity a level of guidance might be necessary to consider, and spatiality supports learning about spatial constructs. Concerning cognitive load and task load it was shown that it is important to consider different features of AR, which can all have an influence on cognitive processes and thus cognitive load. Furthermore, necessary and unnecessary cognitive load should be distinguished for a more complete picture. For both immersion and motivation, theoretical assumptions were made based on the ARcis characteristics. However, these assumptions were only partly examined empirically, so that more research is necessary. However, it can be assumed that subfactors of contextuality have an impact on both constructs. Concerning spatial abilities, the current dissertation shows support for an ability-as-enhancer hypothesis as a potential factor when providing 3D representations for learning about spatial objects.

5.3 Practical Implications

Based on the proposed ARcis framework and the empirical results in the different studies, various practical implications of the results of the current dissertation can be concluded tentatively. Concerning the ARcis framework described in Section 3, the main implication is that the three characteristics can be used to inform the systematic design of AR-based learning applications. The analyses of the three applications concerning their implementation of the characteristics show a range of possible implementations, which should be considered in light of the learning objective or goal of the application. When the learning objective describes the mental integration of information provided through virtual elements, physical objects, and the physical environment, contextuality should be considered. More specifically, the main insights on contextuality from Study 5 (Paper 4, Krüger & Bodemer, subm.) indicate that positioning virtual information close to corresponding physical objects in AR can guide learners' attention towards the physical objects and their relation to the virtually provided information. Designers of AR-based learning environments that have the goal to evoke the creation of combined mental models might thus consider placing virtual elements close to corresponding physical objects as to guide learners' attention towards the physical components and their connection to the virtual elements. When the learning objective describes a discovery process that should be actively controlled by the learners themselves, interactivity should be considered. With regard to interactivity, the main insights from Study 3 (Paper 2, Krüger & Bodemer, 2020) indicate that a certain degree of interaction can be advantageous to learning in AR, but that not everything should be left to the learners, and that guidance can be necessary, especially during initial use. Designers of AR-based learning environments that want to apply an interactive learning process might thus consider the amount of learner control and the amount of guidance in the experience as to not overload learners. When the learning objective focuses on spatial learning, such as knowledge and understanding of the spatial positions and relations of components of 3D objects, spatiality should be considered. The results in Study 4 (Paper 3, Krüger et al., 2022) show that 3D presentation of virtual objects in AR can support learning of spatial structures, but that it is possible that learners' spatial abilities are not always sufficient to process 3D representations. Designers of AR-based learning environments that want to support spatial learning might thus consider an integration of virtual 3D models as to support the building of spatial mental models. Furthermore, they might consider learners’ spatial abilities that they already bring to the learning situation and think about the possibility of training necessary abilities before the learning experience.

When considering the different potential implementations of AR-based learning experiences described in Sections 3.1.3, 3.2.3, and 3.3.3, a range of levels of implementing the three characteristics is illustrated. In the design of contextuality, the general environment of the AR-based learning setting can have no thematical relevance, full thematical relevance, or something in between these extremes. The same is true for the specific physical anchor. Furthermore, the general environment and the physical anchor can be designed as being fully visible, or they can be partly or fully covered by virtual elements. Looking at interactivity, purely virtual interaction, purely physical interaction, and mediated physical-virtual interaction need to be considered in the design of AR applications. All these types of interaction can be more or less elaborate, can require more or less physically big movement, and can be more or less meaningful for the learning objective. When considering spatiality in the design of AR applications, both the anchoring within the physical world and the spatial properties of the virtual objects can be taken into account. For the anchoring, world level and element level linking need to be considered. For the spatial properties of the virtual elements, dimensionality, size, and spatial relations within and between the virtual elements can be applied in the design. When looking at the design of AR, not only the individual ARcis characteristics but also their interplay should be considered, as described in Section 3.4. Spatial anchoring on a bigger level can be implemented to support contextuality, spatial anchoring on a smaller level can be implemented to support mediated interactivity, and contextuality and mediated interactivity can be implemented to support spatial perception. These and other synergies could be considered in the design of AR-based learning experiences.

Concerning the application of multimedia principles in AR described in Sections 3.1.3, 3.2.3, and 3.3.3, contextuality was connected to the coherence principle, interactivity to the learner control and guided activity principle, and spatiality to the spatial contiguity principle. In Study 7 (Paper 5, Krüger & Bodemer, 2022a), the application of the coherence principle was not supported, showing no advantage of not including matching or non-matching sounds in a simulated AR application. Still, the coherence principle should in general be considered when designing learning applications in AR, although the addition of small sounds may not be so harmful as to lead to cognitive overload. More research with different added seductive details should be executed, also taking into account possible motivational effects. In Study 6 (Paper 5, Krüger & Bodemer, 2022a), the application of the spatial contiguity principle was not fully supported. However, the descriptive results suggest that the principle can also be applied to simulated AR applications. The principle should be considered in the design of AR-based learning experiences, although more research is necessary concerning the specific impact when spatially combining virtual and physical elements. The learner control principle and guided activity principle were not specifically tested for AR in the current dissertation, although the results in Study 3 (Paper 2, Krüger & Bodemer, 2020) suggest that some form of guidance is necessary when learner control is implemented. In total it can be said that general design principles such as multimedia design principles, but also AR-specific design decisions based on the ARcis characteristics with the learning objective in mind, should be considered when designing AR-based learning experiences.

5.4 Limitations

There are some limitations that need to be taken into account when interpreting the results of the current dissertation including the seven studies. In general, it needs to be considered that the three ARcis characteristics cannot always be clearly separated from each other in the design of studies. While this describes the interplay mentioned in Section 3.4, it also poses difficulties for systematically manipulating factors of AR in intra-medium studies. However, the ARcis conceptualisation does not strive for a clear-cut distinction between the characteristics. It also does not strive to cover all potential characteristics of AR but gives an indication for relevant characteristics. It can be the basis for a first step in further systemising research on and design of learning in AR that can be extended and improved in future conceptualisations and research. People applying the framework should keep this in mind.

The research on contextuality is still in an early stage due to the mock-up AR design in Study 2 (Paper 1, Krüger et al., 2019) and the exploratory nature of and small sample size in Study 5 (Paper 4, Krüger & Bodemer, subm.). Results concerning this characteristic should be interpreted with caution. More research needs to confirm the insights considering the positioning of virtual elements in AR. Furthermore, as already mentioned above, the current dissertation considers the constructs immersion and motivation for both contextuality and interactivity. The empirical studies, however, only examine immersion and motivation in relation to contextuality. The relation of interactivity and the two processes thus stays quite theoretical in the current dissertation, and more research is necessary.

While in all of the studies presented in the current dissertation cognitive load was conceptualised and operationalised in a nuanced way, the measurements were mostly retrospective and subjective. Only in Study 2 (Paper 1, Krüger et al., 2019), cognitive load was also measured through secondary task execution and continuously during the learning task, and there were some limitations connected to this measurement due to other systematic differences between the two conditions (see Section *4.3 Discussion* in Paper 1, Krüger et al., 2019). The potential to enrich the measurement of cognitive load by adding objective, continuous measures for example through physical measurements like eye tracking is further discussed in Section *4.5. Limitations and future studies* in Paper 3 (Krüger et al., 2022).

When looking at the learner characteristics that were considered theoretically and empirically in the current dissertation, only mental rotation abilities as a specific form of spatial abilities is examined. There are, however, more types of spatial abilities that may play a role here, as shortly introduced in Section 2.6. In another study, we examined both 3D spatial visualisation abilities and 2D spatial memory abilities and their relation to learning with AR and non-AR (Krüger & Bodemer, 2021). The results showed diverging moderating effects of the two types of abilities, suggesting that different spatial abilities may have differing impact on learning processes and outcomes. This needs to be examined further in future research. Furthermore, spatial abilities are not the only learner characteristics that can play a role when learning with AR. In future research, characteristics like prior knowledge, interest in the topic, and learning motivation should be considered, as also suggested by Cheng and Tsai (2013).

When looking at Study 2 (Paper 1, Krüger et al., 2019), Study 6 (Paper 5, Krüger & Bodemer, 2022a) and Study 7 (Paper 5, Krüger & Bodemer, 2022a), all three studies did not apply real AR applications but mock-ups or simulations of these. While this offers the opportunity for more controlled lab-based research, as also described in Section *4.1. Methodological Approach* in Paper 5 (Krüger & Bodemer, 2022a), this shows the necessity to transfer the results from these studies onto real AR-based experiences. In general, all studies took place in a more or less pre-structured laboratory setting. The application of the results in the field still needs to be confirmed in future research.

In general, the complexity of the learning tasks differed a lot between the different studies. While the interactive discovery learning task applied similarly in Study 1 (Paper 1, Krüger et al., 2019) and Study 3 (Paper 2, Krüger & Bodemer, 2020) was quite complex due to many interacting components, all other studies mainly included the task to remember the content displayed during the learning phase. Here, Study 4 (Paper 3, Krüger et al., 2022) and Study 6 (Paper 5, Krüger & Bodemer, 2022a) included less information than Study 5 (Paper 4, Krüger & Bodemer, subm.). This shows that no truly systematic approach was taken when designing the learning tasks in the different studies. This is also apparent in the duration of the learning phases, which were sometimes limited to a specific time, but at other times unrestricted. In Study 5 (Paper 4, Krüger & Bodemer, subm.), for example, the duration of the learning phase was restricted to 40 minutes. In Study 4 (Paper 3, Krüger et al., 2022), no restriction was given for the duration of the learning phase, but learners on average took only a short time. In Study 6 (Paper 5, Krüger & Bodemer, 2022a), the videos presented were only around 3 minutes long. All learning experiences were one-time implementations usually with relatively short durations. When it comes to an implementation of these in formal or informal learning settings, the learning experiences will probably be longer, which may impact attentional processes and cognitive load.

5.5 Future Research and Outlook

The systematic manipulation and experimental analysis of individual, small-scale, AR-specific factors and attributes enables the separate, unconfounded investigation of individual mechanisms, which is crucial for gaining knowledge in the study of AR-based learning experiences. More research on all ARcis characteristics is necessary. Concerning contextuality, the literature shows that AR applications used in a specific, relevant location can lead to learning outcomes and positive learner experiences (e.g., Georgiou & Kyza, 2021; Kamarainen et al., 2013), so that pursuing this research strand may lead to further insights into the precise implementation of location-based AR applications in different settings and with different learning objectives. Looking at interactivity, the literature in general shows positive results of interaction in AR on learning (Johnson-Glenberg & Megowan-Romanowicz, 2017; Lindgren et al., 2016), but more research is needed on the exact factors and types of interaction that play a role in interactive AR applications. Regarding spatiality, studies have also shown the influence of spatial abilities in learning with 3D visualisations (Krüger & Bodemer, 2021; Stull & Hegarty, 2016), but this is not always clear-cut and should be further examined in order to draw conclusions for the effective use of AR applications.

In addition to laboratory studies as mainly applied in the current dissertation and as already described in the limitations, in the further course, the results from this research should be tested in more complex, authentic learning settings to ensure transferability. Future studies should also look at the combination and interaction of different features to provide a basis for the meaningful composition of learning-relevant design variables and criteria for complex AR applications, as AR applications are usually based on all three features. As described in Section 5.1.4 and Section 5.1.5, we have executed a study that examines the interaction of contextuality and interactivity subfactors concerning immersion and motivation in an augmented 360° photo environment (Krüger et al., in press). While this is not a real AR environment, the results may be used as a basis for further research in this area.

Furthermore, the research into the influence of different learning goals, outcomes, and tasks needs to be broadened. In the current dissertation, the theoretical assumptions concerning the ARcis characteristics led the design of the learning material and AR applications used in the studies. More insights into these relations are relevant for practical applications, especially to be able to leverage the characteristics based on their specific mechanisms to support learning. As also already described in the description of the limitations above, focus on different learner characteristics such as motivation or certain abilities is also relevant for future research, as these can have an influence on the processing of content and the learning experience in AR.

In addition to individual learning settings, collaborative learning settings, which are becoming more and more common in education, and their implementation based on the ARcis characteristics are also interesting, as there is a great potential here as well (e.g., learning partners as part of the AR context; collaborative interaction with AR elements; spatial positioning of learning partners around an AR object; see also Paper 1, Krüger et al., 2019). Further down the line, other technologies should also be considered that can bring a better implementation of AR features (e.g., AR glasses to have hands free for interaction). All in all, AR offers many new learning opportunities that need further empirical exploration to provide a basis for the implementation and use of effective and efficient AR applications.

In addition to the theoretical conceptualisation of the ARcis framework to classify and examine specific features of educational AR, first attempts at developing a questionnaire to learn more about learners' experience of these three characteristics have been made. This questionnaire has already been applied as a manipulation check in Study 4 (Paper 3, Krüger et al., 2022) in the current dissertation. In a first evaluation, we tested a first version of this questionnaire in four studies and found an acceptable first fit (Krüger & Bodemer, 2022b). Further studies with a broadened set of items for evaluation of the specific wordings are currently ongoing so that this questionnaire can be used in future studies on learners' experience of the AR environment.

While the research perspective on AR is clearly in need of more systematic experimental studies and results, practitioners like teachers have also shown that they need more support to implement AR in the classroom as indicated by 16 teachers in a survey study (Buchner, Krüger, et al., 2022). Besides the general technological infrastructure that is necessary to implement AR in education, teachers describe that they need general support to create content in a pedagogically informed way, or that they need content that already matches their learning objectives. For this, it is important that research and practice work hand in hand in the research, design, and implementation of AR-based learning environments. The ARcis framework can provide a basis for communication of different stakeholders in this aspect.

6 Conclusion

In the current dissertation the overarching goal of gaining insight into specific characteristics of AR-based learning and how they can be leveraged to support learning processes and outcomes was reached by working on achieving theoretical, empirical, and practical subgoals. A theoretical definition of three characteristics of AR-based learning was achieved, describing contextuality, interactivity, and spatiality as three unique characteristics. These three characteristics can be leveraged to support learning in AR, working through different mechanisms and relating to different learning objectives. Empirical research and practical implementations of AR-based learning can consider this framework for a systematic approach. The studies presented in the current dissertation show that it can be of interest to conduct experimental studies with systematically manipulated variables based on the three characteristics. Concerning contextuality, it can be concluded that the positioning of virtual information in relation to corresponding physical objects should be considered in learning with AR. With regard to interactivity, it can be concluded that the amount and combination of types of interaction and the necessity of guidance during initial use should be considered in learning with AR. Concerning spatiality, it can be concluded that the dimensionality of a virtual representation in AR and its interaction with spatial abilities should be considered in learning with AR. Overall, the results of the empirical studies suggest that variables such as learning behaviour, cognitive load, immersion, motivation, and different types of knowledge can be influenced by the different design of AR-based learning applications, and person variables such as spatial abilities should also be considered in this context. These results can further be used to make certain design decisions for AR-based learning while considering the goal and the pursued learning objective. In conclusion, the theoretical framework and research presented in the current dissertation has provided insights into a meaningful use of AR for learning and forms a solid basis for further research and development in the field of AR-supported learning.

7 References

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, *16*(3), 183–198. https://doi.org/10.1016/j.learninstruc.2006.03.001

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. https://doi.org/10.1016/j.edurev.2016.11.002

Alexander, B., Ashford-Rowe, K., Barajas-Murphy, N., Dobbin, G., Knott, J., McCormack, M., Pomerantz, J., Seilhamer, R., & Weber, N. (2019). *EDUCAUSE Horizon Report: 2019 Higher Education Edition*. EDUCAUSE.

Altmeyer, K., Kapp, S., Thees, M., Malone, S., Kuhn, J., & Brünken, R. (2020). The use of augmented reality to foster conceptual knowledge acquisition in STEM laboratory courses—Theoretical background and empirical results. *British Journal of Educational Technology*, *51*(3), 611–628. https://doi.org/10.1111/bjet.12900

Álvarez-Marín, A., & Velázquez-Iturbide, J. Á. (2021). Augmented reality and engineering education: A systematic review. *IEEE Transactions on Learning Technologies*, *14*(6), 817–831. https://doi.org/10.1109/TLT.2022.3144356

Amores-Valencia, A., Burgos, D., & Branch-Bedoya, J. W. (2022). Influence of motivation and academic performance in the use of augmented reality in education. A systematic review. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.1011409

Arici, F., Yildirim, P., Caliklar, Ş., & Yilmaz, R. M. (2019). Research trends in the use of augmented reality in science education: Content and bibliometric mapping analysis. *Computers & Education*, *142*, Article 103647. https://doi.org/10.1016/j.compedu.2019.103647

Ayres, P., & Sweller, J. (2014). The split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 206–226). Cambridge University Press.

Azuma, R. (1997). A survey of augmented reality. *Presence: Teleoperators and Virtual Environments*, *6*(4), 355–385. https://doi.org/10.1162/pres.1997.6.4.355

Bacca, J., Baldiris, S., Fabregat, R., Graf, S., & Kinshuk. (2014). Augmented reality trends in education: A systematic review of research and applications. *Educational Technology & Society*, *17*(4), 133–149.

Baddeley, A. D. (1999). *Essentials of human memory*. Psychology Press.

Barrett, T. J., Stull, A. T., Hsu, T. M., & Hegarty, M. (2015). Constrained interactivity for relating multiple representations in science: When virtual is better than real. *Computers & Education*, *81*, 69–81. https://doi.org/10.1016/j.compedu.2014.09.009

Becker, S. A., Brown, M., Dahlstrom, E., Davis, A., DePaul, K., Diaz, V., & Pomerantz, J. (2018). *NMC Horizon Report: 2018 Higher Education Edition*. EDUCAUSE.

Billinghurst, M., & Dünser, A. (2012). Augmented reality in the classroom. *Computer*, *45*(7), 56–63. https://doi.org/10.1109/MC.2012.111

Bodemer, D., & Faust, U. (2006). External and mental referencing of multiple representations. *Computers in Human Behavior*, *22*(1), 27–42. https://doi.org/10.1016/j.chb.2005.01.005

Bodemer, D., Janssen, J., & Schnaubert, L. (2018). Group awareness tools for computer-supported collaborative learning. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), *International Handbook of the Learning Sciences* (pp. 351–358). Routledge/Taylor & Francis.

Bodemer, D., Ploetzner, R., Bruchmüller, K., & Häcker, S. (2005). Supporting learning with interactive multimedia through active integration of representations. *Instructional Science*, *33*(1), 73–95. https://doi.org/10.1007/s11251-004-7685-z

Bodemer, D., Ploetzner, R., Feuerlein, I., & Spada, H. (2004). The active integration of information during learning with dynamic and interactive visualisations. *Learning and Instruction*, *14*(3), 325–341. https://doi.org/10.1016/j.learninstruc.2004.06.006

Bogomolova, K., Vorstenbosch, M. A. T. M., El Messaoudi, I., Holla, M., Hovius, S. E. R., van der Hage, J. A., & Hierck, B. P. (2023). Effect of binocular disparity on learning anatomy with stereoscopic augmented reality visualization: A double center randomized controlled trial. *Anatomical Sciences Education*, *16*(1), 87–98. https://doi.org/10.1002/ase.2164

Bölek, K. A., De Jong, G., & Henssen, D. (2021). The effectiveness of the use of augmented reality in anatomy education: A systematic review and meta-analysis. *Scientific Reports, 11*, Article 15292. https://doi.org/10.1038/s41598-021-94721-4

Bower, M., Howe, C., McCredie, N., Robinson, A., & Grover, D. (2014). Augmented reality in education—Cases, places and potentials. *Educational Media International*, *51*(1), 1–15. https://doi.org/10.1080/09523987.2014.889400

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, *18*(1), 32–42. https://doi.org/10.3102/0013189X018001032

Buchner, J., Buntins, K., & Kerres, M. (2021). A systematic map of research characteristics in studies on augmented reality and cognitive load. *Computers and Education Open*, *2*, Article 100036. https://doi.org/10.1016/j.caeo.2021.100036

Buchner, J., Buntins, K., & Kerres, M. (2022). The impact of augmented reality on cognitive load and performance: A systematic review. *Journal of Computer Assisted Learning*, *38*(1), 285–303. https://doi.org/10.1111/jcal.12617

Buchner, J., & Kerres, M. (2023). Media comparison studies dominate comparative research on augmented reality in education. *Computers & Education*, *195*, Article 104711. https://doi.org/10.1016/j.compedu.2022.104711

Buchner, J., Krüger, J. M., Bodemer, D., & Kerres, M. (2022). Teachers’ use of augmented reality in the classroom: Reasons, practices, and needs. In C. Clark, T. Edna, C. Carol, & K. Yael (Eds.), *Proceedings of the 16th International Conference of the Learning Sciences—ICLS 2022* (pp. 1133–1136). International Society of the Learning Sciences.

Cai, Y., Pan, Z., & Liu, M. (2022). Augmented reality technology in language learning: A meta-analysis. *Journal of Computer Assisted Learning*, *38*(4), 929–945. https://doi.org/10.1111/jcal.12661

Carroll, J. B. (1993). Abilities in the domain of visual perception. In J. B. Carroll (Ed.), *Human cognitive abilities: A survey of factor-analytic studies* (pp. 304–363). Cambridge University Press. https://doi.org/10.1017/CBO9780511571312.009

Challenor, J., & Ma, M. (2019). A review of augmented reality applications for history education and heritage visualisation. *Multimodal Technologies and Interaction*, *3*(2), Article 39. https://doi.org/10.3390/mti3020039

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, *8*(4), 293–332. https://doi.org/10.1207/s1532690xci0804_2

Chang, H.-Y., Binali, T., Liang, J.-C., Chiou, G.-L., Cheng, K.-H., Lee, S. W.-Y., & Tsai, C.-C. (2022). Ten years of augmented reality in education: A meta-analysis of (quasi-) experimental studies to investigate the impact. *Computers & Education*, *191*, Article 104641. https://doi.org/10.1016/j.compedu.2022.104641

Chen, P., Liu, X., Cheng, W., & Huang, R. (2017). A review of using augmented reality in education from 2011 to 2016. In E. Popescu, Kinshuk, M. K. Khribi, R. Huang, M. Jemni, N.-S. Chen, & D. G. Sampson (Eds.), *Innovations in Smart Learning, Lecture Notes in Educational Technology* (pp. 13–18). Springer Singapore.

Chen, S.-C., Hsiao, M.-S., & She, H.-C. (2015). The effects of static versus dynamic 3D representations on 10th grade students' atomic orbital mental model construction: Evidence from eye movement behaviors. *Computers in Human Behavior*, *53*, 169–180. https://doi.org/10.1016/j.chb.2015.07.003

Chen, Y.-H., & Wang, C.-H. (2018). Learner presence, perception, and learning achievements in augmented–reality–mediated learning environments. *Interactive Learning Environments*, *26*(5), 695–708. https://doi.org/10.1080/10494820.2017.1399148

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. https://doi.org/10.1007/s10956-012-9405-9

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Chia-Chen, C., Hong-Ren, C., & Ting-Yu, W. (2022). Creative situated augmented reality learning for astronomy curricula. *Educational Technology & Society*, *25*(2), 148–162.

Chiang, F.-K., Shang, X., & Qiao, L. (2022). Augmented reality in vocational training: A systematic review of research and applications. *Computers in Human Behavior*, *129*, 107125. https://doi.org/10.1016/j.chb.2021.107125

Chytas, D., Johnson, E. O., Piagkou, M., Mazarakis, A., Babis, G. C., Chronopoulos, E., Nikolaou, V. S., Lazaridis, N., & Natsis, K. (2020). The role of augmented reality in anatomical education: An overview. *Annals of Anatomy - Anatomischer Anzeiger*, *229*, Article 151463. https://doi.org/10.1016/j.aanat.2020.151463

Clark, R. C., & Mayer, R. E. (2016). Engagement in e-learning. In R. C. Clark & R. E. Mayer (Eds.), *E-Learning and the Science of Instruction: Proven guidelines for consumers and designers of multimedia learning* (4th ed., pp. 219–238). John Wiley & Sons, Inc. https://doi.org/10.1002/9781119239086.ch11

Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research*, *53*(4), 445–459. https://doi.org/10.3102/00346543053004445

Clark, R. E., Howard, K., & Early, S. (2006). Motivational challenges experienced in highly complex learning environments. In J. Elen & R. E. Clark (Eds.), *Handling complexity in learning environments: Research and theory* (pp. 27–43). Elsevier.

Craig, A. B. (2013a). Chapter 2—Augmented reality concepts. In A. B. Craig (Ed.), *Understanding augmented reality* (pp. 39–67). Morgan Kaufmann. https://doi.org/10.1016/B978-0-240-82408-6.00002-3

Craig, A. B. (2013b). Chapter 3—Augmented Reality Hardware. In A. B. Craig (Ed.), *Understanding Augmented Reality* (pp. 69–124). Morgan Kaufmann. https://doi.org/10.1016/B978-0-240-82408-6.00003-5

de Jong, T., & van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research*, *68*(2), 179–201. https://doi.org/10.3102/00346543068002179

De Paolis, L. T., Gatto, C., Corchia, L., & De Luca, V. (2022). Usability, user experience and mental workload in a mobile Augmented Reality application for digital storytelling in cultural heritage. *Virtual Reality*. https://doi.org/10.1007/s10055-022-00712-9

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. https://link.springer.com/book/10.1007/978-1-4899-2271-7

Dengel, A. (2022). What is immersive learning? In A. Dengel, M.-L. Bourguet, D. Pedrosa, J. Hutson, K. Erenli, D. Economou, A. Peña-Rios, & Richter, Jonathon (Eds.), *2022 8th International Conference of the Immersive Learning Research Network (iLRN)* (pp. 1–5). https://doi.org/10.23919/iLRN55037.2022.9815941

Dengel, A., & Mägdefrau, J. (2018). Immersive learning explored: Subjective and objective factors influencing learning outcomes in immersive educational virtual environments. In M. J. W. Lee, S. Nikolic, M. Ros, J. Shen, L. C. U. Lei, G. K. W. Wong, & N. Venkatarayalu (Eds.), *2018*

IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE) (pp. 608–615). https://doi.org/10.1109/TALE.2018.8615281

Diao, P.-H., & Shih, N.-J. (2019). Trends and research issues of augmented reality studies in architectural and civil engineering education—A review of academic journal publications. *Applied Sciences*, *9*(9), Article 9. https://doi.org/10.3390/app9091840

Domagk, S., Schwartz, R. N., & Plass, J. L. (2010). Interactivity in multimedia learning: An integrated model. *Computers in Human Behavior*, *26*(5), 1024–1033. https://doi.org/10.1016/j.chb.2010.03.003

Dunleavy, M., & Dede, C. (2014). Augmented reality teaching and learning. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (4th ed., pp. 735–745). Springer New York.

Dutta, R., Mantri, A., & Singh, G. (2022). Evaluating system usability of mobile augmented reality application for teaching Karnaugh-Maps. *Smart Learning Environments*, *9*(1), Article 6. https://doi.org/10.1186/s40561-022-00189-8

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, *53*(1), 109–132. https://doi.org/10.1146/annurev.psych.53.100901.135153

Elford, D., Lancaster, S. J., & Jones, G. A. (2022). Exploring the effect of augmented reality on cognitive load, attitude, spatial ability, and stereochemical perception. *Journal of Science Education and Technology*, *31*(3), 322–339. https://doi.org/10.1007/s10956-022-09957-0

Feldon, D. F., Callan, G., Juth, S., & Jeong, S. (2019). Cognitive load as motivational cost. *Educational Psychology Review*, *31*(2), 319–337. https://doi.org/10.1007/s10648-019-09464-6

Ferdous, H. S., Hoang, T., Joukhadar, Z., Reinoso, M. N., Vetere, F., Kelly, D., & Remedios, L. (2019). “What’s happening at that hip?”: Evaluating an on-body projection based augmented reality system for physiotherapy classroom. In S. Brewster, G. Fitzpatrick, A. Cox, & V. Kostakos (Eds.), *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–12). Association for Computing Machinery. https://doi.org/10.1145/3290605.3300464

Fidan, M., & Tuncel, M. (2018). Augmented reality in education researches (2012–2017): A content analysis. *Cypriot Journal of Educational Sciences*, *13*(4), Article 4. https://doi.org/10.18844/cjes.v13i4.3487

Garzón, J. (2021). An overview of twenty-five years of augmented reality in education. *Multimodal Technologies and Interaction*, *5*(7), Article 37. https://doi.org/10.3390/mti5070037

Garzón, J., & Acevedo, J. (2019). Meta-analysis of the impact of augmented reality on students’ learning gains. *Educational Research Review*, *27*, 244–260. https://doi.org/10.1016/j.edurev.2019.04.001

Garzón, J., Kinshuk, Baldiris, S., Gutiérrez, J., & Pavón, J. (2020). How do pedagogical approaches affect the impact of augmented reality on education? A meta-analysis and research synthesis.

Educational Research Review, *31*, Article 100334. https://doi.org/10.1016/j.edurev.2020.100334

Garzón, J., Pavón, J., & Baldiris, S. (2019). Systematic review and meta-analysis of augmented reality in educational settings. *Virtual Reality*, *23*(4), 447–459. https://doi.org/10.1007/s10055-019-00379-9

Generosi, A., Agostinelli, T., Mengoni, M., & Ceccacci, S. (2022). Augmented Reality for assembly operation training: Does immersion affect the recall performance? *2022 IEEE International Conference on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE)*, 58–63. https://doi.org/10.1109/MetroXRAINE54828.2022.9967520

Georgiou, Y., & Kyza, E. A. (2017a). A design-based approach to augmented reality location-based activities: Investigating immersion in relation to student learning. *Proceedings of the 16th World Conference on Mobile and Contextual Learning*, 1–8. https://doi.org/10.1145/3136907.3136926

Georgiou, Y., & Kyza, E. A. (2017b). Investigating immersion in relation to students' learning during a collaborative location-based augmented reality activity. In B. K. Smith, M. Borge, E. Mercier, & K. Y. Lim (Eds.), *12th International Conference on Computer Supported Collaborative Learning (CSCL) 2017, Volume 1* (pp. 423–430). International Society of the Learning Sciences. https://doi.org/10.22318/cscl2017.57

Georgiou, Y., & Kyza, E. A. (2017c). The development and validation of the ARI questionnaire: An instrument for measuring immersion in location-based augmented reality settings. *International Journal of Human Computer Studies*, *98*, 24–37. https://doi.org/10.1016/j.ijhcs.2016.09.014

Georgiou, Y., & Kyza, E. A. (2018). Relations between student motivation, immersion and learning outcomes in location-based augmented reality settings. *Computers in Human Behavior*, *89*, 173–181. https://doi.org/10.1016/j.chb.2018.08.011

Georgiou, Y., & Kyza, E. A. (2021). Bridging narrative and locality in mobile-based augmented reality educational activities: Effects of semantic coupling on students' immersion and learning gains. *International Journal of Human-Computer Studies*, *145*, Article 102546. https://doi.org/10.1016/j.ijhcs.2020.102546

Gerup, J., Soerensen, C. B., & Dieckmann, P. (2020). Augmented reality and mixed reality for healthcare education beyond surgery: An integrative review. *International Journal of Medical Education, 11*, 1–18. https://doi.org/10.5116/ijme.5e01.eb1a

Goff, E. E., Mulvey, K. L., Irvin, M. J., & Hartstone-Rose, A. (2018). Applications of augmented reality in informal science learning sites: A review. *Journal of Science Education and Technology, 27*(5), 433–447. https://doi.org/10.1007/s10956-018-9734-4

Gogou, A., & Kasvikis, K. (2022). 'Release orpheus!': Understanding historical time in a mixed/augmented reality environment through embodied learning. *Education 3-13*. https://doi.org/10.1080/03004279.2022.2151317

Guay, F., Vallerand, R. J., & Blanchard, C. (2001). On the assessment of situational intrinsic and extrinsic motivation: The Situational Motivation Scale (SIMS). *Motivation and Emotion*, *24*, 175–213.

Guo, W., & Kim, J. H. (2020). How Augmented Reality Influences Student Workload in Engineering Education. In C. Stephanidis, D. Harris, W.-C. Li, D. D. Schmorrow, C. M. Fidopiastis, P. Zaphiris, A. Ioannou, X. Fang, R. A. Sottilare, & J. Schwarz (Eds.), *HCI International 2020 – Late Breaking Papers: Cognition, Learning and Games* (pp. 388–396). Springer International Publishing. https://doi.org/10.1007/978-3-030-60128-7_29

Hajirasouli, A., & Banihashemi, S. (2022). Augmented reality in architecture and construction education: State of the field and opportunities. *International Journal of Educational Technology in Higher Education*, *19*, Article 39. https://doi.org/10.1186/s41239-022-00343-9

Han, X., Chen, Y., Feng, Q., & Luo, H. (2022). Augmented reality in professional training: A review of the literature from 2001 to 2020. *Applied Sciences*, *12*(3), Article 3. https://doi.org/10.3390/app12031024

Hanid, M. F. A., Said, M. N. H. M., & Yahaya, N. (2020). Learning strategies using augmented reality technology in education: Meta-analysis. *Universal Journal of Educational Research*, *8*(5A), 51–56. https://doi.org/10.13189/ujer.2020.081908

Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(9), 904–908. https://doi.org/10.1177/154193120605000909

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in Psychology* (Vol. 52, pp. 139–183). Elsevier. https://doi.org/10.1016/S0166-4115(08)62386-9

Henssen, D. J. H. A., den Heuvel, L., De Jong, G., Vorstenbosch, M. A. T. M., Cappellen van Walsum, A., Van den Hurk, M. M., Kooloos, J. G. M., & Bartels, R. H. M. A. (2020). Neuroanatomy learning: Augmented reality vs. cross-sections. *Anatomical Sciences Education*, *13*(3), 353–365. https://doi.org/10.1002/ase.1912

Herpich, F., Nunes, F. B., Petri, G., & Tarouco, L. M. R. (2019). How mobile augmented reality is applied in education? A systematic literature review. *Creative Education*, *10*(7), Article 7. https://doi.org/10.4236/ce.2019.107115

Ho, S., Liu, P., Palombo, D. J., Handy, T. C., & Krebs, C. (2022). The role of spatial ability in mixed reality learning with the HoloLens. *Anatomical Sciences Education*, *15*(6), 1074–1085. https://doi.org/10.1002/ase.2146

Höffler, T. N. (2010). Spatial ability: Its influence on learning with visualizations—A meta-analytic review. *Educational Psychology Review*, *22*(3), 245–269. https://doi.org/10.1007/s10648-010-9126-7

Holmes, C. A., Newcombe, N. S., & Shipley, T. F. (2018). Move to learn: Integrating spatial information from multiple viewpoints. *Cognition*, *178*, 7–25. https://doi.org/10.1016/j.cognition.2018.05.003

Huang, H.-M., Huang, T.-C., & Cheng, C.-Y. (2022). Reality matters? Exploring a tangible user interface for augmented-reality-based fire education. *Universal Access in the Information Society*, *21*(4), 927–939. https://doi.org/10.1007/s10209-021-00808-0

Huk, T. (2006). Who benefits from learning with 3D models? The case of spatial ability. *Journal of Computer Assisted Learning*, *22*(6), 392–404. https://doi.org/10.1111/j.1365-2729.2006.00180.x

Ibáñez, M.-B., & Delgado-Kloos, C. (2018). Augmented reality for STEM learning: A systematic review. *Computers & Education*, *123*, 109–123. https://doi.org/10.1016/j.compedu.2018.05.002

Jdaitawi, M., Kan'an, A., Rabab'h, B., Alsharoa, A., Johari, M., Alashkar, W., Elkilany, A., & Abas, A. (2022). The importance of augmented reality technology in science education: A scoping review. *International Journal of Information and Education Technology*, *12*(9), 956–963. https://doi.org/10.18178/ijiet.2022.12.9.1706

Johnson-Glenberg, M. C., & Megowan-Romanowicz, C. (2017). Embodied science and mixed reality: How gesture and motion capture affect physics education. *Cognitive Research: Principles and Implications*, *2*(1), 24. https://doi.org/10.1186/s41235-017-0060-9

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 and Education*, *68*, 545–556. https://doi.org/10.1016/j.compedu.2013.02.018

Keller, J. M. (2010). *Motivational design for learning and performance*. Springer US. https://doi.org/10.1007/978-1-4419-1250-3

Kim, M. J. (2013). A framework for context immersion in mobile augmented reality. *Automation in Construction*, *33*, 79–85. https://doi.org/10.1016/j.autcon.2012.10.020

Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01997

Klepsch, M., & Seufert, T. (2021). Making an effort versus experiencing load. *Frontiers in Education*, *6*, Article 645284. https://doi.org/10.3389/feduc.2021.645284

Kozma, R. B. (1994). Will media influence learning? Reframing the debate. *Educational Technology Research and Development*, *42*(2), 7–19. https://doi.org/10.1007/BF02299087

Krüger, J. M., & Bodemer, D. (2020). Different types of interaction with augmented reality learning material. In D. Economou, A. Klippel, H. Dodds, A. Peña-Rios, M. J. W. Lee, D. Beck, J. Pirker, A. Dengel, T. M. Peres, & J. Richter (Eds.), *2020 6th International Conference of the Immersive*

Learning Research Network (iLRN) (pp. 78–85). IEEE. https://doi.org/10.23919/iLRN47897.2020.9155148

Krüger, J. M., & Bodemer, D. (2021). Space, a central frontier—The role of spatial abilities when learning the structure of 3D AR objects. In D. Economou, A. Peña-Rios, A. Dengel, H. Dodds, M. Mentzelopoulos, A. Klippel, K. Erenli, M. J. W. Lee, & J. Richter (Eds.), *2021 7th International Conference of the Immersive Learning Research Network (iLRN)* (pp. 258–265). IEEE. https://doi.org/10.23919/iLRN52045.2021.9459365

Krüger, J. M., & Bodemer, D. (2022a). Application and investigation of multimedia design principles in augmented reality learning environments. *Information*, *13*(2), Article 2. https://doi.org/10.3390/info13020074

Krüger, J. M., & Bodemer, D. (2022b). Work-in-Progress—Measuring learners' subjective experience in augmented reality: First evaluation of the ARcis questionnaire. In A. Dengel, M.-L. Bourguet, D. Pedrosa, J. Hutson, K. Erenli, A. Peña-Rios, & J. Richter (Eds.), *2022 8th International Conference of the Immersive Learning Research Network (iLRN)* (pp. 1–3). IEEE. https://doi.org/10.23919/iLRN55037.2022.9815900

Krüger, J. M., & Bodemer, D. (subm.). *Positioning augmented reality information for learning in nature: An exploratory pilot study* [Manuscript submitted for publication].

Krüger, J. M., Buchholz, A., & Bodemer, D. (2019). Augmented reality in education: Three unique characteristics from a user's perspective. In M. Chang, H.-J. So, L.-H. Wong, F.-Y. Yu, & J. L. Shih (Eds.), *Proceedings of the 27th International Conference on Computers in Education* (pp. 412–422). Asia-Pacific Society for Computers in Education.

Krüger, J. M., Koch, M., & Bodemer, D. (in press). The role of context and interaction when learning with augmented 360° photos. *2023 9th International Conference of the Immersive Learning Research Network (iLRN)*.

Krüger, J. M., Palzer, K., & Bodemer, D. (2022). Learning with augmented reality: Impact of dimensionality and spatial abilities. *Computers and Education Open*, *3*, Article 100065. https://doi.org/10.1016/j.caeo.2021.100065

Krüger, J. M., Schacht, F., & Bodemer, D. (2023). Aktives Integrieren von Repräsentationen bei interaktiven Augmented Reality-Anwendungen: Betrachtung von kognitiver Belastung und Lernerfolg. *MedienPädagogik: Zeitschrift Für Theorie Und Praxis Der Medienbildung*, *51*.

Lai, A.-F., Chen, C.-H., & Lee, G.-Y. (2019). An augmented reality-based learning approach to enhancing students' science reading performances from the perspective of the cognitive load theory. *British Journal of Educational Technology*, *50*(1), 232–247. https://doi.org/10.1111/bjet.12716

Lai, J. W., & Cheong, K. H. (2022). Educational opportunities and challenges in augmented reality: Featuring implementations in physics education. *IEEE Access*, *10*, 43143–43158. https://doi.org/10.1109/ACCESS.2022.3166478

Law, E. L.-C., & Heintz, M. (2021). Augmented reality applications for K-12 education: A systematic review from the usability and user experience perspective. *International Journal of Child-Computer Interaction, 30,* 100321. https://doi.org/10.1016/j.ijcci.2021.100321

Leppink, J., Gog, T., Paas, F., & Sweller, J. (2015). Cognitive load theory: Researching and planning teaching to maximise learning. In J. Cleland & S. J. Durning (Eds.), *Researching Medical Education* (pp. 207–218). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118838983.ch18

Leppink, J., Paas, F., Van der Vleuten, C. P. M., Van Gog, T., & van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods, 45*(4), 1058–1072. https://doi.org/10.3758/s13428-013-0334-1

Lin, H.-Y., & Tsai, S.-C. (2021). Student perceptions towards the usage of AR-supported STEMUP application in mobile courses development and its implementation into English learning. *Australasian Journal of Educational Technology*, *37*(3), 88–103. https://doi.org/10.14742/ajet.6125

Lin, W., Lo, W.-T., & Yueh, H.-P. (2022). Effects of learner control design in an AR-based exhibit on visitors' museum learning. *PLOS ONE*, *17*(10), Article e0274826. https://doi.org/10.1371/journal.pone.0274826

Lindgren, R., Tscholl, M., Wang, S., & Johnson, E. (2016). Enhancing learning and engagement through embodied interaction within a mixed reality simulation. *Computers and Education, 95*, 174–187. https://doi.org/10.1016/j.compedu.2016.01.001

Liu, Q., Yu, S., Chen, W., Wang, Q., & Xu, S. (2021). The effects of an augmented reality based magnetic experimental tool on students' knowledge improvement and cognitive load. *Journal of Computer Assisted Learning*, *37*(3), 645–656. https://doi.org/10.1111/jcal.12513

Loorbach, N., Peters, O., Karreman, J., & Steehouder, M. (2015). Validation of the Instructional Materials Motivation Survey (IMMS) in a self-directed instructional setting aimed at working with technology. *British Journal of Educational Technology*, *46*(1), 204–218. https://doi.org/10.1111/bjet.12138

López-Belmonte, J., Moreno-Guerrero, A.-J., López Núñez, J. A., & Pozo Sánchez, S. (2019). Analysis of the productive, structural, and dynamic development of augmented reality in higher education tesearch on the Web of Science. *Applied Sciences*, *9*(24), Article 24. https://doi.org/10.3390/app9245306

MacCallum, K., & Jamieson, J. (2017). Exploring augmented reality in education viewed through the affordance lens. *Proceedings of the 8th Annual Conference of Computing and Information Technology Education and Research in New Zealand*.

Makransky, G. (2021). The immersion principle in multimedia learning. In L. Fiorella & R. E. Mayer (Eds.), *The Cambridge Handbook of Multimedia Learning* (3rd ed., pp. 296–303). Cambridge University Press. https://doi.org/10.1017/9781108894333.031

Makransky, G., & Petersen, G. B. (2021). The Cognitive Affective Model of Immersive Learning (CAMIL): A theoretical research-based model of learning in immersive virtual reality. *Educational Psychology Review*, *33*(3), 937–958. https://doi.org/10.1007/s10648-020-09586-2

Mayer, R. E. (2014). Incorporating motivation into multimedia learning. *Learning and Instruction*, *29*, 171–173. https://doi.org/10.1016/j.learninstruc.2013.04.003

Mayer, R. E. (2020a). 2 Science of learning: Determining how multimedia learning works. In *Multimedia learning* (3rd ed., pp. 29–62). Cambridge University Press. https://doi.org/10.1017/9781316941355.004

Mayer, R. E. (2020b). 4 Science of assessment: Determining what was learned in multimedia learning. In *Multimedia learning* (3rd ed., pp. 95–116). Cambridge University Press. https://doi.org/10.1017/9781316941355

Mayer, R. E. (2020c). 6 Coherence principle. In *Multimedia learning* (3rd ed., pp. 143–165). Cambridge University Press. https://doi.org/10.1017/9781316941355.009

Mayer, R. E. (2020d). 9 Spatial contiguity principle. In *Multimedia learning* (3rd ed., pp. 207–226). Cambridge University Press. https://doi.org/10.1017/9781316941355.012

Mayer, R. E. (2020e). 18 Immersion principle. In *Multimedia learning* (3rd ed.). Cambridge University Press. https://doi.org/10.1017/9781316941355.023

Mayer, R. E. (2020f). *Multimedia Learning* (3rd ed.). Cambridge University Press. https://doi.org/10.1017/9781316941355

Mazzuco, A., Krassmann, A. L., Reategui, E., & Gomes, R. S. (2022). A systematic review of augmented reality in chemistry education. *Review of Education*, *10*(1), Article e3325. https://doi.org/10.1002/rev3.3325

McBain, K. A., Habib, R., Laggis, G., Quaiattini, A., M. Ventura, N., & Noel, G. P. J. C. (2022). Scoping review: The use of augmented reality in clinical anatomical education and its assessment tools. *Anatomical Sciences Education*, *15*(4), 765–796. https://doi.org/10.1002/ase.2155

McCord, K. H., Ayer, S. K., Perry, L. A., Patil, K. R., London, J. S., Khoury, V., & Wu, W. (2022). Student approaches and performance in element sequencing tasks using 2D and augmented reality formats. *Education Sciences*, *12*(4), Article 4. https://doi.org/10.3390/educsci12040247

Meister, P., Miller, J., Wang, K., Dorneich, M. C., Winer, E., Brown, L. J., & Whitehurst, G. (2022). Designing three-dimensional augmented reality weather visualizations to enhance general aviation weather education. *IEEE Transactions on Professional Communication*, *65*(2), 321–336. https://doi.org/10.1109/TPC.2022.3155920

Milgram, P., Takemura, H., Utsumi, A., & Kishino, F. (1994). Augmented Reality: A class of displays on the reality-virtuality continuum. *SPIE 2351: Telemanipulator and Telepresence Technologies*, 282–292. https://doi.org/10.1117/12.197321

Moreno, R. (2006). Instructional technology: Promise and pitfalls. In L. M. PytlikZillig, M. Bodvarsson, & R. Bruning (Eds.), *Technology-Based Education: Bringing Researchers and Practitioners Together* (pp. 1–20). IAP.

Moreno, R. (2010). Cognitive load theory: More food for thought. *Instructional Science*, *38*(2), 135–141. https://doi.org/10.1007/s11251-009-9122-9

Moreno, R., & Mayer, R. E. (2007). Interactive multimodal learning environments. *Educational Psychology Review*, *19*(3), 309–326. https://doi.org/10.1007/s10648-007-9047-2

Mystakidis, S., Christopoulos, A., & Pellas, N. (2021). A systematic mapping review of augmented reality applications to support STEM learning in higher education. *Education and Information Technologies, 27*, 1883–1927. https://doi.org/10.1007/s10639-021-10682-1

Nilsson, N. C., Nordahl, R., & Serafin, S. (2016). Immersion revisited: A review of existing definitions of immersion and their relation to different theories of presence. *Human Technology*, *12*(2), 108–134. https://doi.org/10.17011/ht/urn.201611174652

O'Brien, J. (2020). *2020 EDUCAUSE Horizon Report: Teaching and Learning Edition*. EDUCAUSE.

Olympiou, G., & Zacharia, Z. C. (2012). Blending physical and virtual manipulatives: An effort to improve students' conceptual understanding through science laboratory experimentation. *Science Education*, *96*(1), 21–47. https://doi.org/10.1002/sce.20463

Paas, F., Tuovinen, J. E., van Merriënboer, J. J. G., & Aubteen Darabi, A. (2005). A motivational perspective on the relation between mental effort and performance: Optimizing learner involvement in instruction. *Educational Technology Research and Development*, *53*(3), 25–34. https://doi.org/10.1007/BF02504795

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford University Press ; Clarendon Press.

Papakostas, C., Troussas, C., Krouska, A., & Sgouropoulou, C. (2021). Exploration of augmented reality in spatial abilities training: A systematic literature review for the last decade. *Informatics in Education*, *20*(1), 107–130. https://doi.org/10.15388/infedu.2021.06

Park, B., Flowerday, T., & Brünken, R. (2015). Cognitive and affective effects of seductive details in multimedia learning. *Computers in Human Behavior*, *44*, 267–278. https://doi.org/10.1016/j.chb.2014.10.061

Parmaxi, A., & Demetriou, A. A. (2020). Augmented reality in language learning: A state-of-the-art review of 2014–2019. *Journal of Computer Assisted Learning*, *36*(6), 861–875. https://doi.org/10.1111/jcal.12486

Parsons, D., & MacCallum, K. (2021). Current perspectives on augmented reality in medical education: Applications, affordances and limitations. *Advances in Medical Education and Practice*, *12*, 77–91. https://doi.org/10.2147/AMEP.S249891

Pellegrino, J. W., Alderton, D. L., & Shute, V. J. (1984). Understanding spatial ability. *Educational Psychologist*, *19*(3), 239–253. https://doi.org/10.1080/00461528409529300

Pelletier, K., Brown, M., Brooks, D. C., McCormack, M., Reeves, J., & Arbino, N. (2021). *2021 EDUCAUSE Horizon Report: Teaching and Learning Edition*. EDUCAUSE.

Pelletier, K., McCormack, M., Reeves, J., Robert, J., & Arbino, N. (2022). *2022 EDUCAUSE Horizon Report: Teaching and Learning Edition*. EDUCAUSE.

Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., & Richardson, C. (1995). A redrawn Vandenberg and Kuse mental rotations test—Different versions and factors that affect performance. *Brain and Cognition*, *28*(1), 39–58. https://doi.org/10.1006/brcg.1995.1032

Plant, R. W., & Ryan, R. M. (1985). Intrinsic motivation and the effects of self-consciousness, self-awareness, and ego-involvement: An investigation of internally controlling styles. *Journal of Personality*, *53*(3), 435–449. https://doi.org/10.1111/j.1467-6494.1985.tb00375.x

Radu, I. (2012). Why should my students use AR? A comparative review of the educational impacts of augmented-reality. *ISMAR 2012 - 11th IEEE International Symposium on Mixed and Augmented Reality 2012, Science and Technology Papers*, 313–314. https://doi.org/10.1109/ISMAR.2012.6402590

Radu, I. (2014). Augmented reality in education: A meta-review and cross-media analysis. *Personal and Ubiquitous Computing*, *18*(6), 1533–1543. https://doi.org/10.1007/s00779-013-0747-y

Radu, I., & Schneider, B. (2022). How augmented reality (AR) can help and hinder collaborative learning: A study of AR in electromagnetism education. *IEEE Transactions on Visualization and Computer Graphics*, Advance online publication. https://doi.org/10.1109/TVCG.2022.3169980

Rau, M. A. (2020). Comparing Multiple Theories about Learning with Physical and Virtual Representations: Conflicting or Complementary Effects? *Educational Psychology Review*, *32*(2), 297–325. https://doi.org/10.1007/s10648-020-09517-1

Reeves, T. (1995). Questioning the questions of instructional technology research. *Proceedings of the 1995 Annual National Convention of the Association for Educational Communications and Technology (AECT)*, 459–470. https://www.semanticscholar.org/paper/Questioning-the-Questions-of-Instructional-Reeves/0249127a9935711117fb8676629fd3c3e0846130

Reid, J., Cater, K., Fleuriot, C., & Hull, R. (2005). *Experience design guidelines for creating situated mediascapes* (HPL-2005-181; p. 71). Mobile and Media Systems Laboratory, HP Laboratories.

Rodríguez-Abad, C., Fernández-de-la-Iglesia, J.-C., Martínez-Santos, A.-E., & Rodríguez-González, R. (2021). A systematic review of augmented reality in health sciences: A guide to decision-making in higher education. *International Journal of Environmental Research and Public Health*, *18*(8), Article 4262. https://doi.org/10.3390/ijerph18084262

Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology*, *43*, 450–461. https://doi.org/10.1037/0022-3514.43.3.450

Ryan, R. M., Mims, V., & Koestner, R. (1983). Relation of reward contingency and interpersonal context to intrinsic motivation: A review and test using cognitive evaluation theory. *Journal of Personality and Social Psychology*, *45*(4), 736–750.

Salar, R., Arici, F., Caliklar, S., & Yilmaz, R. M. (2020). A model for augmented reality immersion experiences of university students studying in science education. *Journal of Science Education and Technology*, *29*(2), 257–271. https://doi.org/10.1007/s10956-019-09810-x

Scheiter, K. (2021). The learner control principle in multimedia learning. In L. Fiorella & R. E. Mayer (Eds.), *The cambridge handbook of multimedia learning* (3rd ed., pp. 418–429). Cambridge University Press. https://doi.org/10.1017/9781108894333.043

Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, *13*(2), 141–156. https://doi.org/10.1016/S0959-4752(02)00017-8

Schroeder, N. L., & Cenkci, A. T. (2018). Spatial contiguity and spatial split-attention effects in multimedia learning environments: A meta-analysis. *Educational Psychology Review*, *30*(3), 679–701. https://doi.org/10.1007/s10648-018-9435-9

Seufert, T., & Brünken, R. (2006). Cognitive load and the format of instructional aids for coherence formation. *Applied Cognitive Psychology*, *20*(3), 321–331. https://doi.org/10.1002/acp.1248

Shaghaghian, Z., Burte, H., Song, D., & Yan, W. (2022). Learning spatial transformations and their math representations through embodied learning in augmented reality. In P. Zaphiris & A. Ioannou (Eds.), *Learning and Collaboration Technologies. Novel Technological Environments* (pp. 112–128). Springer International Publishing. https://doi.org/10.1007/978-3-031-05675-8_10

Shepard, R. N., & Metzler, J. (1971). Mental totation of three-dimensional objects. *Science*, *171*(3972), 701–703. https://doi.org/10.1126/science.171.3972.701

Sırakaya, M., & Alsancak Sırakaya, D. (2020). Augmented reality in STEM education: A systematic review. *Interactive Learning Environments*, *123*, 109–123. https://doi.org/10.1080/10494820.2020.1722713

Slater, M., & Wilbur, S. (1997). A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, *6*(6), 603–616. https://doi.org/10.1162/pres.1997.6.6.603

Sommerauer, P., & Müller, O. (2014). Augmented reality in informal learning environments: A field experiment in a mathematics exhibition. *Computers & Education*, *79*, 59–68. https://doi.org/10.1016/j.compedu.2014.07.013

Sommerauer, P., & Müller, O. (2018). Augmented reality for teaching and learning—A literature review on theoretical and empirical foundations. *Twenty-Sixth European Conference on Information Systems (ECIS2018)*. https://aisel.aisnet.org/ecis2018_rp/31

Stull, A. T., & Hegarty, M. (2016). Model manipulation and learning: Fostering representational competence with virtual and concrete models. *Journal of Educational Psychology*, *108*(4), 509–527. https://doi.org/10.1037/edu0000077

Sugiura, A., Kitama, T., Toyoura, M., & Mao, X. (2019). The use of augmented reality technology in medical specimen museum tours. *Anatomical Sciences Education*, *12*(5), 561–571. https://doi.org/10.1002/ase.1822

Sundararajan, N., & Adesope, O. (2020). Keep it coherent: A meta-analysis of the seductive details effect. *Educational Psychology Review*, *32*(3), 707–734. https://doi.org/10.1007/s10648-020-09522-4

Surry, D. W., & Ensminger, D. (2001). What's wrong with media comparison studies? *Educational Technology*, *41*(4), 32–35.

Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251–296. https://doi.org/10.1023/A:1022193728205

Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, *31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

Tang, K. S., Cheng, D. L., Mi, E., & Greenberg, P. B. (2020). Augmented reality in medical education: A systematic review. *Canadian Medical Education Journal*, *11*(1), e81–e96. https://doi.org/10.36834/cmej.61705

Thees, M., Altmeyer, K., Kapp, S., Rexigel, E., Beil, F., Klein, P., Malone, S., Brünken, R., & Kuhn, J. (2022). Augmented reality for presenting real-time data during students' laboratory work: Comparing a head-mounted display with a separate display. *Frontiers in Psychology*, *13*, Article 804742. https://doi.org/10.3389/fpsyg.2022.804742

Thees, M., Kapp, S., Strzys, M. P., Beil, F., Lukowicz, P., & Kuhn, J. (2020). Effects of augmented reality on learning and cognitive load in university physics laboratory courses. *Computers in Human Behavior*, *108*, Article 106316. https://doi.org/10.1016/j.chb.2020.106316

Theodoropoulos, A., & Lepouras, G. (2021). Augmented Reality and programming education: A systematic review. *International Journal of Child-Computer Interaction*, *30*, Article 100335. https://doi.org/10.1016/j.ijcci.2021.100335

Uriarte-Portillo, A., Ibáñez, M.-B., Zataraín-Cabada, R., & Barrón-Estrada, M.-L. (2022). Higher immersive profiles improve learning outcomes in augmented reality learning environments. *Information*, *13*(5), Article 5. https://doi.org/10.3390/info13050218

Urlings, J., Sezer, S., ter Laan, M., Bartels, R., Maal, T., Boogaarts, J., & Henssen, D. (2022). The role and effectiveness of augmented reality in patient education: A systematic review of the literature. *Patient Education and Counseling*, *105*(7), 1917–1927. https://doi.org/10.1016/j.pec.2022.03.005

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, *47*(2), 599–604. https://doi.org/10.2466/pms.1978.47.2.599

Vásquez-Carbonell, M. (2022). A systematic literature review of augmented reality in engineering education: Hardware, software, student motivation & development recommendations. *Digital Education Review*, *41*, 249–267. https://doi.org/10.1344/der.2022.41.249-267

Volmer, B., Baumeister, J., Von Itzstein, S., Bornkessel-Schlesewsky, I., Schlesewsky, M., Billinghurst, M., & Thomas, B. H. (2018). A comparison of predictive spatial augmented reality cues for procedural tasks. *IEEE Transactions on Visualization and Computer Graphics*, *24*(11), 2846–2856. https://doi.org/10.1109/TVCG.2018.2868587

Wang, X.-M., Hu, Q.-N., Hwang, G.-J., & Yu, X.-H. (2022). Learning with digital technology-facilitated empathy: An augmented reality approach to enhancing students' flow experience, motivation, and achievement in a biology program. *Interactive Learning Environments*, Advance online publication. https://doi.org/10.1080/10494820.2022.2057549

Weerasinghe, M., Biener, V., Grubert, J., Quigley, A., Toniolo, A., Pucihar, K. Č., & Kljun, M. (2022). VocabulARy: Learning vocabulary in AR supported by keyword visualisations. *IEEE Transactions on Visualization and Computer Graphics*, *28*(11), 3748–3758. https://doi.org/10.1109/TVCG.2022.3203116

Wei, X., Weng, D., Liu, Y., & Wang, Y. (2015). Teaching based on augmented reality for a technical creative design course. *Computers and Education*, *81*, 221–234. https://doi.org/10.1016/j.compedu.2014.10.017

Wen, Y., & Looi, C.-K. (2019). Review of augmented reality in education: Situated learning with digital and non-digital resources. In P. Díaz, A. Ioannou, K. K. Bhagat, & J. M. Spector (Eds.), *Learning in a digital world* (pp. 179–193). Springer Singapore. https://doi.org/10.1007/978-981-13-8265-9_9

Wetzel, R., Blum, L., Broll, W., & Oppermann, L. (2011). Designing mobile augmented reality games. In B. Furht (Ed.), *Handbook of augmented reality* (pp. 513–539). Springer New York. https://doi.org/10.1007/978-1-4614-0064-6_25

Wigfield, A. (1994). Expectancy-value theory of achievement motivation: A developmental perspective. *Educational Psychology Review*, *6*(1), 49–78. https://doi.org/10.1007/BF02209024

Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology*, *25*(1), 68–81. https://doi.org/10.1006/ceps.1999.1015

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review*, *9*(4), 625–636. https://doi.org/10.3758/BF03196322

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, *7*(3), 225–240. https://doi.org/10.1162/105474698565686

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. https://doi.org/10.1016/j.compedu.2012.10.024

Wu, H.-K., & Shah, P. (2004). Exploring visuospatial thinking in chemistry learning. *Science Education*, *88*(3), 465–492. https://doi.org/10.1002/sce.10126

Xu, W.-W., Su, C.-Y., Hu, Y., & Chen, C.-H. (2022). Exploring the effectiveness and moderators of augmented reality on science learning: A meta-analysis. *Journal of Science Education and Technology*, *31*(5), 621–637. https://doi.org/10.1007/s10956-022-09982-z

Yang, S.-Y., & Tsai, C.-H. (2020). Deep motivation or surface strategy: Effects of authentic technologies on scientific learning outcomes and study approaches. *Journal of Internet Technology*, *21*(6), Article 6.

Young, M. F. (1993). Instructional design for situated learning. *Educational Technology Research and Development*, *41*(1), 43–58. https://doi.org/10.1007/BF02297091

Zhang, J., Li, G., Huang, Q., Feng, Q., & Luo, H. (2022). Augmented reality in K–12 education: A systematic review and meta-analysis of the literature from 2000 to 2020. *Sustainability*, *14*(15), Article 9725. https://doi.org/10.3390/su14159725

Zhao, S., Ni, Y., Dong, G., Tian, J., & Chen, Y. (2023). Comparing three XR technologies in reviewing performance-based building design: A pilot study of façade fenestrations. *Computer Animation and Virtual Worlds*, Article e2139. https://doi.org/10.1002/cav.2139

Zumbach, J., von Kotzebue, L., & Pirklbauer, C. (2022). Does augmented reality also augment knowledge acquisition? Augmented reality compared to reading in learning about the human digestive system? *Journal of Educational Computing Research*, *60*(5), 1325–1346. https://doi.org/10.1177/07356331211062945

8 Appendix

8.1 Paper 1 – Krüger, Buchholz & Bodemer, 2019

Krüger, J. M., Buchholz, A., & Bodemer, D. (2019). Augmented reality in education: three unique characteristics from a user's perspective. In M. Chang, H.-J. So, L.-H. Wong, F.-Y. Yu, & J. L. Shih (Eds.), *Proceedings of the 27th International Conference on Computers in Education* (pp. 412-422). Asia-Pacific Society for Computers in Education. https://apsce.net/icce/icce2019/04_Proceedings.html

Augmented Reality in Education: Three Unique Characteristics from a User's Perspective

Jule M. KRÜGER[a*], Alexander BUCHHOLZ[a] & Daniel BODEMER[a]
[a]*Media-based Knowledge Construction, University of Duisburg-Essen, Germany*
*jule.krueger@uni-due.de

Abstract: In this paper, three technological characteristics of augmented reality (AR) are reframed from a perceptual, user's perspective and discussed concerning their potential for education and in the context of research on technology-supported learning. The first characteristic, *contextuality*, describes that users of AR can experience the real world and virtual elements simultaneously. The second characteristic, *interactivity*, includes the possibilities to interact with AR through the manipulation of both real objects and virtual properties, which offers novel possibilities for interaction. The third characteristic, *spatiality*, focusses on the linking of virtual objects to specific points in space and the more realistic three-dimensionality that AR visualizations offer. It is proposed that these three characteristics can provide a way to structure the broad research landscape of AR in education and form a basis for future research projects. Two studies are presented and linked to the three characteristics. In the first study, the comparison of a desktop simulation and an AR simulation in an individual learning setting is linked to the characteristics of interactivity and spatiality. In the second study, the contextuality of AR is systematically varied and exploited to present group awareness information about other learners next to these learners instead of separated from them. The results of the studies are discussed in the context of the three characteristics and the paper concludes that there are a lot of different educational settings in which AR could be beneficial. The classification of and systematic variation in research based on the three characteristics may form a basis to systematize educational AR research. Furthermore, the results of this research and the three characteristics themselves can inform the design of AR applications to support learning.

Keywords: Augmented reality, Contextuality, Interactivity, Spatiality, Technology-supported education, Multimedia learning

1. Introduction and background

During the past centuries, augmented reality (AR) has turned from a technological vision of the future, which could often be found in science fiction movies, to a technological achievement of the present, which can now be created by the smart technological devices we have in our pockets. This development concerning the access to the necessary technology creates novel opportunities for applying AR in different fields. One area that many recent studies concerning AR focus on is education (Cipresso, Giglioli, Raya, & Riva, 2018). Education may also be one of the most promising areas for applying AR and there is an increasing number of studies that focus on the opportunities that AR as a way of visualizing information has to offer for both individual and collaborative learning settings (Akçayır & Akçayır, 2017; Phon, Ali, & Halim, 2014; Radu, 2014; Wu, Lee, Chang, & Liang, 2013). In most of these studies, advantages of AR in comparison to more traditional learning settings are examined. Positive effects that have been found when using AR in education are enhanced learning performance and motivation, higher enjoyment and engagement, more positive attitudes towards the learning material, and a better collaboration between learners (Akçayır & Akçayır, 2017; Bower, Howe, McCredie, Robinson, & Grover, 2014; Chen, Liu, Cheng, & Huang, 2017; Dunleavy & Dede, 2014; Phon et al., 2014; Radu, 2014; Saidin, Halim, & Yahaya, 2015; Wu et al., 2013). Challenges that were discovered are for example technical limitations, the use of the application being too complicated and mentally overloading, the amount of time that has to be invested to develop the applications, and pedagogical issues when trying to integrate AR into the classroom (Akçayır & Akçayır, 2017; Bower et al., 2014; Dunleavy & Dede, 2014; Radu, 2014).

Over the years, various definitions of AR have been used in different areas of research. A rather general definition describes AR as "technology which overlays virtual objects (augmented components) into the real world" (Akçayır & Akçayır, 2017, p. 1) and in earlier definitions, AR is often linked to head-mounted displays, which were the preferred display devices before smartphones and tablets were available (Azuma, 1997). One of the most commonly used definitions by Azuma (1997) defines AR as systems with three characteristics: (1) combination of the real world and virtual elements, (2) real-time interactivity, and (3) registration in 3D. The definition is used in papers by Azuma (1997) and Azuma et al. (2001), which are the two most cited papers in AR as of 2016 (Cipresso et al., 2018). This underlines the importance of this definition and those three characteristics in AR research. In the current paper, Azuma's definition is employed because of its use in the educational field (e.g., Bower et al., 2014; Radu, 2014), its scope (not too broad or too narrow), and its independence of a technological device.

In addition to different definitions, there have also been various attempts to classify AR applications and technologies (see Normand, Servières, & Moreau, 2012 for an overview). In the most known taxonomy, the Reality-Virtuality Continuum, AR is placed between the two extremes of real and virtual environment, leaning towards the side of the real environment (Milgram & Kishino, 1994). A newer taxonomy by Normand et al. (2012) classifies AR applications based on four axes, namely tracking (degrees of freedom and accuracy), augmentation type (optical see-through, video see-through, spatial augmentation), temporal base (past, present, future, time independent) and rendering modalities (beyond visual augmentation). With this taxonomy, AR applications can be classified depending on their goal and independent of the technology or the device used (Normand et al., 2012).

While the different definitions and taxonomies are often used independently of the research area, in the educational AR literature there have been attempts to connect AR to different learning theories and pedagogical approaches. Bower et al. (2014) and Dunleavy and Dede (2014) connect AR to situated and constructivist learning by assessing that learning with AR can take place at a relevant location and a deeper learning can occur with the support of AR. Game-based learning, in which immersion in the learning material is important, and inquiry-based learning, in which a scientific data gathering process is enacted, are also mentioned in connection to AR (Bower et al., 2014). In a review of the usage of learning theories to support the design of educational AR applications, Sommerauer and Müller (2018) mainly found that Mayer's multimedia principles from his Cognitive Theory of Multimedia Learning (Mayer, 2009), situated learning, game-based learning and simulations, and experiential learning were used in studies. Based on their findings, they furthermore developed a design framework that can be used for designing educational AR applications (Sommerauer & Müller, 2018).

While research on AR in education has been conducted for some time now, it is still not completely obvious how exactly AR is better for supporting learning than other learning technologies like tablet-based simulations or desktop learning environments. One key affordance of AR that Bower et al. (2014) mention is that with AR, students can rescale virtual objects of all sizes in order to better understand them. It is, however, not evident, how this is better than executing the same action on a tablet or desktop screen. Affordances of AR that are mentioned by Wu et al. (2013) and might also be true for technologies other than AR (for example a normal smartphone app), are ubiquity and situatedness, the visualization of the invisible and the bridging of formal and informal learning.

Although it is evident that these affordances all have the potential to support learning, it is not completely clear how exactly AR as a form of visualizing information plays a unique role in them. That is why, in the remainder of this paper, we aim to present and discuss three characteristics of AR that have been identified to be important factors in supporting learning. We describe how they are in this specific way only found in AR and not in other learning technologies, and thus reveal unique values that AR has for education, as proposed by Wu et al. (2013). Furthermore, we suggest that these three characteristics might provide a structure and a focus for educational AR research, to examine when and how the implementation of AR is most beneficial for education. This may help to develop a systematic research agenda for the use of AR in education scenarios and thus also support instructors and designers in developing effective AR-based learning experiences for various target groups and learning objectives in formal and informal learning settings. After the introduction and discussion of the three characteristics from a user's perspective in the next section, two studies that have been conducted on AR-supported learning are presented and discussed in the context of the characteristics. These studies exemplify how the three characteristics can be used for classifying and planning empirical research. A conclusion for the three characteristics and future research is drawn at the end of the paper.

2. Three Characteristics of AR from a User's Perspective

As stated by Hugues, Fuchs, and Nannipieri (2011), augmenting reality in itself is not possible, so that in AR a person's perception of reality is augmented. Therefore, we chose to look at the characteristics that AR possesses from a perceptual, user's perspective. In order to do this, we considered the three characteristics in the aforementioned definition of AR by Azuma (1997): (1) combination of the real world and virtual elements, (2) real-time interactivity, and (3) registration in 3D. The technology that delivers the AR experience to the user must possess these properties. In order to reframe the characteristics from a user's perspective, we looked at how they affect the user's experience of AR and propose three characteristics of the experience of using AR that cannot be found in this specific form in other technologies: *contextuality*, *interactivity*, and *spatiality*. In the following paragraphs, these three characteristics are described and their value for technology-enhanced learning is discussed. Also, four interesting research areas are given for each of the characteristics: two concerning individual learning, and two concerning collaborative learning. Table 1 shows an overview of the three characteristics.

2.1 Contextuality

In Azuma's (1997) definition, the first characteristic is that real world and virtual elements are combined in AR. From a technological perspective this means that virtual and real elements are displayed simultaneously. The displaying device must be context sensitive and aware of its location to show the user the digital content that is relevant at that place in that moment (Dunleavy & Dede, 2014).

When looking at this characteristic from a perceptual, user's perspective, this means that the user perceives the displayed virtual elements (e.g., objects, pictures, text) in the context of the real world around them (e.g., physical objects, other learners). In contrast to virtual reality, the context is not completely covered by the virtual elements, and in contrast to information on a screen, the virtual element and the context are not separated from each other (Rekimoto & Nagao, 1995). With this, novel opportunities and challenges to link the context and the virtual elements appear. Therefore, the first AR-specific characteristics reframed from a user's perspective is "contextuality".

Concerning the benefits that contextuality has for learning, it can be said that with AR it is possible to situate learning in a relevant context, which may increase the authenticity and ground students in reality (Wu et al., 2013). Even though it may also be possible to look up information that is relevant to the place where the user is at that moment with mobile devices, in AR the possibility to overlay visual virtual information over the environment gives additional potential for "perfectly situated scaffolding" (Bower et al., 2014, p. 6). Here, the relationship between the real world and the virtual information is closer than when just looking at relevant information on a mobile device. Bower et al. (2014) call the ability to contextually overlay information onto the real world one of the key pedagogical affordances of AR and Dunleavy and Dede (2014) state that embedding learning within relevant environments is very likely to enhance learning. In scientific literature, there is furthermore a connection made between the contextuality of AR and Mayer's (2009) multimedia principles of spatial and temporal contiguity (Akçayır & Akçayır, 2017; Radu, 2014). Through contextuality, instructional information can be made available at the right place and time and can this way be situated inside the real world. This implements the contiguity principles, which state that information that belongs together should be presented in an integrated way and at the same time (Mayer, 2009) in order to avoid split attention and thus increased cognitive load (Ayres & Sweller, 2014). When working in a collaborative learning setting, the contextuality of AR can also be beneficial. In co-located collaboration, contextuality means that because the virtual elements do not occlude other learners and the context, virtual information can be added to face-to-face collaborative learning settings. Learners can then perceive virtual information, the other learners, and the context around them at the same time. Here, it must be considered that a complex interplay between the three elements takes place, which might have an influence on the collaboration between the learners and their references to learning material or other external artifacts (see Bodemer, Janssen, & Schnaubert, 2018; Stahl, 2006). In general, through the characteristic of contextuality, AR has the potential to apply some of the multimedia principles onto the real world and support especially the situating of learning in a relevant environment. This provides interesting opportunities for applying AR to support learning both inside and outside the classroom.

Different questions concerning the contextuality of AR that still need to be answered through empirical research are, for example: (a) Do people indeed learn better when they are in a relevant context than when they are not and which (cognitive, motivational, and emotional) factors play a role in this?, (b) How closely must the context and the virtual information be thematically related for the overlaying of information to be beneficial?, (c) How does the interplay between learners, contexts, and virtual material have an influence on the interactions between two or more learners learning collaboratively?, (d) What are the advantages and challenges of placing group awareness information (see Bodemer et al., 2018) about other learners directly next to the respective learner? Concerning this last question, a study is presented later in this paper (study 2).

2.2 Interactivity

The second characteristic that Azuma (1997) mentions in his definition of AR is that AR elements are interactive in real time. From a technological perspective this means that the elements must be programmed to react to input that the user or – in a collaborative setting – the users give.

From a perceptual, user's perspective this entails that users experience the virtual elements reacting to their and other learners' actions. In turn, all users can react to the element's actions. In AR, virtual elements have two interactive sides. Because virtual objects in AR are placed inside the real world, they lend themselves to natural and intuitive interaction that is not possible with screen-bound virtual objects (e.g., "real" touching, gesture-based interaction). On the other hand, users can manipulate the virtual AR objects in other ways than purely physical objects (e.g., input of new data to change simulations, control through input devices) and can receive realistic and immediate feedback upon their input. This way, the interactive capabilities of real and virtual elements are combined in AR. Billinghurst and Dünser (2012), for example, state that in AR books, different forms of interaction are possible, like turning real pages to change the virtual scenery or tilting and rotating the pages to view the virtual elements from different angles. Hence, users can interact with the digital content by manipulating real objects, using a tangible interface metaphor. Therefore, a second AR-specific characteristic reframed from a user's perspective is its "interactivity".

Concerning the benefits that interactivity has for learning, it was found that even the most intuitive form of interaction with an object (i.e., perspective changing by walking around it) can be advantageous for learning (Holmes, Newcombe, & Shipley, 2018). Following embodied cognition theory, whole-body interaction with AR learning material can also lead to better learning outcomes (Johnson-Glenberg & Megowan-Romanowicz, 2017). Concerning collaborative learning settings, it can be said that in AR all learners can interact with the virtual elements in the same way and can watch how other learners interact with them. With other learning technologies, one person controls the mouse and keyboard and others watch, or everybody uses their own device to collaborate online. In AR, learners and their actions can directly be linked to each other, which may support the forming of a mental model of the other learners and thus group awareness. In general, AR's interactivity provides interesting new ways to interact with learning material, supporting learning in different settings.

Questions that still need to be answered with empirical research concerning the interactivity of AR are for example: (a) How does AR-based interaction (using a tangible interface metaphor in which interaction with an AR marker in the real world leads to manipulation of virtual objects) have a different effect on learning especially the connections between objects in comparison to a more familiar touch-based interaction with virtual objects?, (b) How must interaction with the material be designed to evoke higher order thinking processes?, (c) What influence does the collaborative interaction with the AR material have on the interaction between learners?, (d) How does watching other group members interact with the material support understanding and for example grounding processes in the group?

2.3 Spatiality

The third characteristic mentioned in the definition is that virtual elements must be registered (i.e., placed) inside the 3D real world (Azuma, 1997). From a technological perspective this means that the real world must be tracked continuously, so that the virtual element can be pinned to a specific point in space. Also, the spatial specifics like the dimensionality of the element itself need to be defined.

From a perceptual, user's perspective this means that the virtual elements should seem to exist in the same space as the real world. When virtual objects are placed inside the 3D real world, they can appear to have more spatial depth than virtual objects shown purely on flat screens. Pseudo-spatial visualizations are possible when using monocular depth cues on AR flat screens, while even true spatial visualizations can be created with the aid of binocular disparity when using AR glasses (Jeřábek, Rambousek, & Wildová, 2015). The third AR-specific characteristic reframed from a user's perspective, is thus its "spatiality".

Concerning the benefits of spatiality in educational settings it can be said that physical 3D objects were found to be better for learning than 3D computer models (Preece, Williams, Lam, & Weller, 2013). When looking at the spatial properties of 3D AR models, they lie between physical and computer models, so that they may also be more beneficial for learning than normal computer models. Advantages concerning the mental load of participants using a 3D visualization to learn a visual motor task over using a 2D visualization could also be found (Dan & Reiner, 2017). AR might be especially useful for learning the spatial structure of 3D material (Radu, 2014) and subjects with a spatial component are learned more effectively with AR (Billinghurst & Dünser, 2012). In collaborative learning settings, an example of how the fixation of an AR object to a point in space can be used is through knowledge sharing by tagging and annotating objects (Specht, Ternier, & Greller, 2011). The objects over which the learners collaborate or which they create collaboratively can also be three-dimensional and fixed to one point in space. This may offer various advantages over working together on two-dimensional screen-based material. In general, it can be said that learners may especially benefit from AR's spatiality when learning about spatial structures and relationships.

Questions that arise and should be answered through empirical research are for example: (a) Is using a three-dimensional AR object as beneficial for learning spatial structures as real objects are, in comparison to screen-based objects?, (b) How much does the use of stereoscopic AR glasses in comparison to screen-based monoscopic AR influence the spatial perception of an object and what are the advantages concerning the spatial understanding the user acquires about it?, (c) Does the collaborative creation of a three-dimensional artefact lead to better learning than the creation of a two-dimensional artefact?, (d) How exactly does using a whole room as a space to learn in together instead of a shared screen influence the interaction with the material and between the learners?

Table 1

Three Characteristics of AR from a User's Perspective

Azuma's characteristic	**User perspective characteristic**	**Description**
Combination of the real world and virtual elements	Contextuality	▪ users perceive virtual elements simultaneously with real world (including other users) around it ▪ users do not perceive virtual elements and context (including other users) separately
Real-time interactivity	Interactivity	▪ users experience virtual elements reacting to them and other users, and experience themselves and other users reacting to actions of the elements ▪ interactive properties of physical AND virtual elements
Registration in 3D	Spatiality	▪ virtual elements placed inside the 3D real world appear as if they were really there ▪ virtual elements appear more spatial than if shown on screen

2.4 Interplay of the three characteristics

The three characteristics of AR and their advantages for educational settings are not only interesting on their own, but also in their combination into one experience. Moving around a virtual AR object and looking at it from all perspectives, for example, concerns both interactivity and spatiality of AR. When the object stays in one place, it reacts to the user's movement (interactivity), which is possible because the object is fixed to a point in 3D space (spatiality). The authenticity of an experience can also be influenced by all three characteristics. Authenticity can imply the placement of a virtual object in a relevant, authentic environment (contextuality). It can also refer to the authenticity of the object itself, including its 3D presentation (spatiality). Furthermore, authenticity may imply authentic interaction with the virtual object (interactivity). An authentic virtual object placed in a relevant, authentic real-world environment and with authentic interactive properties, may provide the most authentic experience for learners.

This shows that the three characteristics cannot always be considered separately but can interact with each other. It is important to examine them through experimental research both separately and in interaction, to get an overarching picture of how AR can be used best in educational settings. In the following sections, we present two experimental studies that we executed concerning the use of AR in different educational settings: study 1 as an example of considering different characteristics (interactivity and spatiality) in an individual setting, study 2 as an approach of systematically varying one of the proposed characteristics (contextuality) in a collaborative setting. This way, two quite different ways of using the characteristics to structure and design empirical research are presented.

3. Study 1: Interactivity and spatiality in an individual setting

The first experimental study is based on research about learning with computer simulations. Using computer simulations paired with inquiry-based learning instructions like scientific discovery learning proved to be valuable in many ways for the learner to comprehend complex concepts in research contexts and practical applications (de Jong, 1991; de Jong & van Joolingen, 1998). AR applications for learning purposes can also be understood as (interactive) computer simulations or visualizations, but research about learning with AR applications rarely explored the fact that traditional and AR simulations share common concepts but differ in various aspects. It is unclear whether the learning benefits in working with AR applications found in these studies were due to the AR aspect of the application or because the learning material was a simulation or interactive visualization instead of traditional paper and text. The aim of this study was to compare a traditional (tablet-based) computer simulation with an AR version of the application with regards to their effects on conceptual knowledge, cognitive load, motivation, and spatial abilities of the learners. Although the study was not planned based on the three proposed characteristics, when comparing the AR and non-AR applications used, it shows that interactivity and spatiality differ between them. Concerning interactivity, it can be said that while in AR the users moved around the simulation and interacted with a real object (AR marker) to manipulate it, in the traditional simulation they used touch-based drag-and-drop on a tablet. Spatiality differed in the two applications in that virtual AR objects appear to be more spatial because the user has the reference of the real world, while this is not the case in a normal screen-based simulation.

3.1 Method

For this study, two almost identical computer simulations were developed and compared in an experimental laboratory setting: a normal computer simulation of a power plant on a tablet, and an AR simulation with AR markers and the tablet as a video-see-through display for AR elements. The two simulations differed regarding their interactivity and spatiality as described in the previous paragraph. During the experiment the participants (N = 56) followed a scientific discovery-based learning script with the goal of comprehending the underlying concept of power plants by building their own, changing the composition of the plant components, and first hypothesizing and then observing the outcome. The participants were randomly assigned to use either the traditional (n_t = 28) or the AR simulation (n_{AR} = 28). It was hypothesized that after the interaction with the material, participants have equivalent conceptual knowledge and cognitive load during the learning process as well as improved spatial abilities and motivation when learning with the AR simulation compared to the traditional simulation. Based on this, three TOST equivalence tests and five t-tests were executed to analyze the data.

3.2 Results

The equivalence tests were all executed for the equivalence bounds Cohen's d = +/-0.67, based on the smallest detectable effect with this sample size. The hypothesis that conceptual knowledge was equivalent in the two simulations could be supported (M_t = 12.79, SD_t = 3.06; M_{AR} = 12.64, SD_{AR} = 2.84), 90% CI for d [-0.40;0.49], lower bound, $t(54)$ = 2.69, p = .005, upper bound, $t(54)$ = -2.33, p = .012. An equivalence of intrinsic cognitive load in the simulations was also found (M_t = 4.88, SD_t = 1.95; M_{AR} = 5.02, SD_{AR} = 1.98), 90% CI for d [-0.52;0.38], lower bound, $t(54)$ = 2.25, p = .014, upper bound, $t(54)$ = -2.76, p = .004. For extraneous cognitive load, equivalence in the simulations could not be concluded (M_t = 1.21, SD_t = 1.35; M_{AR} = 1.58, SD_{AR} = 1.42), 90% CI for d [-0.72;0.18], lower bound,

$t(54) = 1.50$, $p = .070$, upper bound, $t(54) = -3.52$, $p < .001$. Based on three t-tests, no significant differences were found between the groups for these three variables.

The hypothesis concerning the difference in the resulting spatial abilities was not supported, as no significant difference between the traditional ($M_t = 7.96$, $SD_t = 4.15$) and the AR simulation ($M_{AR} = 9.07$, $SD_{AR} = 5.00$) was found, $t(54) = -0.90$, $p = .371$, $d = -0.25$. Contrary to expectations, motivation did also not differ between the two forms of simulation: intrinsic motivation ($M_t = 5.51$, $SD_t = 1.18$; $M_{AR} = 5.78$, $SD_{AR} = 0.94$), $t(54) = 0.94$, $p = .352$, $d = 0.26$, identified regulation ($M_t = 4.61$, $SD_t = 1.30$; $M_{AR} = 5.14$, $SD_{AR} = 0.99$), $t(54) = -1.73$, $p = .089$, $d = -0.47$, external regulation ($M_t = 4.69$, $SD_t = 0.98$; $M_{AR} = 4.46$, $SD_{AR} = 1.00$), $t(54) = 0.88$, $p = .382$, $d = 0.24$, and amotivation ($M_t = 2.58$, $SD_t = 1.18$; $M_{AR} = 2.19$, $SD_{AR} = 1.18$), $t(54) = 1.25$, $p = .217$, $d = 0.34$.

3.3 Discussion

The results of this study indicate that just transferring a desktop simulation into an AR simulation and thus manipulating interactivity and spatiality together might not be enough to be more beneficial for the learner regarding conceptual knowledge, motivation, cognitive load and spatial abilities. After using the application, the participants learning with the AR simulation had equal conceptual knowledge and intrinsic cognitive load and nearly equal extraneous cognitive load as the participants using the traditional simulation. The groups did not differ in motivational aspects and spatial abilities. Still, this experiment can serve as an initial study to find out more about how the three characteristics influence learning. In this study, both interactivity and spatiality were manipulated in the applications. To find out more about the specific benefits the two characteristics and their interaction have on learning processes and outcomes, more systematic studies are necessary in which interactivity and spatiality are varied separately. Furthermore, AR offers other possibilities than the ones varied in this study. A procedural simulation or visualization where the learner can use the application directly in the environment where the knowledge domain is registered (based on the characteristic of contextuality) might be more beneficial to the learner regarding learning outcomes and learning related variables. This also requires more research in the form of an experiment with systematically manipulated predictor variables.

4. Study 2: Contextuality in a collaborative setting

A further experimental study that was systematically planned and executed based on one of the three characteristics has focused on how to use the potential of AR's contextuality in a collaborative setting. Due to contextuality, the user can perceive virtual information, other learners, and the environment simultaneously. This way, virtual information can be shown exactly at the right time and place. As suggested by Radu (2014), this characteristic can be connected to Mayer's (2009) multimedia principles of spatial and temporal contiguity which state that information that belongs to each other should be presented at the same time and close to each other, preventing the splitting of attention and decreasing extraneous cognitive load. In computer-supported collaborative learning (CSCL), group awareness tools (GATs) can be used to support collaborative learning processes (Bodemer et al., 2018). As GATs provide contextual information about the social learning environment, it is crucial that they do not divert attention from germane learning activities. When group awareness (GA) information about other learners is visualized in face-to-face collaborative settings, this information is often printed out or shown on a screen, which means that the given information is separated from the context in which it is relevant (i.e., the collaboration with the other person) due to the medium that delivers it. This could especially be a problem in bigger groups of learners, because the correct GA information must still be connected to the right person. AR's unique characteristic of contextuality provides the opportunity to show GA information directly next to the corresponding person. Similar to the work of Holstein, Hong, Tegene, McLaren, & Aleven (2018), where teachers were provided with real-time information about their students' learning process through augmented reality glasses, this GA information could be presented directly over or next to the corresponding student. In this study, the systematic variation in the two conditions was thus based on contextuality so that in the AR condition the information and the context were integrated, while in the non-AR condition they were separated from each other. The aim of the study was to find out whether placing information about people directly next to them in comparison to placing it further away has an influence on cognitive load and retention of the information.

4.1 Method

To compare the visualization of GA information next to people and further away from them, we used pictures instead of a real implementation in AR to investigate the characteristic of contextuality in a controlled laboratory setting. In the study, the participants (N = 38) worked on tasks in which they had to form study groups of the people shown to them in pictures based on the GA information given about them. The participants were randomly assigned to one of two conditions: GA information visualized directly next to the corresponding person in the picture (AR mockup; n_{AR} = 18) or GA information shown separately below the pictures (n_{nonAR} = 20). In the different tasks given to the participants, the number of people shown to them was varied between two and ten people to see if an effect of the proximity of the information differs with a differing number of people. The two independent variables were thus the proximity of the information to the people (between-subject) and the number of people displayed (within-subject). It was hypothesized that these two factors and their interaction influence the cognitive load of the participants as measured continuously through a secondary reaction task and the efficiency in executing the task as measured by their time spent on the task. Furthermore, it was expected that the proximity of the information influences the participants' self-reported extraneous cognitive load and their recall of the GA information. Two mixed-design ANOVAs and two independent samples t-tests were used to analyze the data based on these hypotheses.

4.2 Results

The hypothesis that the proximity of the information has an influence on the continuously measured cognitive load could be supported with a significantly slower reaction time (ms) in the group where picture and information were shown further apart (M_{nonAR} = 2197.83, SD_{nonAR} = 1863.29; M_{AR} = 1153.49, SD_{AR} = 744.64), $F(1,36) = 4.93$, $p = .033$, $\eta_p^2 = 0.12$. The same pattern was found for the time spent on the task, where the group with the separate information presentation needed more time (s) to solve the tasks than the group with the integrated visualization (M_{nonAR} = 62.86, SD_{nonAR} = 19.76; M_{AR} = 50.93, SD_{AR} = 11.11) , $F(1,36) = 5.11$, $p = .030$, $\eta_p^2 = 0.12$. Concerning the within-subject factor (number of people), it can be said that even though more people shown generally meant both a longer time spent on the task, $F(1.24, 44.44) = 21.11$, $p < .001$, $\eta_p^2 = 0.37$, and a longer reaction time in the secondary task, $F(2.531, 91.13) = 4.933$, $p = .005$, $\eta_p^2 = 0.12$, this pattern was not found for all pairwise comparisons. No significant interaction effect was found for reaction time, $F(2.53, 91.13) = 1,47$, $p = .233$, $\eta_p^2 = 0.04$, or time spent on the task, $F(1.24, 44.44) = 1.25$, $p = .279$, $\eta_p^2 = 0.03$.

Concerning the variables that were not measured for every single task, no significant difference was found in either self-reported extraneous cognitive load (M_{AR} = 4.07, SD_{AR} = 1.68; M_{nonAR} = 4.25, SD_{nonAR} = 1.67), $t(36) = -0.32$, $p = .748$, $d = -0.11$, or recall of the GA information between the two groups (M_{AR} = 2.39, SD_{AR} = 1.29; M_{nonAR} = 2.20, SD_{nonAR} = 1.11), $t(36) = 0.49$, $p = .630$, $d = 0.16$.

4.3 Discussion

In this study, in which contextuality was varied systematically, significant differences between the two groups concerning the reaction time in a secondary task and the time on task were found. The participants in the AR mockup group needed less time for solving the tasks and reacted faster on the secondary task, which shows that they were more efficient and less cognitively occupied in their task of forming study groups based on the information about the people. However, these results could not be supported by the results in the self-reported cognitive load and recall of the information, which did not differ between the groups. A confounding variable that might have led to the differences in the timings between the groups was that the participants from the non-AR group had to scroll down on the pages with the tasks, while the others did not. In a future study, this factor must be held stable between the groups. Also, other objective measures for cognitive load, which should not be influenced by scrolling (e.g., eye-tracking metrics), might be used to compare the two forms of visualization in a future study. A factor that may have led to less differences between the groups is that the tasks could be solved without even looking at the pictures of the people. This way, the participants might not even have made the connection between the people and the information. Split attention only happens when one part of the material is not understandable without the other (Ayres & Sweller, 2014). This was not the case here and an adapted study design should be considered for future studies.

5. Conclusion

In this paper, three characteristics of AR are reframed from a user's perspective and discussed in relation to their potential for supporting individual and collaborative learning. It is proposed that these three characteristics can be used as a basis for researching AR in educational settings and two studies which have been executed with the three characteristics in mind are presented.

The two studies differed considerably in their usage of the characteristics. In study 1, the experimental manipulation can be classified into two of the characteristics, namely interactivity and spatiality. Concerning this study, we conclude that to get a more complete picture, follow-up studies are necessary in which the two characteristics are varied separately and systematically. This way, their influence on learning processes and outcomes can be determined. In study 2, a systematic experimental variation based on the characteristic of contextuality took place and positive effects on efficiency and cognitive load could be found. Due to confounding variables, the results of the study should be interpreted with caution. Follow-up studies that control for these factors are needed to confirm the results concerning the increased efficiency and decreased cognitive load in the setting.

While contextuality, interactivity, and spatiality all seem to be important for using AR in educational settings, more systematic empirical research concerning their potentials, their impact and their interplay is necessary. Based on the two presented studies, which initialized the research on AR in education at our lab, more empirical studies with systematic variations based on the three characteristics are currently conducted and planned, such as two experimental studies that intend to systematically disentangle the characteristics of interactivity and spatiality.

AR-supported learning experiences have the potential to be applied in different settings and with various goals, which can also be seen in the differences between the two presented studies. Thus, systematic AR-related research findings can enrich the design of formal and informal educational environments for individual and social learning of diverse students. In order to provide a structuring basis for this heterogeneous research field, the three characteristics contextuality, interactivity, and spatiality are proposed to serve as common denominators for the users' experience of AR in a wide range of learning settings.

Acknowledgements

We would like to thank Sophie-Marie Zentarra for collecting the data of the second study in the course of her Bachelor's Thesis.

References

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. https://doi.org/10.1016/j.edurev.2016.11.002

Ayres, P., & Sweller, J. (2014). The split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 206–226). Cambridge: Cambridge University Press.

Azuma, R. (1997). A survey of augmented reality. *Presence: Teleoperators and Virtual Environments*, *6*(4), 355–385. https://doi.org/10.1162/pres.1997.6.4.355

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. https://doi.org/10.1109/38.963459

Billinghurst, M., & Dünser, A. (2012). Augmented reality in the classroom. *Computer*, *45*(7), 56–63. https://doi.org/10.1109/MC.2012.111

Bodemer, D., Janssen, J., & Schnaubert, L. (2018). Group awareness tools for computer-supported collaborative learning. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), *International Handbook of the Learning Sciences* (pp. 351–358). New York, NY: Routledge/Taylor & Francis.

Bower, M., Howe, C., McCredie, N., Robinson, A., & Grover, D. (2014). Augmented Reality in education—Cases, places and potentials. *Educational Media International*, *51*(1), 1–15. https://doi.org/10.1080/09523987.2014.889400

Chen, P., Liu, X., Cheng, W., & Huang, R. (2017). A review of using augmented reality in education from 2011 to 2016. In E. Popescu, Kinshuk, M. K. Khribi, R. Huang, M. Jemni, N.-S. Chen, & D. G. Sampson (Eds.), *Innovations in Smart Learning, Lecture Notes in Educational Technology* (pp. 13–18). Singapore: Springer Singapore.

Cipresso, P., Giglioli, I. A. C., Raya, M. A., & Riva, G. (2018). The past, present, and future of virtual and augmented reality research: A network and cluster analysis of the literature. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.02086

Dan, A., & Reiner, M. (2017). EEG-based cognitive load of processing events in 3D virtual worlds is lower than processing events in 2D displays. *International Journal of Psychophysiology*, *122*, 75–84. https://doi.org/10.1016/j.ijpsycho.2016.08.013

de Jong, T. (1991). Learning and instruction with computer simulations. *Education and Computing*, *6*(3), 217–229. https://doi.org/10.1016/0167-9287(91)80002-F

de Jong, T., & van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research*, *68*(2), 179–201. https://doi.org/10.3102/00346543068002179

Dunleavy, M., & Dede, C. (2014). Augmented reality teaching and learning. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of Research on Educational Communications and Technology* (4th ed., pp. 735–745). New York, NY: Springer New York.

Holmes, C. A., Newcombe, N. S., & Shipley, T. F. (2018). Move to learn: Integrating spatial information from multiple viewpoints. *Cognition*, *178*, 7–25. https://doi.org/10.1016/j.cognition.2018.05.003

Holstein, K., Hong, G., Tegene, M., McLaren, B. M., & Aleven, V. (2018). The classroom as a dashboard: Co-designing wearable cognitive augmentation for K-12 teachers. *LAK'18: 8th International Conference on Learning Analytics and Knowledge*, 79–88. https://doi.org/10.1145/3170358.3170377

Hugues, O., Fuchs, P., & Nannipieri, O. (2011). New augmented reality taxonomy: Technologies and features of augmented environment. In B. Furht (Ed.), *Handbook of Augmented Reality* (pp. 47–63). New York, NY: Springer New York.

Jeřábek, T., Rambousek, V., & Wildová, R. (2015). Perceptual specifics and categorisation of augmented reality systems. *Procedia - Social and Behavioral Sciences*, *191*, 1740–1744. https://doi.org/10.1016/j.sbspro.2015.04.419

Johnson-Glenberg, M. C., & Megowan-Romanowicz, C. (2017). Embodied science and mixed reality: How gesture and motion capture affect physics education. *Cognitive Research: Principles and Implications*, *2*(1), 24. https://doi.org/10.1186/s41235-017-0060-9

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

Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems*, *E77-D*(12), 1321–1329.

Normand, J., Servières, M., & Moreau, G. (2012). A new typology of augmented reality applications. *Augmented Human*, 1–8. https://doi.org/10.1145/2160125.2160143

Phon, D. N. E., Ali, M. B., & Halim, N. D. A. (2014). Collaborative augmented reality in education: A review. *2014 International Conference on Teaching and Learning in Computing and Engineering*, 78–83. https://doi.org/10.1109/LaTiCE.2014.23

Preece, D., Williams, S. B., Lam, R., & Weller, R. (2013). "Let's Get Physical": Advantages of a physical model over 3D computer models and textbooks in learning imaging anatomy. *Anatomical Sciences Education*, *6*(4), 216–224. https://doi.org/10.1002/ase.1345

Radu, I. (2014). Augmented reality in education: A meta-review and cross-media analysis. *Personal and Ubiquitous Computing*, *18*(6), 1533–1543. https://doi.org/10.1007/s00779-013-0747-y

Rekimoto, J., & Nagao, K. (1995). The world through the computer: Computer augmented interaction with real world environments. *Proceedings of the 8th Annual ACM Symposium on User Interface and Software Technology - UIST '95*, 29–36. https://doi.org/10.1145/215585.215639

Saidin, N. F., Halim, N. D. A., & Yahaya, N. (2015). A review of research on augmented reality in education: Advantages and applications. *International Education Studies*, *8*(13), 1–8. https://doi.org/10.5539/ies.v8n13p1

Sommerauer, P., & Müller, O. (2018). Augmented reality for teaching and learning—A literature review on theoretical and empirical foundations. *Twenty-Sixth European Conference on Information Systems (ECIS2018)*. Retrieved from https://aisel.aisnet.org/ecis2018_rp/31

Specht, M., Ternier, S., & Greller, W. (2011). Dimensions of mobile augmented reality for learning: A first inventory. *Journal of the Research for Educational Technology*, *7*(1), 117–127.

Stahl, G. (2006). *Group cognition: Computer support for building collaborative knowledge*. Cambridge, Mass: MIT Press.

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. https://doi.org/10.1016/j.compedu.2012.10.024

8.2 Paper 2 – Krüger & Bodemer, 2020

Krüger, J. M., & Bodemer, D. (2020). Different types of interaction with augmented reality learning material. In D. Economou, A. Klippel, H. Dodds, A. Peña-Rios, M. J. W. Lee, D. Beck, J. Pirker, A. Dengel, T. M. Peres, & J. Richter (Eds.), *2020 6th International Conference of the Immersive Learning Research Network (iLRN)* (pp. 78–85). Immersive Learning Research Network. https://doi.org/10.23919/iLRN47897.2020.9155148

Different Types of Interaction with Augmented Reality Learning Material

Jule M. Krüger
Media-based Knowledge Construction
University of Duisburg-Essen
Duisburg, Germany
jule.krueger@uni-due.de

Daniel Bodemer
Media-based Knowledge Construction
University of Duisburg-Essen
Duisburg, Germany
bodemer@uni-due.de

***Abstract*—In this paper, a study with the focus on interactivity in augmented reality (AR) applications concerning the influence of different forms of interaction with AR learning material is presented. While research on multimedia learning often distinguishes between mental and physical interaction with learning material, other research fields state that physical interaction is necessary to interact mentally. To look at how this distinction may play a role in AR-based learning material, an experimental study with a 2x2 design manipulating mental and physical interaction was conducted, including learning material on the topic of power plants. The data ($N = 128$) were collected and analyzed, showing that, although not expected, learning was better in groups in which either more physical or more mental interaction was applied, but not in groups in which both were high. The results are discussed under the potential idea of cognitive overload.**

***Index Terms*—augmented reality, multimedia learning, interactivity, interactive learning, technology-enhanced learning**

I. Introduction

Augmented reality (AR) is a form of visualizing information in which real and virtual elements are combined. Through this combination, a lot of new possibilities arise for the design of educational material, so that there are many variables to consider when using AR to support education. Most research on AR applications includes the comparison of an AR setting with a traditional learning setting. This way, it is difficult to deduce why exactly an AR setting may lead to increased knowledge, because it is not clear whether observed differences are elicited, for example, by differences in dynamicity, interactivity or dimensionality of the media of presentation or even differences in the information which is made available through the different media [1]. To structure AR from a human-centered perspective, and to initiate more systematic research and design activities regarding educational AR applications, the consideration of three characteristics has been suggested: contextuality (c), interactivity (i) and spatiality (s; ARcis characteristics) [2]. While the focus of contextuality is on the real environment in which the learners are located and its (thematic) relationship to the perceived virtual elements, interactivity focusses mainly on the interaction of the learners with both real elements (including the ones anchoring virtual elements to the real world) and virtual elements. Spatiality especially considers the spatial properties of the virtual elements and their spatial connections to the real environment. In the following sections, a short overview over the three characteristics and examples of their role in typical AR learning applications are given, with a focus on interactivity. Afterwards, an experimental study concerning interaction with AR learning material is described and its results are presented and discussed.

II. Three AR Characteristics

Of the three characteristics discussed in [2], contextuality focusses on the real-world environment that is the basis for the AR experience and its relation to the virtual elements. While the real-world context can be relevant for the virtual elements or the learning material in general, it may also have no thematic relevance and only be there as a space in which the virtual elements are viewed [2]. For example, context plays an important role when cultural heritage sites are augmented with educational information. Learners can be immersed in the real environment at a historical site and still receive additional information that helps them understand historic events. In this sense, [3] present an application in which the Bergen-Belsen memorial site is augmented with virtual content. Learners can be at the real-world place and view a virtual reconstruction of the buildings of 1945. The meaning of the virtual information in the AR application would not be the same if used somewhere else, showing the thematic relevance of the real-world context for the virtual elements.

Spatiality focusses on the virtual elements that are placed inside the real world through AR. The virtual elements have spatial properties and can be two- or three-dimensional (e.g., a picture or a model of an object). Also, the virtual elements are linked to a point in the 3D real world, for example through an AR marker or object surface recognition [2]. When spatial object knowledge should be acquired the spatial properties of the virtual elements in AR are important. This is, for example, the case during learning about human anatomy. [4] present an application in which learners can view a 3D model of a human skeleton. This way, spatial relations between bones can be seen, so that viewing the same object in only two dimensions would not offer the same amount of information.

Interactivity, the characteristic in focus in the present paper, focusses on the learners and their connection to the information. AR can provide learners with the opportunity to interact with virtual elements, the real world, and real-world anchors of virtual elements (e.g., AR markers) [2]. This interaction can be more or less elaborate. Less elaborate interaction can mean that learners are just able to walk around an object, which is the case in most AR applications in which the virtual element is linked to a specific point in the 3D real world. Already this basic form of interaction may lead to enhanced learning by enabling the learner to view an object from different perspectives [5]. But AR also has the potential for more elaborate interaction with

978-1-7348995-0-4/20/$31.00 ©2020 Immersive Learning Research Network
6th International Conference of the Immersive Learning Research Network (iLRN 2020)
Online, June 21-25, 2020.

learning material, which motivates the present paper and study. In the following section, information on interactivity in learning contexts and specifically in AR-based settings is provided.

III. Interactivity in Learning Contexts

A. Interactive Learning

In multimedia learning, interactivity is an important topic, although it is defined in many different ways [6]. Moreno and Mayer [7] define interactive multimodal learning environments as environments in which actions of learners induce events. They state that it is important to design the interaction in a way that it does not overload the learners' cognitive capacities in order to induce learning. In research on multimedia-based learning, it is often stated that mental engagement or interaction with learning material leads to better and deeper information processing and learning than physical interaction [8]. Physical interaction with learning material may even obstruct learning processes through the induction of cognitive processes not relevant for learning. In [9], for example, learning with a learner-generated graphic organizer and learning with a graphic organizer that was given to the learners were compared. The results showed that the learners who were provided with graphic organizers they had not made themselves, scored higher on a transfer test. The authors explain this result on the grounds of cognitive load theory, stating that the extraneous processing required when generating the graphic organizers comes at the cost of generative processing and thus deeper learning. In cognitive load theory, three types of cognitive load are differentiated: intrinsic, extraneous and germane cognitive load [10]. While intrinsic cognitive load (ICL) focusses on the intrinsic difficulty of the learning task, extraneous cognitive load (ECL) depends on the design of the instructional material and has nothing to do with the actual learning task. Germane cognitive load (GCL) is defined as the load that has to do with activities leading to learning. While ECL should be decreased, germane cognitive load should be increased within the limits of working memory.

In a different approach towards interaction with learning material, embodied cognition theory states that cognitive processes are rooted within the interaction of the body with the physical world, so that learning is supported by physical interaction with learning material [11]. The theory thus implies that the "body plays a central role in shaping the mind" [12, p. 970], although more theory and empirical research is still needed in this area. In the ICAP-Framework by Chi and Wiley [13] it is also postulated that overt physical behavior is an indicator for engagement with learning material. The authors say that the most learning takes place when learners are more engaged with learning material (e.g., constructing, interacting).

Based on the described theories and frameworks, physical interaction with learning material might either hinder or explicitly support learning processes. The different views can be brought together in their agreement that physical interaction with learning material must induce germane cognitive processes and not just extraneous load to support learning processes. As AR offers new ways of interacting with physical objects to manipulate virtual objects, it is important to find out when the exploitation of this possibility leads to the desired learning.

B. Interaction in AR

In AR, the visualized information can be based in multiple media, representing a specific case of multimedia-based learning. Systematic research concerning physical and mental interaction has been executed with traditional multimedia learning material (e.g., [9], [14], see also [6], [7]), but not as systematically with AR multimedia learning material. Interacting with AR can differ from interacting with other multimedia learning material through the unique combination of virtual elements and the real world, as it enables interaction with the learning material on different levels. With AR-based learning material, quite elaborate interaction is possible through the combination of interactive properties of reality and virtuality [2]. On the real level, learners can interact with the real-world environment by moving around freely or moving real-world elements just as they would in non-augmented reality. On the virtual level, learners can interact with virtual elements through the devices they use for the visualization of the elements (e.g., through a controller or touch-based interaction with a tablet). A third level of interaction that needs to be considered in AR is the link between the real and the virtual level. Through this connection, manipulation of elements on the real level can lead to interaction with elements on the virtual level. In practice, this connection can be maintained through the use of AR markers.

AR markers are real-world anchors for virtual elements in AR. Those markers can, for example, be 2D photos or drawings, but also 3D objects. Markers can be made specifically for the purpose of linking the real to the virtual (e.g., marker cubes as the ones used in [15]), but they can also be already existing pictures or objects that have been repurposed for this goal (e.g., the wooden statue in [16]). In a quite elaborate way of interacting with AR learning material in [1], a physical model of an audio speaker is connected to 3D visualizations of invisible phenomena like electric current, magnetic fields and soundwaves. The virtual visualizations react to manipulations of the physical model in real time, so that learners should be able to easily link their actions to normally invisible physical phenomena. Because so many different forms of interaction with AR learning material are possible, it is important to investigate them more closely in systematic research.

In the present experimental study, physical interaction with the AR learning material mainly takes place as a manipulation of virtual elements through an interaction with real-world objects, thus using a tangible interface metaphor [17]. Here, learners can move around AR paper markers that show different components of a power plant when scanned with a tablet application. This way, moving the physical, real object (paper marker) leads to movement of the virtual object (power plant component) connected to it. Furthermore, when combining the paper markers and thus power plant components in the right pattern, a working power plant with animations and connections between the components is shown. The learners can then retrieve information on how much electricity the built power plant is generating and how efficient it is, so that real-time feedback based on the position of the physical markers is given via the virtual channel both by animations and information display.

C. Research Question and Hypotheses

The goal of the experimental study is to take a closer look at how learners learn with AR-based learning material that differs in its interactive possibilities and demand. This was accomplished by providing learners with different material (physical interaction: high vs. low) and different task information (mental interaction: high vs. low) during a learning task. As not a lot of research including systematically varied interaction within AR learning material has been executed, the research question that should be answered in the present paper asks how mental and physical interaction with AR learning material have an influence on learning processes and outcomes.

Based on the literature, a first set of hypotheses on the influence of the type of interaction on learning outcomes was formulated. It was hypothesized that, by eliciting germane cognitive processes, both high mental interaction (Hyp. 1a) and high physical interaction (Hyp. 1b) would lead to more knowledge compared to respectively low mental and low physical interaction. Based on the idea that physical interaction could also hinder learning by inducing extraneous cognitive processes, it was further hypothesized that the positive effect of higher mental interaction on knowledge would be bigger than the positive effect of higher physical interaction (Hyp. 1c).

In a second set of hypotheses, the influence of the type of interaction on the above-mentioned two types of cognitive load was focused. Analogous to the first set of hypotheses, it was hypothesized that both high mental interaction (Hyp. 2a) and high physical interaction (Hyp. 2b) would lead to higher GCL compared to respectively low mental and low physical interaction. It was furthermore suggested that the positive effect of higher mental interaction on GCL would be bigger than the positive effect of higher physical interaction (Hyp. 2c), due to potentially more elaborate elicitation of germane cognitive processes. Also analogous to the first set of hypotheses, it was hypothesized that high physical interaction would lead to higher ECL than low physical interaction (Hyp. 2d), due to a potential elicitation of extraneous cognitive processes.

In addition to these hypotheses, the influence of mental and physical interaction on different types of task load will be examined in an explorative way. Furthermore, correlations between the mentioned variables knowledge, cognitive load types and task load will be explored. Although data on more outcome variables were collected during the study, this paper will only focus on the ones mentioned here.

IV. Method

A. Participants

In total, 136 people took part in the study. Four participants were filtered out due to technical errors in the first trials, after which the AR application was changed a little. Two participants were filtered out because their proficiency of the German language was not elaborate enough for the learning material. Two other participants were filtered out because their study courses were too closely related to the learning topic. The final sample consists of 128 people (39 males, 89 females) who were between 18 and 40 years old ($M = 22.55$, $SD = 3.90$). Most of them (96.09 %) were enrolled in study programs not related to the learning topic. The other participants were either employed, pre-university students or unemployed. As the participants used an AR application to learn in the study, they were asked about their experience with using general mobile applications, learning applications, general AR applications and AR learning applications on tablets or smartphones. On a scale from "never" (1) to "regularly" (5), the participants answered how often they had used those kinds of applications in the past. While they had on average used general mobile applications quite regularly ($M = 4.70$, $SD = 0.61$), they had not used learning applications that often ($M = 2.48$, $SD = 1.07$). Their experience with AR applications was even lower, with very rare use of general AR applications ($M = 1.58$, $SD = 0.84$) and nearly any experience in using AR learning applications ($M = 1.09$, $SD = 0.31$). This shows that although the participants had a lot of experience in general with using a technological device like the one used in the study and some experience with using learning applications, they hardly had any experience in using AR applications. For most of the participants, the use of an AR application (58.59 %) and specifically the use of an AR learning application (92.19 %) in the study was their first use ever. No significant differences on the different forms of experience with mobile technologies between the groups were found in the data. The participants received either course credit or money for their participation in the study. The study was assessed and approved by the ethics committee of the University of Duisburg-Essen.

B. Design

In the study, a 2x2 between-subject design was administered. One factor was the physical interaction, which could be either low (p–) or high (p+). The other factor was the mental interaction, which could also be either low (m–) or high (m+). The participants were distributed randomly into one of the four groups (p–/m–, p–/m+, p+/m–, p+/m+).

C. Materials and Apparatus

During the study, the participants used an application on an android tablet (HUAWEI MediaPad M5; 10.8-inch display, 2560 x 1600 pixels). The application was programmed in the Vuforia plugin in Unity 3D and installed on the tablet through an APK-file. With the application, printed paper markers can be scanned to view different components of a combined cycle power plant. Those components can be adjoined, whereupon animations are added in accordance with their functions. When a correctly built, working power plant is scanned completely, its electricity output and efficiency are shown in the right top corner on the tablet screen (see Fig. 1 for a screenshot).

To control for potential differences between the four groups concerning pre-study knowledge about and interest in the topic, a questionnaire about ability beliefs, expectancy concerning task execution and perceived usefulness, and importance and interest concerning the topic of power plants was administered. For this, a translated version of the questionnaire used to measure ability beliefs (3 items), expectancies for success (2 items) and values (6 items) in [18], adapted to the topic of power plants, was used. The items could be rated on a scale from 1 (low) to 5 (high).

The learning material consisted of two preparatory texts, one about energy and energy transformation in general and one about the different components of a power plant and their functions. These texts were presented to the participants before the learning task in order to align their pre-knowledge and provide them with

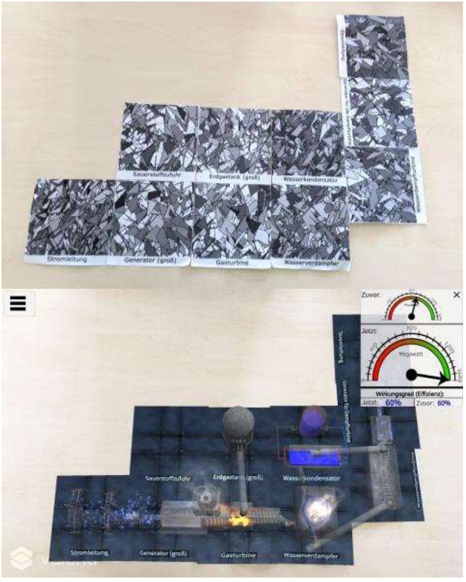


Fig. 1. Photo of a marker cluster and a screenshot from the application scanning that same marker cluster.

some basic information that was necessary for the task execution. After the participants read the texts and before starting with the task, a pre-knowledge test with 8 questions was administered in order to see if the knowledge alignment had worked. In the learning task itself, the participants were provided with seven hypotheses about the electricity output and efficiency of differently composed power plants, which they could test with help of the paper markers and the AR tablet application. In order to test the hypotheses, the participants needed to compare different power plant compositions concerning their energy output and efficiency. Possible variations were different types of power plants (gas, steam or combined) and changes in both the size of the generator and the amount of natural gas input on three levels (small, middle, big). The participants answered a knowledge test with 24 questions at the end of the study. Both the pre-knowledge test and the knowledge test were composed for the present study based on the learning material. The knowledge test consisted of four parts. In part one, there were 7 questions that were not directly connected to the learning task and could be answered completely based on the preparatory texts the participants read before. In part two, 5 questions about the different components of the power plants were asked. The components were all described in the second preparatory text, but knowledge about them may have also been supported through the learning task. In part three, there were 5 questions about efficiency and energy output of the power plants, which could only be answered based on the learning task. Finally, in part four, 7 questions were asked that went further than the texts and the task, so that knowledge transfer was necessary. These can be split into 3 near transfer questions, which still had something to do with the hypotheses tested in the learning task but were not directly answerable based on it, and 4 far transfer questions, which went even further beyond that. As the manipulation between the conditions happened during the learning task, part one of the knowledge test, which did not focus on information provided in the task, was excluded from the analysis concerning knowledge differences between the groups.

To measure the participants' load during task execution, two different questionnaires were used. A cognitive load questionnaire [19] with scales concerning ECL (3 items) and GCL (2 items) based on cognitive load theory was answered on a scale from 1 (low) to 7 (high). Furthermore, the participants answered the NASA TLX [20] in a version translated to German, to measure their task load on six one-item scales (mental demand, physical demand, temporal demand, performance, effort, frustration) between 1 (low) and 21 (high), with an opposite scale for performance (1: high, 21: low).

D. Procedure

At the start of the study, the participants were welcomed and received an explanation of procedure and content of the study. They signed an informed consent form after all their questions were answered. In a questionnaire that was started on a computer screen, the participants subjectively rated their ability beliefs, expectancies for success and subjective task values concerning the topic and task on combined gas-steam power plants.

Following this, the learning phase of the study started. First, the participants read two texts with information about power plants, the first about energy transformation in general and the second about components of power plants and their functions. After that, they answered 8 questions in a pre-knowledge test. Then the learning task started. On the computer screen, the participants were provided with hypotheses about gas-, steam- and combined cycle power plants, which they tested on their correctness. This was done with the help of AR learning material on a table next to the computer. When the participants scanned paper markers representing different components of power plants with a tablet, virtual models of those components appeared upon the markers on the tablet screen. From the individual markers, marker clusters with all necessary components representing the whole power plants could be build and scanned to find out about their electricity output and efficiency (pictures in Fig. 1). The difference between the four conditions lay in both the physical interaction (low vs. high) and the mental interaction (low vs. high) with the learning material. The physical interaction the participants had with the AR learning material was manipulated through the markers. In the conditions with low physical interaction (p–), the participants received AR markers of power plant components already built into complete power plants. In the high physical interaction conditions (p+), those component markers were not built together, so that the participants needed to be more physically active in order to answer the hypotheses in the learning task. In Fig. 2, a visualization of the physical interaction conditions in comparison is shown. The mental interaction the participants had with the learning material was manipulated through the instructions given in the learning task. In the conditions with low mental interaction (m–), the participants received instructions in which they were given information on which kinds of power plants to compare in order to answer the hypotheses. In the high mental interaction conditions (m+), they did not receive that kind of information so that they needed to be more mentally active in order to answer the hypotheses. In Fig. 3, a visualization of the mental interaction conditions in comparison is shown. In Table I, more detailed descriptions of the four conditions can be found.

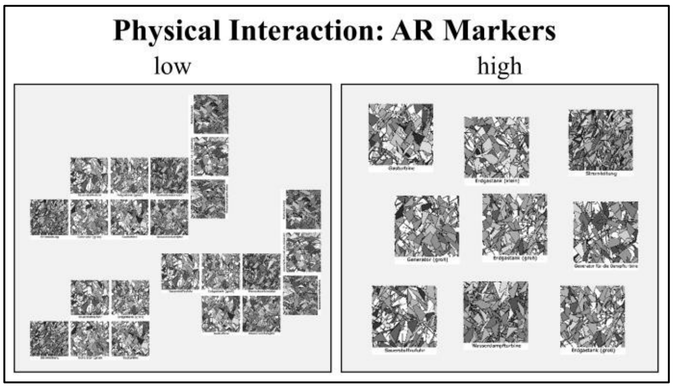


Fig. 2. Comparison of the low and high physical interaction conditions.

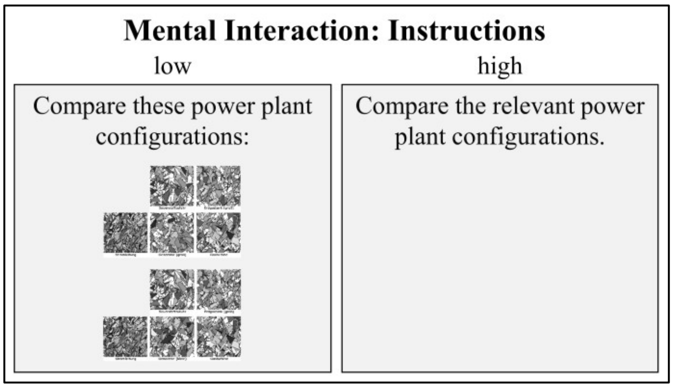


Fig. 3. Comparison of the low and high mental interaction conditions.

TABLE I. DESCRIPTIONS OF THE FOUR CONDITIONS IN THE STUDY

		Physical Interaction	
		p+	*p–*
Mental Interaction	***m+***	The learners are provided with individual markers for the different components of the power plants (e.g., generator, gas turbine). They are only provided with the hypotheses about the power plants they should answer and choose the markers and build the power plants that they need to test it themselves.	The learners are provided with clusters of markers that are already combined into power plants (e.g., gas-fired power plant with a small generator and a small natural gas input). They are only provided with the hypotheses about the power plants they should answer and choose the power plants that they need to test it themselves.
	m–	The learners are provided with individual markers for the different components of the power plants (e.g., generator, gas turbine). They are provided with the hypotheses about the power plants they should answer and descriptions of the power plants (+ pictures of their marker clusters) that they need to test them, so that they only need to build the necessary marker clusters.	The learners are provided with clusters of markers that are already combined into power plants (e.g., gas-fired power plant with a small generator and a small natural gas input). They are provided with the hypotheses about the power plants they should answer and descriptions of the power plants (+ pictures of their marker clusters) that they need to test them, so that they only need to find the necessary marker clusters.

In the last part of the study, the participants answered the remaining questionnaires concerning their cognitive load and task load and answered a knowledge test about the learned topics. They also filled in a questionnaire about their demographic data at the end of the study. After that, they were asked if they had any questions concerning the design of the study and received their compensation.

V. RESULTS

A. *Belief, Expectancy and Value*

To make sure that the groups did not differ concerning their pre-study knowledge about and interest in the topic, the groups' subjective rating of their knowledge on the topic, their expectancy on how well they would solve the tasks and their perceived usefulness, importance and interest concerning knowledge on power plants were compared. A scale of 1 (low) to 5 (high) was administered. Means and standard deviations split by group can be found in Table II. In pairwise comparisons, no significant differences between the groups on their subjective pre-knowledge and their task performance expectancy were found. However, two groups (p–/m– and p+/m–) differed quite substantially in their perceived value of the topic, $t(61.93) = 2.51$, $p = 0.015$, $d = 0.63$.

B. *Pre-Knowledge*

To test if the groups' pre-knowledge basis after reading the preparatory texts was the same, pairwise comparisons of the groups concerning their pre-knowledge test score were executed. The participants could receive a maximum of eight points in the test. Means and standard deviations split by group can be found in Table III. No significant differences were found between the groups, so that it can be assumed that they did not differ in their pre-knowledge after reading the preparatory texts.

C. *Knowledge*

In order to test hypotheses 1a-1c on the influence of the types of interaction on knowledge, a factorial 2x2 ANOVA was administered with physical and mental interaction as factors and knowledge test score as outcome variable. The 17 questions of part two, three and four of the knowledge test were used, because those were the questions for which the task execution was relevant while part one could be answered on the basis of the preparatory texts. The means and standard deviations of the

TABLE II. MEANS AND SDS OF BELIEF, EXPECTANCY AND VALUE

Mean and SD per Group	**Group**			
	p–/m–	*p+/m–*	*p–/m+*	*p+/m+*
Belief	1.17 (0.29)	1.26 (0.40)	1.19 (0.32)	1.22 (0.29)
Expectancy	2.27 (0.58)	2.44 (0.55)	2.44 (0.59)	2.33 (0.69)
Value	1.84 (0.59)	2.21 (0.61)	1.94 (0.50)	2.12 (0.65)

TABLE III. MEANS AND SDS OF PRE-KNOWLEDGE TEST SCORE

Mean (SD) Knowledge Test Score per Group		**Physical Interaction**	
		Low (p–)	***High (p+)***
Mental interaction	***Low (m–)***	5.16 (1.63)	5.28 (1.46)
	High (m+)	5.28 (1.51)	4.88 (1.64)

TABLE IV. MEANS AND SDS OF KNOWLEDGE TEST SCORE

Mean (SD) Knowledge Test Score per Group		Physical Interaction	
		Low (p-)	*High (p+)*
Mental Interaction	***Low (m–)***	7.66 (2.87)	9.34 (3.32)
	High (m+)	8.84 (2.83)	7.242.87)

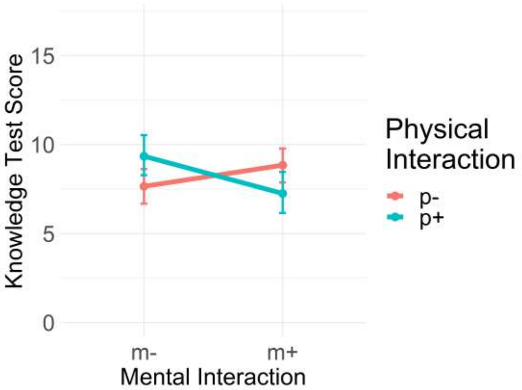


Fig. 4. Interaction effect of mental and physical interaction on knowledge

knowledge test scores can be found in TABLE IV. Group p+/m– has the highest average knowledge test score, followed in this order by group p–/m+, group p–/m– and group p+/m+. As tested by Levene's test, the groups' variances in knowledge test score were found to be homogeneous [$F(3,124) = 0.67$, $p = 0.575$] so that the ANOVA could be used to test the hypothesis. No significant main effect for mental interaction was found [$F(1,124) = 0.67$, $p = 0.416$, $\omega_p^2 = 0.003$], showing that the data did not support hypothesis 1a that mental interaction has an influence on knowledge. Also, no significant main effect for physical interaction was found, [$F(1,124) = 0.01$, $p = 0.933$, $\omega_p^2 = 0.008$], showing that the data did also not support hypothesis 1b that physical interaction has an influence on knowledge. Table IV shows an interaction effect between the two types of interaction, which was found to be statistically significant with a medium effect size, $F(1,124) = 8.72$, $p = 0.004$, $\omega_p^2 = 0.057$. The found interaction does not support hypothesis 1c that the influence of mental interaction on knowledge is bigger than the influence of physical interaction. It shows that knowledge is higher in the groups in which either physical or mental interaction is high than in the groups in which both mental and physical interaction are either high or low.

D. Cognitive Load

To test the second set of hypotheses on the influence of the types of interaction on the two types of cognitive load, different analyses were administered. To test hypothesis 2a-2c that GCL is influenced by both physical and mental interaction, although with a higher influence through mental interaction, a factorial 2x2 ANOVA was administered with physical and mental interaction as factors and GCL as outcome variable. In the data it was found that group p+/m+ has the highest GCL ($M = 4.48$, $SD = 1.64$), followed by group p–/m– ($M = 4.38$, $SD = 1.58$) and group p+/m– ($M = 4.38$, $SD = 1.52$) with the same mean score, and group p–/m+ ($M = 4.16$, $SD = 1.59$) with the lowest score. As tested by Levene's test, the groups' variances in GCL were found to be homogeneous [$F(3,124) = 0.34$, $p = 0.795$], so that the ANOVA could be used to test the hypothesis. No significant main effect for either physical [$F(1,124) = 0.04$, $p = 0.846$, $\omega_p^2 = 0.008$] or mental interaction [$F(1,124) = 0.34$, $p = 0.559$, $\omega_p^2 = 0.003$] was found. Also, no interaction effect was found, $F(1,124) = 0.34$, $p = 0.559$, $\omega_p^2 = 0.008$. Hypotheses 2a-2c could thus not be supported by the data.

To test hypothesis 2d concerning the influence of physical interaction on ECL, an independent samples t-test with physical interaction as grouping variable and ECL as outcome variable was administered. As expected, the p+ groups stated a slightly higher ECL ($M = 3.42$, $SD = 1.32$) than the p– groups ($M = 3.29$, $SD = 1.28$). As tested by Levene's test, the groups' variances in ECL were found to be homogeneous [$F(1,126) = 0.13$, $p = 0.722$] so that the t-test could be used to test the hypothesis. No significant difference between the groups was found, $t(126) = -0.57$, $p = 0.572$, $d = -0.10$. Hypothesis 2d could thus not be supported by the data.

E. Task Load

To explore effects of mental and physical interaction with AR learning material on different kinds of task load, factorial 2x2 ANOVAs with physical and mental interaction as factors and the different kinds of task load measured with the NASA TLX were calculated. The means and standard deviations of the different subscales per group can be found in Table V. For all of the subscales and the total score, the groups p–/m– and p+/m+ have the two highest mean scores, while the p+/m– and p–/m+ groups have the two lowest scores. As tested by Levene's test, the groups' variances were found to be homogeneous for all variables, so that the ANOVAs could be used to test the hypotheses. No significant main or interaction effects were found. Table VI shows the results of the ANOVAs.

TABLE V. MEANS AND SDS OF NASA TLX SCALES

Mean and SD per Group[a]	Group			
	p-/m–	*p+/m–*	*p-/m+*	*p+/m+*
Men. Dem.	11.22 (4.76)	*10.75 (5.65)*	11.06 (5.00)	**12.75 (4.47)**
Phys. Dem.	11.78 (5.79)	11.44 (6.49)	*11.38 (5.66)*	**13.34 (5.45)**
Temp. Dem.	8.28 (4.93)	8.03 (4.68)	*7.94 (5.79)*	**9.53 (4.89)**
Performance	**12.38 (5.72)**	*9.69 (5.87)*	11.38 (6.62)	12.16 (5.94)
Effort	**12.13 (4.48)**	*11.13 (5.42)*	11.38 (4.90)	12.09 (4.60)
Frustration	12.09 (5.43)	*11.19 (5.76)*	11.84 (5.84)	**12.13 (5.60)**
Total	11.31 (3.06)	*10.37 (3.73)*	10.83 (3.69)	**12.00 (2.89)**

[a] Highest mean per subscale in **bold**, lowest mean per subscale in *italic*.

TABLE VI. RESULTS OF ANOVAS CONCERNING NASA TLX

Mean and SD per Group	Ment. Int.		Phys. Int.		Ment. * Phys.	
	F	*p*	*F*	*p*	*F*	*p*
Men. Dem.	1.09	0.298	0.48	0.491	1.49	0.224
Phys. Dem.	0.52	0.470	0.62	0.434	1.25	0.266
Temp. Dem.	0.41	0.524	0.55	0.459	1.04	0.310
Performance	0.47	0.494	0.79	0.374	2.63	0.107
Effort	0.02	0.899	0.03	0.870	1.00	0.320
Frustration	0.12	0.732	0.10	0.755	0.35	0.554
Total	0.93	0.337	0.04	0.847	3.17	0.078

F. Exploratory Correlations

To find out more about the relationships between the different variables that were the focus of this study, correlations between all the variables were calculated and the p-values were corrected with help of the Holm method. A moderate, positive correlation was found between GCL and mental demand [$r(126) = 0.34$, $p < 0.01$] and GCL and effort [$r(126) = 0.36$, $p < 0.01$], which shows that a higher experienced GCL during learning task execution came with a higher experienced mental demand and effort. Furthermore, ECL was correlated moderately and positively with frustration [$r(126) = 0.37$, $p < 0.01$], showing that the higher the experienced ECL during the learning task execution, the higher the experienced frustration. Also, a negative, moderate correlation was found between mental demand and knowledge test score [$r(126) = -0.37$, $p < 0.01$], which shows that when the mental demand during task execution was experienced as higher, the participants scored lower on the knowledge test. These correlations show at least some interesting associations that will be discussed in the light of the other results.

VI. Discussion

The goal of the study presented in this paper was to find out more about the potential influence of different forms of interaction with AR learning material on learning processes and outcomes. For this, participants received AR learning material which required either low or high physical interaction and instructions which required either low or high mental interaction. Unfortunately, none of the hypotheses could be supported. In the first set of hypotheses, the focus was on learning outcomes. It was hypothesized that both high mental and high physical interaction lead to increased knowledge, although with a higher increase through mental interaction. This pattern could not be found in the data. Instead, it was found that knowledge was high when either mental or physical interaction was high (p+/m– and p–/m+), but not when both were high (p+/m+), or both were low (p–/m–). The expectations were met partly, because it was hypothesized that both forms of interaction would increase learning, and it was found that when both types of interaction were low, the resulting knowledge was also lower. However, the second part of our findings, that the resulting knowledge was lower when both types of interaction were high, was not expected. A possible explanation for this pattern is that the combination of mental and physical interaction with AR learning material led to cognitive overload so that learning was hindered. In the exploratory analyses concerning the participants' task load, it was found that participants in the groups p+/m+ and p–/m– had the two highest loads of the four groups on all scales. Those effects were not found to be significant, so that they should be interpreted with care, but they point in the direction that some kind of overload may have hindered learning in these groups. Although it is expected that group p+/m+ scores high in task load, it is not expected that group p–/m– does. This result may be explained by the fact that the group had to handle big marker clusters in their first-time use of an AR learning application while looking back and forth at the desktop screen to see which clusters they needed. This may have led to a feeling of more task load. In general, it may be concluded from the results that some interaction with the AR learning material can support learning while still some support is given to solve the task, so that the focus can be on the content and not just the interaction.

In the second set of hypotheses with the focus on subjective rated cognitive load it was hypothesized that both high mental and high physical interaction with AR learning material would lead to an increased GCL, although with a higher influence by mental interaction. Furthermore, it was hypothesized that high physical interaction would lead to an increase in ECL. These hypotheses could also not be supported by the data. It seems that, although differences in knowledge were found between the conditions, no subjective differences between GCL and ECL were perceived by the participants. This can be due to different reasons. Although the questionnaire used to assess GCL and ECL was validated in the original study [19], the tripartition of cognitive load as it is used here is discussed quite controversial. Leppink, for example, has developed a questionnaire in which there are only two scales (one for ECL and one for ICL), based on the idea that GCL is the load that is induced when dealing with intrinsic cognitive load [21], [22]. The scores on the different subscales of NASA TLX, of which some correlated with ECL and GCL, also showed no significant main or interaction effects. It might thus also be the case that the mental demand of the different conditions may have led to different learning outcomes (which is supported by the data showing a moderate, negative correlation between mental demand and knowledge), although this difference was not detectable on a subjective level. In a future study, cognitive load of the learners could be measured in a more objective way, for example through a secondary task or a physiological measure like eye-tracking related values [22]. Unfortunately, those measures may disrupt the learning task itself and should thus be used with care. Furthermore, the manipulation of the different forms of interaction could aim at providing not just two, but more levels of interaction, to test if it has a systematic influence on cognitive and task load which are then in turn connected to learning outcome.

Some interesting results were found in the exploratory analyses of correlations between the measured variables. Mental demand could be shown to be negatively associated with knowledge, so that higher experienced task load may have led to less knowledge. Also, a positive association between GCL and mental demand and effort could be found, while ECL was associated with frustration.

In the present study, not all potential confounding factors that may have an influence on the learning are controlled for, but the experimental manipulation is fairly systematic, so that differences in the groups can be attributed quite certainly to those differences. A limitation of the study is that it tried to separate mental and physical interaction, although that may not be that easily possible. Although physical interaction can be limited and monitored when watching learners, this is not possible for mental interaction. It was tried to give participants the necessary additional information in the task instructions, so that they would only have to be physically active in copying what they saw on the screen, but especially in the condition with high physical and low mental interaction, mental processes might have been induced. It was also not made sure that the participants really interacted with the learning material. They were not monitored or controlled, so that some participants may have not interacted with the material at all.

In future studies, it is important to transfer the manipulations that were done in this study onto different AR learning material because the types of interaction manipulated in this study were very specific for the learning material (i.e. physical interaction: building power plants vs. not building them; mental interaction: figuring out how to answer the hypotheses vs. receiving this information). Also, different forms of mental and/or physical interaction should be investigated in studies with systematic experimental manipulation of AR learning material. Furthermore, it should be considered to increase the interaction even more, because the interaction in this study was rather low, so that in a future study a condition with higher interaction could be added. To transfer the study's design to a more applied and realistic setting, it should be examined if using different forms of interaction with AR learning material throughout a longer time period shows the same results, as the present study took place only in a short time with one application in a laboratory.

In conclusion it can be said that although the outcomes were not as expected and not all questions could be answered, this study revealed interesting results on how different forms of interaction with AR can lead to different knowledge, cognitive load and task load. The experimental study presented here pursues a systematic approach to vary and examine learning with AR on the basis of an ARcis characteristic in order to learn more about the specific factors that influence how and when AR can be used in an effective way in educational settings. This way, it tries to overcome limitations of other studies in which research on AR applications is executed by comparing an AR setting with a traditional learning setting, making it difficult to deduce why exactly the AR setting may lead to an increased knowledge. This systematic approach is important to identify specific variables for the design of successful applications, although a lot of research is still necessary. This study focused on one of the three ARcis characteristics (interactivity). Combined with systematic research on the other two characteristics (contextuality and spatiality) and the interaction of all three characteristics, insights into AR functions can be gained that help to understand the integrated mechanisms when learners deal with these technologies and thus to develop learner- and context-specific AR tools that can support elaborated learning processes.

References

[1] I. Radu and B. Schneider, "What Can We Learn from Augmented Reality (AR)?," in *Proc. 2019 CHI Conf. Hum. Fact. in Comput. Syst. - CHI '19*, Glasgow, Scotland UK, 2019, pp. 1–12, doi: 10.1145/3290605.3300774.

[2] J. M. Krüger, A. Buchholz, and D. Bodemer, "Augmented reality in education: three unique characteristics from a user's perspective," in *Proc. 27th Int. Conf. on Comput. in Educ.*, Taiwan, 2019, pp. 412–422.

[3] D. Pacheco *et al.*, "A location-based Augmented Reality system for the spatial interaction with historical datasets," in *2015 Digit. Heritage*, Granada, Spain, 2015, pp. 393–396, doi: 10.1109/DigitalHeritage.2015.7413911.

[4] S. S. Jamali, M. F. Shiratuddin, K. W. Wong, and C. L. Oskam, "Utilising Mobile-Augmented Reality for Learning Human Anatomy," *Procd. Soc. Behv. Sci.*, vol. 197, pp. 659–668, 2015, doi: 10.1016/j.sbspro.2015.07.054.

[5] C. A. Holmes, N. S. Newcombe, and T. F. Shipley, "Move to learn: integrating spatial information from multiple viewpoints," *Cognition*, vol. 178, pp. 7–25, 2018, doi: 10.1016/j.cognition.2018.05.003.

[6] S. Domagk, R. N. Schwartz, and J. L. Plass, "Interactivity in multimedia learning: An integrated model," *Comp. Hum. Behav.*, vol. 26, no. 5, pp. 1024–1033, 2010, doi: 10.1016/j.chb.2010.03.003.

[7] R. Moreno and R. E. Mayer, "Interactive Multimodal Learning Environments," *Educ. Psychol. Rev.*, vol. 19, no. 3, pp. 309–326, 2007, doi: 10.1007/s10648-007-9047-2.

[8] R. C. Clark and R. E. Mayer, "Engagement in e-Learning," in *e-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning*, 4th ed., R. C. Clark and R. E. Mayer, Eds. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2016, ch. 11, pp. 219–238.

[9] A. T. Stull and R. E. Mayer, "Learning by doing versus learning by viewing: Three experimental comparisons of learner-generated versus author-provided graphic organizers.," *J Educ Psychol*, vol. 99, no. 4, pp. 808–820, 2007, doi: 10.1037/0022-0663.99.4.808.

[10] J. Sweller, "Cognitive Load Theory: Recent Theoretical Advances," in *Cognitive Load Theory*, J. L. Plass, R. Moreno, and R. Brunken, Eds. Cambridge: Cambridge University Press, 2010, pp. 29–47.

[11] M. Wilson, "Six views of embodied cognition," *Psychon. B. Rev.*, vol. 9, no. 4, pp. 625–636, 2002, doi: 10.3758/BF03196322.

[12] P. Kosmas and P. Zaphiris, "Embodied Cognition And Its Implications In Education: An Overview Of Recent Literature," *Int. J. Educ. Pedagog. Sci.*, vol. 12, no. 7, pp. 2018, doi: 10.5281/ZENODO.1340510.

[13] M. T. H. Chi and R. Wylie, "The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes," *Educ. Psychol.*, vol. 49, no. 4, pp. 219–243, 2014, doi: 10.1080/00461520.2014.965823.

[14] D. Bodemer, R. Ploetzner, K. Bruchmüller, and S. Häcker, "Supporting learning with interactive multimedia through active integration of representations," *Instr. Sci.*, vol. 33, no. 1, pp. 73–95, 2005, doi: 10.1007/s11251-004-7685-z.

[15] A. Protopsaltis, M. Mentzelopoulos, J. Ferguson, and K. Kaloyan, "Quiz Cube: An AR mobile learning application," in *Proc. - 11th Int. Workshop Semant. Soc. Med. Adapt. Pers., SMAP 2016*, pp. 151–155, 2016, doi: 10.1109/SMAP.2016.7753401.

[16] C. Fenu and F. Pittarello, "Svevo Tour: The Design and the Experimentation of an Augmented Reality Application for Engaging Visitors of a Literary Museum," *Int. J. Hum.-Comput. St.*, vol. 114, 2018, pp. 20–35, 2018, doi: 10.1016/j.ijhcs.2018.01.009.

[17] M. Billinghurst and A. Dünser, "Augmented reality in the classroom," *Computer*, vol. 45, no. 7, pp. 56–63, 2012, doi: 10.1109/MC.2012.111.

[18] J. S. Eccles and A. Wigfield, "Motivational Beliefs, Values, and Goals," *Annu. Rev. Psychol.*, vol. 53, no. 1, pp. 109–132, 2002, doi: 10.1146/annurev.psych.53.100901.135153.

[19] M. Klepsch, F. Schmitz, and T. Seufert, "Development and Validation of Two Instruments Measuring Intrinsic, Extraneous, and Germane Cognitive Load," *Front. Psychol.* vol. 8, 2017, doi: 10.3389/fpsyg.2017.01997.

[20] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," in *Adv. Psychol.*, vol. 52, Elsevier, 1988, pp. 139–183.

[21] J. Leppink, T. Gog, F. Paas, and J. Sweller, "Cognitive load theory: researching and planning teaching to maximise learning," in *Researching Medical Education*, J. Cleland and S. J. Durning, Eds. Chichester, UK: John Wiley & Sons, Ltd, 2015, ch. 18, pp. 207–218.

[22] J. Leppink, "Cognitive load theory: Practical implications and an important challenge," *J. Taibah Univ. Med. Sci.*, vol. 12, no. 5, pp. 385–391, 2017, doi: 10.1016/j.jtumed.2017.05.003.

8.3 Paper 3 – Krüger, Palzer & Bodemer, 2022

Krüger, J. M., Palzer, K., & Bodemer, D. (2022). Learning with augmented reality: Impact of dimensionality and spatial abilities. *Computers and Education Open, 3*, Article 100065. https://doi.org/10.1016/j.caeo.2021.100065

[Krüger, J. M., Palzer, K., & Bodemer, D. (2023). Corrigendum to 'Learning with augmented reality: Impact of dimensionality and spatial abilities' [Computers and Education Open, Volume 3 (December 2022), Article 100065]. *Computers and Education Open*, *4*, Article 100127. https://doi.org/10.1016/j.caeo.2023.100127]

Changes of the corrigendum to Figure 1 and Figure 8 directly implemented in the appended article.

Computers and Education Open 3 (2022) 100065

Contents lists available at ScienceDirect

Computers and Education Open

journal homepage: www.sciencedirect.com/journal/computers-and-education-open

Learning with augmented reality: Impact of dimensionality and spatial abilities

Jule M. Krüger *, Kevin Palzer, Daniel Bodemer

University of Duisburg-Essen, Lotharstrasse 65, Duisburg 47057, Germany

ARTICLE INFO

Keywords:
Augmented and virtual reality
Human-computer interface
Media in education
Mobile learning

ABSTRACT

Three-dimensional (3D) representations are often more effective for learning about spatial objects than two-dimensional (2D) representations. In augmented reality (AR), which can include 2D and 3D visualizations, learners' perceptions might differ from other visual media. To examine the specific influence of the dimensionality of AR visualizations on learning the spatial structure of components, 3D and 2D AR representations of the human heart were compared in an experimental laboratory study, otherwise keeping the conditions as comparable as possible. The participants (N = 150) received the respective AR representation and were instructed to look for the hearts' components mentioned in an informational text. As expected, learning with the 3D compared to the 2D representation resulted in higher germane cognitive load and knowledge about spatial relations of components. Proposed effects on extraneous cognitive load, knowledge about spatial positions of components and mediation effects were not found. Higher mental rotation abilities were found to be more beneficial for learning with the 3D visualization, suggesting that these learners were better equipped for this task and supporting an ability-as-enhancer hypothesis. Overall, the study revealed that even in AR scenarios, 3D visualizations may be better to convey knowledge about spatial structures than 2D visualizations. Moreover, the results emphasize the specific moderating role of spatial abilities when learning with 3D AR material. In future research with various spatial learning material and spatial abilities, the generalizability of these results needs to be examined, so that ultimately more insights can be gained for the design of optimal AR learning experiences.

1. Introduction

Augmented reality (AR) is a form of visualizing virtual in combination with physical, real-world information, which has been shown to be beneficial, for example, for learning performance, motivational factors and attitudes in different educational settings (see reviews of AR in education, e.g., [1,6,9,11,18,30,40,39,85,90,105,107]). In the studies described in these reviews, diverse AR applications were used to support teaching different topics with various learning objectives. Although this shows that AR may be useful in diverse learning settings, the research often focused on field studies in which an AR application was compared to a traditional learning setting without observing the specific factors that may lead to its superiority. Critique on this kind of media comparison studies includes that (1) the medium is only a way to deliver the information and instructional method, which are primarily important for learning achievement, (2) the specific personal and media attributes leading to increased learning are unknown, and (3) confounding variables cannot be controlled for [94]. Surry and Ensminger further propose that alternative research should focus on intra-medium studies concerning specific individual attributes of the medium itself and on aptitude-treatment-interaction studies concerning learners' characteristics in interaction with specific medium-related variables. These systematic research approaches can also help examine and disentangle how specific characteristics and variables of AR presentations and experiences influence learning processes and outcomes in interaction with learner characteristics and skills. The insights gained can support teachers and designers in deciding when it is useful to implement AR and how applications should be designed and applied for their specific (learning) goals. In the present paper, we follow the above-mentioned suggestions for alternative research by Surry and Ensminger and execute an intra-medium comparison study including an aptitude-treatment-interaction, which we specify further in the following paragraph.

In AR, virtual elements are combined with real elements by presenting them like they are placed in the real world. The line between virtual and real is blurred so that real-world physical elements and

* Corresponding author.
E-mail addresses: jule.krueger@uni-due.de (J.M. Krüger), kevin.palzer@stud.uni-due.de (K. Palzer), bodemer@uni-due.de (D. Bodemer).

https://doi.org/10.1016/j.caeo.2021.100065
Received 12 June 2021; Received in revised form 24 October 2021; Accepted 7 December 2021
Available online 20 December 2021
2666-5573/

environments can be enriched with virtual elements, and vice versa virtual elements can be linked to real-world elements and environments. This can be achieved through three characteristics of AR systems: they combine real and virtual elements, they allow for interaction in real-time and they place the virtual elements inside the three-dimensional (3D) real world [5]. While these characteristics define the technological features of AR systems, it is also important to view AR as an experience of humans using those systems. For educational settings, Krüger, Buchholz and Bodemer [2] defined three characteristics of how humans experience AR, analogous to Azuma's system-focused characteristics: contextuality, interactivity, and spatiality. (1) Contextuality means that in AR, virtual elements are experienced in the context of a real-world environment, (2) interactivity includes experiences of interacting with both real and virtual elements in AR, and (3) spatiality describes that virtual elements are experienced spatially inside the 3D real world, even when the device used has a two-dimensional (2D) screen. One factor that is a part of spatiality is the dimensionality of the presentation of graphical learning material, which can be in both 2D and 3D in AR. Because AR representations enable the visualization of virtual 2D pictures and 3D models in the context of real-world objects and an intuitive interaction with virtual elements, users' experience of the dimensionality differs from other virtual media (e.g., virtual reality, desktop-based virtual objects), so that general research concerning 3D representations in education may not be completely transferable to AR. Furthermore, systematic empirical research on 3D AR visualizations that explores the influence of cognitive load and spatial abilities is scarce. The present paper provides novel answers to the research questions "How does the dimensionality of the visualization of a 3D object in AR influence cognitive load and learning outcomes, and which role do spatial abilities play in this relationship?". It reports an experimental study concerning learning processes and outcomes when learning about the spatial structure of a 3D object, focusing on an intra-medium comparison isolating dimensionality as a variable that can be varied in AR representations through using either 2D or 3D virtual visualizations, while otherwise keeping the compared AR applications as similar and thus comparable as possible. Additionally, the study examines an aptitude-treatment-interaction effect focusing on the role that mental rotation abilities as a specific form of spatial abilities play in this context. Answering these research questions can provide a foundation for decisions in the design of effective and efficient AR applications with regard to the dimensionality of the visualization for the learning of spatial structures, and decisions concerning the groups of learners that should receive those applications with regard to their spatial abilities.

1.1. Dimensionality of representations in education

It seems obvious that when it comes to the dimensionality of a visual representation, a 3D representation of a 3D object delivers a more correct and complete picture of the object than a 2D representation. Wu and Shah [106] identified that learners may have difficulties with the identification of depth cues in 2D visualizations and the formation of 3D mental images on the basis of 2D structures. Further, it has been proposed that learning with 3D representations supports the development of more accurate mental models than learning with 2D representations [19]. Empirical studies in various domains in which spatial elements are crucial show that using a 3D representation positively influences learning outcomes.

One of the educational domains in which the dimensionality of a representation can play an important role is chemistry, especially when the focus is on learning about molecules (e.g., [24,28,106]). Dori and Barak [29] suggest that both virtual and physical models should be used to support a spatial understanding of molecular structures. Stull and Hegarty [92] found that training translation between different formats of 2D diagrams in organic chemistry using physical or virtual 3D models in comparison to only using 2D diagrams led to higher translation accuracy in subsequent tests with the 3D model available, but also in a delayed test without a model. They explained these positive effects on both immediate and long-term learning with a decreased cognitive load when a 3D model is available to support translation, leading to an internalization of a mental model of the transformation that can be used in future tasks.

Further educational domains in which 3D representations have been compared to 2D representations are astronomy (e.g., [12,17,51]) and geometry (e.g., [35,42,58]). In astronomy, for example, teaching about moon phases and relative positions of the sun and the moon using a desktop-based 3D virtual environment led to better learning results than using only 2D photographs [93]. In a study on education about geometric figures, students who learned with desktop-based 3D virtual models also scored better in subsequent questions in which the visualization was critical, but not in questions in which it was noncritical, in comparison to students in a conventional learning setting [91].

Studies in which virtual or physical 3D models are used to support learning are often executed in the domain of (bio-)medical education, especially with material concerning anatomy where the structures that are taught are inherently spatial. Spatial visualization and thus 3D learning have been identified as very important in the anatomy domain [4], which is fundamental for medical education because it forms a structural basis for diagnosis and therapeutic procedures [73]. In a review by Triepels et al. [98], it was shown that many but not all studies concerning learning of anatomy found advantages for students learning with virtual 3D visualizations compared to more traditional methods. In another meta-review on the usage of different 3D visualization technologies in the domain of anatomy, benefits for learners' performance and cognitive load were found [44]. Some specific measures of performance that benefitted from the 3D visualizations in the studies were the identification and localization of and knowledge about spatial relationships between anatomical structures. This is important information that needs to be stored when learning about a spatial object. As part of a cognitive task analysis, Berney et al. [8], for example, describe the identification and reconstruction of position and location of 3D structures in relation to their surroundings as steps in learning functional anatomy. In order to recall the spatial structure of the object, a comprehensive internal representation including this knowledge about the spatial position of and spatial relations between components must be established, which may be supported by the usage of 3D representations. In order to memorize the general spatial structure, a pictorial mental representation may be enough. Once it comes to communication about the structures and components, the correct terms are also part of the knowledge that learners should have memorized and be able to recall. In one study, Zinchenko et al. [108] found that immersive 3D VR visualizations of a human heart led to more knowledge than paper-based 2D and screen-based 3D visualizations. This study thus highlights that there are not only differences between 2D and 3D but also between different 3D visualizations concerning beneficial effects on learning outcomes. In the present study, we examine if the positive influence of 3D visualizations on learning outcomes when learning about a spatial object can also be found when learning with AR visualizations, which differ from other media in their approach to display virtual 3D models. Specifically, we examine if the knowledge about spatial positions of components within a 3D object and the knowledge about spatial relations between those components are supported through the 3D visualization. Insights from this are important to identify whether a more comprehensive internal representation can be established and whether 3D visualizations should be used for instructional AR material about spatial characteristics of objects.

1.1.1. Augmented reality visualizations

3D representations can be displayed through different forms of visualization and with different technologies. Examples for display variants are monoscopic 3D displays, such as desktop-based 3D, stereoscopic 3D displays, such as 3D glasses, autostereoscopic displays, such as parallax barrier displays, and AR or mixed reality displays [44].

Craig [25] describes that the virtual elements that are shown on AR displays and thus placed in the real-world environment can be presented in one, two, three or more dimensions, although the display showing the visualizations is often two dimensional. AR displays can be further differentiated into stereoscopic and monoscopic displays. Stereoscopic displays, which show two different pictures to the eyes using binocular disparity, include optical see-through head-mounted displays (HMDs). Monoscopic displays, which show only one picture and do not rely on stereoscopic but only monoscopic depth cues, include video see-through handheld devices, such as tablets. Stereoscopic depth cues can lead to a better perception of spatial depth than monoscopic cues alone, especially for nearby objects [25].

Although the 3D models themselves may be the same, viewing them in monoscopic AR still differs from viewing them in usual monoscopic non-AR desktop- or tablet-visualizations. While non-interactive traditional visualizations only include static monoscopic depth cues, AR visualizations can rely on additional motion-based depth cues, which are derived from position changes in relation to the object [25]. Due to the fixation of the virtual AR object to a point in the real world, it is possible for learners to move in relation to the object so that active motion parallax, sometimes also called motion perspective, can arise [68,83]. This way, 3D objects viewed in handheld AR can appear more spatial than in non-AR even without adding stereopsis, although they are still classified as pseudo spatial and not true spatial visualizations [83]. Through the additional motion depth cues in monoscopic AR visualizations shown on handheld devices which are available to more people and not as expensive as HMDs, we believe that handheld AR may be a very effective and valuable alternative to usual 3D visualization for learning about 3D objects. In addition to motion-based depth cues for depth perception, AR enables learners to intuitively move around objects, which differs from a mouse- or touch-based interaction with non-AR displays and enables realistic perspective changing, which is in accordance with the AR-characteristic interactivity [2]. Compared to 3D objects shown in virtual reality (VR), which may also benefit from motion-based depth cues and are often viewed stereoscopically, the main difference in AR is that the virtual objects are perceived in spatial relation to real-world objects, which is in accordance with the AR-characteristic contextuality [2]. Through the placement of virtual AR objects in the real world, learners can thus intuitively change perspectives around 3D objects and spatially relate virtual to physical elements, so that the perception of a virtual 3D object in the best case is very similar to the perception of a real, physical 3D object. Added advantages in comparison to physical objects are that virtual models can be widely shared and have no material costs.

With regard to learning, dimensionality as a characteristic of an (AR) representation may be especially influential when a 3D object is the topic of learning material and its spatial structure including the arrangement of individual components should be learned. Cheng and Tsai [20] described different applications used in studies on AR in education, in which image-based AR is used to support spatial abilities, practical skills, and conceptual understanding. The AR application used by Martín-Gutiérrez et al. [75], for example, displays 3D virtual objects to support the visualization of engineering graphics and improve spatial abilities. In an AR application on the topic of inorganic chemistry, the goal was to help students understand the 3D arrangements of presented structures [80]. Furthermore, two AR applications were concerned with astronomical concepts which focus on spatial relationships between planets and stars [55,88]. Compared to traditional learning media, such as textbooks, videos or even desktop applications, it seems that learning with AR leads to an increased understanding of spatial structures. Radu [85], for example, included a special category of studies on learning about spatial structures and function in his meta-review of papers on AR in education. Studies described in this review were executed with learning applications in different spatial domains, such as geometry, chemistry, mechanics, astronomy, and anatomy. Positive effects on learning outcomes of an AR application in comparison to other displays of learning material were shown, for example, in astronomy education (e.g., [36,66,67]), mechanical education (e.g., [71,100,102]), and anatomy education (e.g., [7,59]). Many studies on AR in education seem to focus on applications that are concerned with spatial learning topics and use 3D representations. Furthermore, many of those applications showed positive effects on learning, which additionally emphasizes the importance of the dimensionality of AR representations and their use for spatial learning topics for educational applications.

Although the value of AR for education especially in spatial domains is apparent, most of the studies compare AR with a traditional and often very different form of visualization. In addition to differences in the spatiality of the visualization, there are also often differences in, for example, the device used, where using a tablet in comparison to a book may also have an influence on learning due to novelty and motivational effects, and the interactive possibilities, where a potentially interactive AR visualization may convey more information than a non-interactive textbook; these differences may also have an influence on learning. Surry and Ensminger [94] describe the potential for confounding variables as one of the critical points of media comparison studies and, as mentioned above, describe intra-medium comparisons as valuable alternative studies, which is the approach we use in the present study. Because dimensionality of representation plays an important role in learning about spatial structures with AR learning material, it needs to be examined more systematically and in empirical research settings. While systematic empirical research on the dimensionality of the visual representation of learning material has already been executed with physical and non-AR virtual 3D models (see Section 1.1), systematic research comparing 2D and 3D representations in handheld video see-through AR is still missing. To see if 3D presentation as an individual, controlled variable supports learning about spatial objects in AR, the present study specifically focuses on this comparison of different representations in AR-based learning experiences.

3D AR differs from other forms of virtual 3D representation through the direct spatial relation to the real world and the possibility to move around the virtual object to view it from different perspectives. This can also have a specific influence on learning processes, which may differ in AR from learning with other visual media. Viewing 3D objects in AR seems to be similar to viewing physical 3D objects and it has been found that learning with physical 3D objects can lead to better spatial understanding especially of more complex spatial structures [82]. Specifically, the possibility to move around an object, which is given for both physical and AR 3D models, may have an influence on learning about 3D objects. Learner control of perspective changes around a 3D object was found to be an important factor for spatial learning [38] and walking around an object actively instead of passive movement of the object was found to support the flexibility of spatial memory [48]. Despite these similarities, there are clear differences between AR-based and physical 3D objects. The most prevalent difference is that an AR object can only be moved with an anchor like an AR marker and cannot be touched directly, so that learning with a 3D AR object may also differ from learning with physical 3D objects. To get a more complete picture on the specific case of learning with 3D AR representations and see if the results of other 3D representations can be transferred, systematic empirical research with a focus on both learning outcomes, such as spatial object knowledge, and learning processes, such as the processing of information in working memory, is necessary. In the present study, we thus do not only focus on knowledge as learning outcomes, but also take a closer look at cognitive load that the specific variable of dimensionality in AR may elicit. This way, potential overload that different representations might evoke can be detected and avoided when designing an AR application.

1.1.2. *Cognitive load*

In addition to findings concerning increased knowledge, the dimensionality of visualizations was also found to have an influence on the usage of learners' working memory resources. As mentioned in

Section 1.1, Stull and Hegarty [92] attributed the improved learning results they found to a decrease in cognitive load, but there are also studies that explicitly measured mental or cognitive load when comparing 3D and 2D visualizations. In a study by Dan and Reiner [26], for example, participants learning the visual motor task of origami paper-folding using a 3D visualization compared to a 2D visualization had a lower mental load measured through electroencephalography (EEG) recordings, lower self-reported mental load and also better learning outcomes. It was also found that using a stereoscopic 3D display option in a training with a surgical robot led to lower mental workload scores than using a 2D display version [56]. In their above-mentioned literature review, Hackett and Proctor [44] also reported studies on 3D visualizations that showed a decrease in cognitive load for the specific case of anatomy learning. Foo et al. [37], for example, found a significantly lower mental demand measured by the NASA Task Load Index (see [45]) for learners locating anatomical structures in 3D representations than for learners using 2D representations.

Concerning learning with AR, results are not conclusive with regard to cognitive load, as studies showing a decrease in cognitive load and studies showing cognitive overload have been found [1]. In a study by Lai et al. [60], for example, the authors found that self-rated mental effort was lower in a group of students learning geography with an AR book compared to students learning with traditional multimedia learning material. Cognitive overload was suggested to arise due to a lot of material and task complexity in AR [20]. Here, very different elements can be combined (e.g., real, virtual, static, dynamic, interactive), so that it is particularly important to consider cognitive load when developing AR-based learning environments. In the design, the characteristics of virtual elements, real elements, and their combination must be taken into account. In a systematic mapping review, 64 studies that measured cognitive load in AR experiences were identified in studies from 2007 to 2019 [13]. Most of the studies were media comparison studies (73%) and most used the NASA Task Load Index by Hart and Staveland [45] to measure cognitive load. Only one study distinguished between intrinsic, extraneous, and germane cognitive load, which may be useful to separate cognitive load that is either detrimental or crucial for learning.

Concerning the influence of instructional design on cognitive processing, it is important to acknowledge that measured cognitive load may not only be a sign of detrimental cognitive processing, but also processing that is crucial for learning. A framework to further specify types of cognitive load based on this notion is cognitive load theory (CLT), which describes cognitive processing of learning material based on multiple assumptions about the human cognitive architecture [95, 97]. CLT assumes that to store knowledge in long-term memory and thus to learn, learners process the information that they receive in working memory. This cognitive processing and subsequent storing of information is essential for learning, makes use of the limited working memory resources of the learners and thus has an influence on their cognitive load. Specifically, the parts of cognitive load that are characterized as essential for learning are referred to as intrinsic cognitive load (ICL) and germane cognitive load (GCL) in CLT. ICL is determined by cognitive processing that depends on the complexity of the content of the learning material and the prior knowledge of the learner, where more complex content without the appropriate level of knowledge leads to higher ICL. GCL is determined by cognitive processing dependent on the design of the learning material that is directly relevant for learning. The third type of cognitive load in CLT which is specified as not essential for learning is extraneous cognitive load (ECL). ECL is determined by cognitive processing dependent on the design of the learning material that is not directly relevant for learning. ECL may even hinder learning when all cognitive resources are exhausted by a high amount of cognitive processing before even reaching the point when the content of the learning material can be processed [95,97].

While in the 1998 conception by Sweller and colleagues the three types of cognitive load are described as independently adding up to total cognitive load, this view has changed over time. Today, GCL is not described as an independent cognitive load component, but as a component that "redistributes working memory resources from extraneous activities to activities relevant to learning by dealing with information intrinsic to the learning task" ([96], p. 264) and is thus closely related to ICL. Independent of this reframing of GCL, it is still assumed that the processes associated with GCL are crucial for learning and from a measuring perspective it is important that all three aspects of cognitive load are understood in a learning situation [57]. From a design perspective it is also important to understand how learning material can be designed to decrease cognitive processing that is not relevant or even detrimental to learning as much as possible, while still supporting and increasing germane cognitive processing that is essential for learning within the capacity limits, which are goals proposed in CLT [101]. This is also supported by other researchers, who additionally state that the reduction of ECL in learning tasks is not sufficient but the focus should also be on fostering GCL when designing learning material [87]. In their attempt to connect CLT and human-computer interaction, Hollender et al. [47] transferred this demand onto learning technologies by concluding that a primary goal of educational software should be to foster GCL.

Based on the presented studies showing that (extraneous) cognitive load can be reduced and germane processing of information into mental models and thus learning can be supported when using 3D visualizations, it can be assumed that the dimensionality of a visualization has an influence on the distribution of cognitive load. Further, when looking at the specific case of 3D representations in AR, it is important to examine cognitive load due to the number of interacting elements and thus potential overload. In the present study, we want to examine if the presumptions that 3D visualizations are beneficial to cognitive load can also be confirmed for learning with AR visualizations. We focus on the distinction between ICL, ECL, and GCL to gain insights into the specific allocation of cognitive resources to relevant and non-relevant tasks when learning with 3D or 2D AR. This way, potential differences in learning outcomes might be explained and designers of AR applications can take this into account. Through the intra-medium comparison, differences in cognitive load can be attributed to the dimensionality of the visualization and confounding variables are limited. Although the literature review suggests that presenting 3D learning material in three dimensions in AR is beneficial, it may be necessary to take a more nuanced look at this, especially in respect to learners' spatial abilities. Cheng and Tsai [20] described spatial abilities as relevant learner characteristics that should be examined in image-based AR learning environments because they might have an influence on learning processes and outcomes.

1.1.3. *Spatial abilities*

Perceptual abilities, including spatial abilities, in general "have to do with individuals' abilities in searching the visual field, apprehending the forms, shapes, and positions of objects as visually perceived, forming mental representations of those forms, shapes, and positions, and manipulating such representations 'mentally'" ([15], p. 304). One important type of spatial abilities is the ability to mentally rotate figures and know what they look like from different perspectives. This is relevant when learning about 3D objects because 3D mental models of those objects must be kept in memory and mentally rotated if they need to be recalled from different viewpoints.

A meta-review by Höffler [46] emphasizes that spatial abilities play a role in learning with pictorial visualizations. A significant difference in effect size when comparing learners with higher and lower spatial abilities was found between studies that used 2D learning materials and those that used 3D learning materials. In studies in which 2D learning materials were used, the effect size was larger than in studies in which 3D learning materials were used, showing that spatial abilities are more influential when the materials are in 2D than when they are in 3D. This suggests that the effectivity of the dimensionality of learning materials is moderated by spatial abilities in accordance with an ability-as-compensator hypothesis [49] showing that for learners with low spatial abilities the support through 3D visualizations is important, while this may not necessarily be the case for learners with high spatial abilities. This can be explained by the additional depth cues in 3D visualizations, which support learners with low spatial abilities in building a correct 3D mental model. Stull and Hegarty [92] found that using a 3D model in a task for translating molecules between different 2D representations predicted learning results better than spatial abilities did. They reason that the direct representation of 3D space eliminated learners' need to imagine it, which might be especially difficult for learners with low spatial abilities. In a study in which both interactivity and stereopsis were manipulated, students with low visuospatial abilities profited more from a freely rotatable virtual, stereoscopic 3D model of the lower abdominal anatomy instead of three 2D pictures of cross-sections than students with high visuospatial abilities [70].

Some evidence has also been found for an ability-as-enhancer hypothesis which implies that especially learners with high spatial abilities benefit from using 3D visualizations because they can mentally handle and process those visualizations more easily than learners with low spatial abilities [49]. An empirical study by Huk that compared how learning outcomes were influenced by the dimensionality of the visualization compared to spatial abilities found a significant interaction effect, showing that learners with high spatial abilities benefitted from 3D models, while learners with low spatial abilities did not. These results are thus in contrast with the findings of the meta-review and the other studies described in the previous paragraph. In a pilot study concerning the relevance of spatial abilities when learning with 3D AR visualizations, Krüger and Bodemer [3] found inconclusive results, showing support for an ability-as-compensator hypothesis with regard to 3D spatial visualization abilities and a learning task and support for an ability-as-enhancer hypothesis with regard to 2D spatial memory abilities and a spatial knowledge test. As research with a focus on spatial abilities and AR 3D visualizations, which, as described above, may differ from other forms of visualization, is still scarce, more research is necessary to come to more conclusive results. In the present study, we shed more light on the interaction between the dimensionality of the visualization and mental rotation abilities as a specific form of spatial abilities when using AR learning materials, thus including an aptitude-treatment-interaction analysis into the intra-medium study. We focus on 3D mental rotation abilities because it is related to building a correct mental model of an object from different perspectives, that is, in 3D. The results can provide a basis concerning which population benefits from a 3D visualization and which may not, so that decisions concerning target populations for applications can be made. Furthermore, insight into this specific aptitude-treatment-interaction can be gained.

1.2. The present study: goal and hypotheses

The goal of the present study is to examine the influence of the dimensionality of a visual representation in AR on learning processes and outcomes, answering the research questions "How does the dimensionality of the visualization of a 3D object in AR influence cognitive load and learning outcomes, and which role do spatial abilities play in this relationship?". We take an intra-medium comparison approach, including an aptitude-treatment-interaction, to gain more specific insight into an effective use of AR in education and its underlying mechanisms. As described in Section 1.1, the literature on dimensionality of representations in education shows that learning outcomes in spatial domains and of spatial content can be supported when using 3D visualizations. The literature on learning about spatial objects in AR further supports the idea that 3D AR visualizations are beneficial here (see Section 1.1.1). This seems to be especially the case for learning about the spatial position of and the spatial relations between components in a spatial object. Knowledge about the spatial position of the components concerns the attributes of individual objects independent of other objects – if the position of one component is unknown, the position of another can still be known. For knowledge about spatial relations, on the other hand, information about the position of the component, the position of the other component, and the relation between those two positions must be stored and recalled. In the present study, we want to differentiate between those two kinds of knowledge, to reveal potential differences. Because both kinds of knowledge are dependent on the spatial representation of the object, we expect that learning with a 3D AR visualization leads to increased learning of both kinds of spatial aspects of the object that is visualized (spatial position of components of the object in H1a and spatial relations between components of the object in H1b). We do not expect this difference for aspects that are not related to the spatial aspects of the visualization but concern the general knowledge about the topic, because the visualization should not have an influence here (H1c). The specific hypotheses concerning these different types of knowledge as learning outcomes are:

H1a. Learning with a 3D AR visualization leads to higher resulting knowledge concerning the spatial position of components of the material than learning with a 2D AR visualization.

H1b. Learning with a 3D AR visualization leads to higher resulting knowledge concerning spatial relations between components of the material than learning with a 2D AR visualization.

H1c. Learning with a 3D AR visualization leads to an equal amount of resulting knowledge concerning general (not specifically spatial) aspects of the material as learning with a 2D AR visualization.

A second set of hypotheses focuses on the influence of the dimensionality of visualization on learners' cognitive processing of the content. The three types of cognitive load as defined in CLT are considered in the present study: 1) ICL - load that is elicited by cognitive processing of the content, 2) ECL - load that is elicited by cognitive processing that depends on the design of the material, but is not necessarily relevant for or can even be detrimental to learning, and 3) GCL - load that is elicited by cognitive processing that depends on the design of the material and is relevant for learning, like the forming of mental models. The literature on cognitive load (see Section 1.1.2) shows that extraneous cognitive processing can be decreased when 3D visualizations are used for learning. We hypothesize that learners using a 3D AR visualization in comparison to a 2D AR visualization need to engage in less extraneous cognitive processing and thus have less ECL because they do not need to first mentally transform the 2D visualization before being able to build a 3D mental model of the object (H2a). Because it was also found in the literature that cognitive processing which is relevant for learning can be supported by 3D visualizations, we also hypothesize that using a 3D AR visualization leads to more germane cognitive processing and thus GCL than using a 2D AR visualization (H2b) because a 3D AR visualization should enable and even encourage learners to create a more complete and correct spatial mental model of the object than a 2D visualization. Because ICL is not influenced by the design but by the content of the material, and we would expect the complexity of the content itself to be the same when only the dimensionality of the visualization changes, we expect ICL to not differ on this basis (H2c). The specific hypotheses concerning the different types of cognitive load are:

H2a. Learning with a 3D AR visualization leads to lower ECL during learning than learning with a 2D AR visualization.

H2b. Learning with a 3D AR visualization leads to higher GCL during learning than learning with a 2D AR visualization.

H2c. Learning with a 3D AR visualization leads to equal ICL during learning as learning with a 2D AR visualization.

Many of the studies in the literature (especially those in Section 1.1) also state that the increased learning outcomes from using 3D instead of 2D visualizations resulted from the decrease of extraneous cognitive processing that we hypothesize in H2a and the increase of germane cognitive processing that we hypothesize in H2b. Based on this, we also formulated two mediation hypotheses:

H2d. The effect of the dimensionality of the visualization on knowledge concerning spatial aspects of the material is mediated by the elicited ECL. ECL is lower when learning with a 3D AR instead of a 2D AR visualization, and knowledge is in turn higher.

H2e. The effect of the dimensionality of the visualization on knowledge concerning spatial aspects of the material is mediated by the elicited GCL. GCL is higher when learning with a 3D AR instead of a 2D AR visualization, and knowledge is in turn also higher.

In a third set of hypotheses, we focus on a moderation effect that mental rotation abilities may have on learning with 3D AR learning material. As the empirical support is higher for the ability-as-compensator hypothesis than the ability-as-enhancer hypothesis concerning spatial abilities (see Section 1.1.3), our hypotheses are formulated in agreement with the former. We would thus expect that learners with lower 3D mental rotation abilities especially benefit from receiving a 3D AR visualization in comparison to a 2D AR visualization, because they are relieved of the task of forming a 3D mental model by themselves due to the already three-dimensional presentation of the object. This way, they need to execute less extraneous cognitive processing and can process the object more easily. It is expected that learners with higher 3D mental rotation abilities have less trouble mentally visualizing a 3D object from a 2D visualization, so that they do not need a 3D AR visualization because their ECL is already kept low. This assumption translates to the mediated moderation hypothesis H3a. To examine if a 3D AR visualization in comparison to a 2D AR visualization also increases germane cognitive processing and thus GCL for especially learners with lower 3D mental rotation abilities, an exploratory research question concerning a second moderated mediation is proposed (RQ3b). The two following hypotheses describe moderations of the path from the dimensionality of the visualization to the cognitive load in the mediations proposed in H2d and H2e (see Fig. 1):

H3a. In the mediation by ECL of the effect of dimensionality on knowledge concerning spatial aspects (H2d), the influence of the dimensionality of visualization on ECL is moderated by mental rotation abilities – learners with lower mental rotation abilities benefit more from the 3D AR visualization in comparison with the 2D AR visualization than learners with higher mental rotation abilities.

RQ3b. In the mediation by GCL of the effect of dimensionality on knowledge concerning spatial aspects (H2e), is the influence of the dimensionality of visualization on GCL moderated by mental rotation abilities? Do learners with lower mental rotation abilities benefit more from the 3D AR visualization in comparison with the 2D AR visualization than learners with higher mental rotation abilities?

2. Method

2.1. Design

In this study, a randomized between-subjects design with two conditions was implemented in an experimental laboratory study. The manipulated, independent variable was the dimensionality of a visual representation of a human heart in AR, which the participants received as part of a learning task. In the 3D condition, the participants received a 3D AR model and in the 2D condition, they received a 2D AR graphic, which are described in more detail below. We kept all other factors, such as device used, interaction, and context, as similar and comparable between the conditions as possible, so that an influence of variables other than the dimensionality of the representation could be ruled out. The main variables that were measured to answer the hypotheses in this study are spatial positions knowledge (H1a, H2d-H2e H3a-RQ3b), spatial relations knowledge (H1b, H2d-H2e H3a-RQ3b), general knowledge (H1c), extraneous cognitive load (H2a, H2d, H3a), germane cognitive load (H2b, H2e, RQ3b), and intrinsic cognitive load (H2c). Also, we measured learners' mental rotation abilities as a potential moderator variable (H3a-RQ3b).

2.2. Participants

In total, the study had N = 150 participants (109 female and 41 male). The age ranged from 17 to 31 years with a mean of M = 21.81 (SD = 2.98). The allocation to the conditions was quasi-randomized with the goal of evenly distributing male and female participants between the groups, due to potential differences in spatial abilities between men and women, especially mental rotation abilities [16,79,81]. n = 75 (54 female, 21 male) participants were placed in the 3D and n = 75 (55 female, 20 male) participants in the 2D condition. All participants indicated their language level to be at least competent (C1-C2), with most (97.3%) indicating "native language". The sample mainly consisted of students (96.0%), the majority of whom were enrolled in the bachelor's degree program Applied Cognitive and Media Science (84.7%). The other participants were either pre-university students or employed. The participants could receive course credit for taking part in the study.

The participants were asked about how often they had used general mobile applications, mobile learning applications, general mobile AR applications, and mobile AR learning applications on tablets or smartphones in the past, and answered in a five-point response format: "never" (1), "rarely" (2), "now and then" (3), "often" (4), "regularly" (5). The participants indicated to have used general mobile applications quite regularly (M = 4.65, SD = 0.86), but learning applications had not been used that often (M = 2.37, SD = 1.11). The participants also did not have a lot of experience with using AR applications (M = 1.77, SD = 0.75) or specifically AR learning applications (M = 1.21, SD = 0.53). The participants were thus in general familiar with using mobile devices, but not with using AR applications on smartphones or tablets. Many participants (37%) reported never having used a general AR application on a mobile device, while 83% indicated never having used an AR learning application on such a device before. The participants in the 3D condition indicated a significantly higher amount of usage of general learning applications on mobile devices (M_{3D} = 2.57, SD_{3D} = 1.16) than participants in the 2D condition (M_{2D} = 2.17, SD_{2D} = 1.03), U = 2274.50, p = .037, d = 0.36. In the other three categories, including usage of AR learning applications which is most relevant for the present study, no differences between the groups were found. This study with the ID psychmeth_2019_AR4_56 was conducted in accordance with ethical guidelines and was approved by the ethics committee of the Computer Science and Applied Cognitive Science department under ethics vote ID 1905PFPK3747.

2.3. Material and apparatus

2.3.1. AR applications

For this study, two AR applications were developed with the Unity[1] software (version 2018.2.11f1, [99]) and the Vuforia Augmented Reality Development Kit version 7.5 from PTC Inc. [84], one including a 2D AR and one a 3D AR representation of a model of the human heart. A virtual 3D object of the human heart was obtained from Remix 3D [77], a free online library for 3D objects. This model was used for the 3D version and for a 2D image of the heart model's cross section. For the purpose of the study, the names of components of the human heart and connecting lines were added as labels to both graphics. The AR marker that was used in both applications showed the 2D image of the cross section, but without these labels. In the tablet applications, scanning the AR marker with the camera brought up the 2D image or 3D model with labels on top of the marker, and a white background was added covering the marker to decrease visual clutter. The virtual representations were fixed to the point of the visual marker on the paper, so that when the participants moved around, the representation stayed in the same spot.

[1] This research is not sponsored by or affiliated with Unity Technologies or its affiliates. "Unity" is a trademark or registered trademark of Unity Technologies or its affiliates in the U.S. and elsewhere.

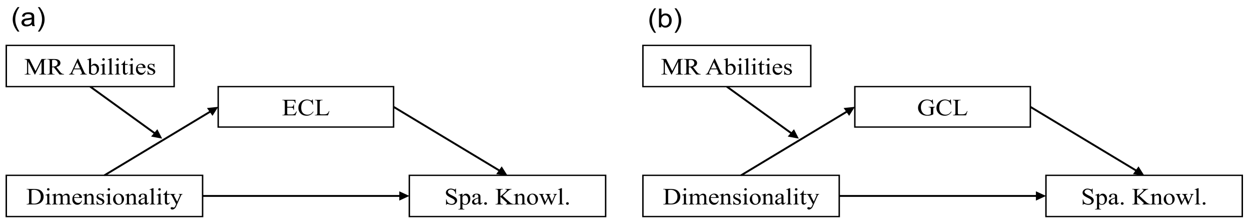


Fig. 1. Moderated mediation in H3a (a) and RQ3b (b). In H3a, the influence of dimensionality on ECL in the mediation model proposed in H2d is hypothesized to be moderated by mental rotation abilities. In RQ3b, the influence of dimensionality on GCL in the mediation model proposed in H2e is hypothesized to be moderated by mental rotation abilities.

This way, participants could either rotate the piece of paper or move around the virtual representation to view it from different perspectives. No additional interaction with the virtual representation was possible and the actual interaction that the participants executed was neither tracked nor observed. The applications were exported as Android packages (APKs), installed and used on a tablet with a 10.8-inch IPS display with a resolution of 2560 × 1600 pixels and about 500g weight. It had 4 GB RAM and a HiSilicon Kirin 960 eight-core processor. The camera had 13 megapixels and a resolution of 1080p at 30 frames per second. Examples of how the application looked during use can be seen in Fig. 2.

2.3.2. *Learning material*

In the present study, an approach is used in which textual and pictorial information is enriched with either a 3D AR model or a 2D AR graphic, following an approach called AR books or sometimes MagicBook [10]. As the literature review shows, anatomy is a domain in which knowledge about spatial arrangement of components of an object, including knowledge about the names of the objects for a common communication ground, is important. This is why we used textual and pictorial material and a model of the human heart as learning material in the present study. The human heart is a spatial structure, so that its components and their spatial relationship are suitable to be displayed in a 3D representation and participants can be tested on their knowledge about the positions and relations of components afterwards. In the study design, there were two experimental conditions that differed with respect to the dimensionality of the presented virtual object in AR. In both conditions, the participants received an informational text, which extended over two paper-based pages and described the human heart and its components. The names of the components were highlighted through italicization. The information was taken from two schoolbooks [41,78] and a reference book [14], which were integrated into one informational text for this study. The text contained two images, the second being the AR marker. To ensure systematic comparability between the two conditions, every part of the learning material was the same except for the additional virtual representation of the human heart. Fig. 2 shows how the labelled 3D model and 2D image are overlaid in the two applications. Although it would not have been necessary to add an AR application for the 2D condition because the picture with the labels could just have been on the printout, we wanted to keep the two conditions very similar, including the use of a tablet to reveal further information. In the 2D condition this additional information was only the textual labels of the components, while in the 3D condition the 3D model and the labels were the additional information. This way, we made sure that the students in the 2D condition also needed to look through the tablet to receive all the necessary information. The participants were instructed to thoroughly read the informational text and at the same time use the tablet for scanning the graphic of the human heart. Further instructions were to look for the structures mentioned in the text in the tablet-application and try to understand their relations. This way, the learning of the spatial arrangement of the components should be encouraged, while also laying a focus on the connection of the textual description and thus the terms for the different components and the spatial arrangement.

2.3.3. *Manipulation check*

To see if the manipulation of the AR learning material worked as intended, a manipulation check was administered. For this, we constructed a questionnaire called the ARcis Questionnaire with the goal of measuring learners' perception of the representation in AR. This questionnaire was developed with three subscales on the basis of the three human-centered characteristics of AR experiences contextuality, interactivity, and spatiality (ARcis characteristics; [2]). The participants rated six statements per subscale in a seven-point response format ranging from 1 (not at all true) to 7 (very true) and a mean score was calculated for each subscale. Examples of statements are "I perceived the

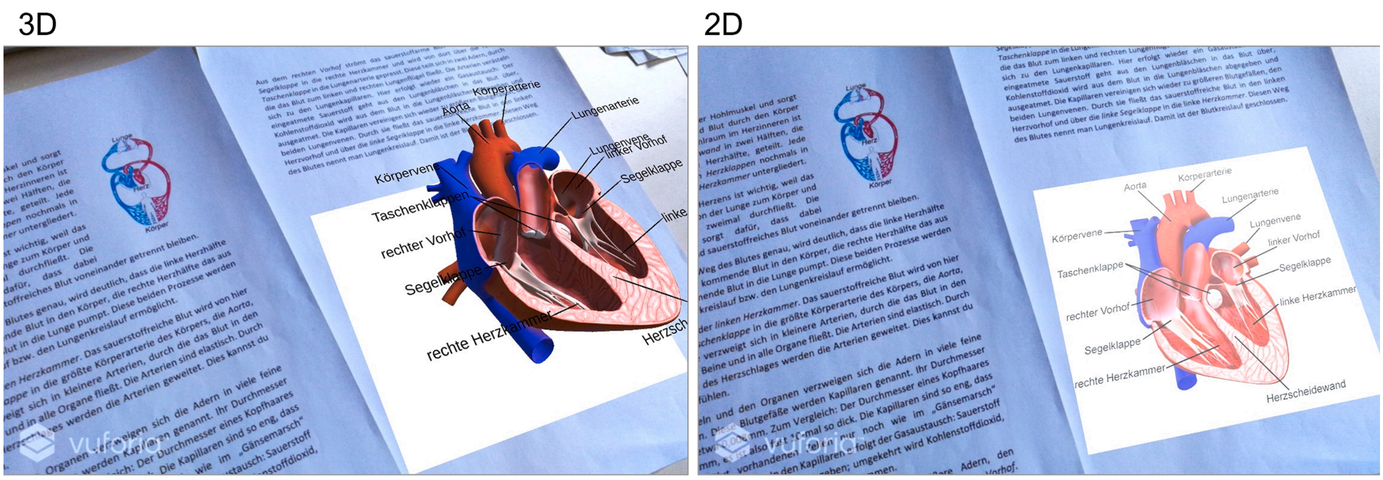


Fig. 2. Screenshot from the application in the 3D condition (left) and the 2D condition (right).

virtual element in the context of the real world" (*contextuality*), "The virtual element was very interactive" (*interactivity*) and "The virtual element has a spatial depth similar to the depth of a real object" (*spatiality*). The virtual element was specified as the model of the human heart in the instructions to the questionnaire. Internal consistency measured through Cronbach's alpha was acceptable for the *contextuality* subscale (.74) and good for the *interactivity* (.81) and *spatiality* (.84) subscales. Because the dimensionality of the representation is defined as part of the AR characteristic spatiality, it was constructed into the questionnaire as part of the *spatiality* subscale. The assumption implicit in the manipulation was thus that the spatiality of the 3D AR representation would be perceived as higher than the spatiality of the 2D AR representation, while interactivity and contextuality would not necessarily be perceived as different.

2.3.4. Tests and questionnaires

Expectancy-value questionnaire. To learn more about the sample and determine potential pre-task differences between the groups concerning knowledge beliefs; competence expectancy; and perceived usefulness, importance and interest regarding the learning material, the Expectancy-Value Questionnaire by Wigfield and Eccles [104] was translated and adapted to the learning topic in the study. It is divided into three subscales and a five-point response format with different wording was used to answer each question (1 was low, 5 was high). The *Knowledge Beliefs* scale was reformulated from the original ability beliefs scale to ask about perceived knowledge and not abilities. It includes three items and was used to assess the participants' self-rated prior knowledge. The *Expectancy* subscale comprises two items that measure expected personal performance during the learning task. The third subscale, *Usefulness, Importance, and Interest*, comprises six items and measures the motivation to acquire knowledge on the topic. For each subscale, a mean score was calculated. Internal consistency of the three scales measured through Cronbach's alpha was acceptable for the scales *Knowledge Beliefs* (.73) and *Expectancy* (.68), with a high value for the *Usefulness, Importance and Interest* scale (.86).

Knowledge test. The resulting knowledge after the execution of the learning task was measured through a knowledge test with three different parts (1 – *spatial: components,* 2 – *spatial: relations*, and 3 - *general*) which was developed based on the learning material. We classified both the *spatial: components* part and the *spatial: relations* part as knowledge concerning spatial aspects of the human heart, but we examined them separately with different forms of tests to test hypotheses H1a and H1b. The *general* part is the focus of the analysis concerning H1c. First, in the *spatial: components* part of the test, the participants identified the positions of components of the human heart. They located components in 2D pictures which were captured from different perspectives of the 3D model used in the application (see Fig. 3) and presented on separate pages of an online questionnaire. In each of the four pictures, four small numbers were pinned to parts of the picture and participants filled in blank fields naming the components. For each correctly named component, the participants received one point, so that a score of 0 to 16 was possible for this part of the test. Points were given for all instances when it was clear that the correct component was meant, even though it was not written completely correctly (e.g., "arota" also gave a point for "aorta"). This way the focus lies more on the placement of the components and less on correctly remembering the spelling of the terms. None of the four pictures was the exact same visualization as the 2D graphic from the learning task, so that some mental transformation of the 3D object and not just recognition of the picture was necessary especially for the two pictures which showed the heart from the back (picture 2 and 4 in Fig. 3). This way, just memorizing the names of the components was not enough and was not the focus, but the spatial positions had to be remembered. Second, in the *spatial: relations* part of the test, the participants received five multiple-choice questions concerning spatial relationships between different components of the heart, such as "Which component separates atrium and heart ventricle?" with the answer possibilities "atrioventricular valve", "cardiac septum", "aorta", and "semilunar valve". The multiple-choice questions had one correct and three incorrect answer options and a point was given for each correct answer, so that a score of 0 to 5 was possible. For this part of the test, the terms needed to be recognized correctly to know which components the question and the answers referred to, and the learners needed to know the spatial relations of the components to correctly answer the questions. Third, in the *general* part of the test, five multiple-choice questions concerning general information on the human heart that was provided through the informational text and did not have a specific link to the provided visualization were answered. An example is "What is the diameter of a capillary?" with the answer possibilities "0.008mm", "0.5mm", "0.07mm", and "1.0mm". Again, the multiple-choice questions had one correct and three incorrect answer options so that a score of 0 to 5 was possible. The learners did not need to have remembered the (spatial) information from the visualization and only questions without relation to the visualization were asked. The kind of visualization should thus not play a role for answering these questions.

Cognitive load questionnaire. Cognitive load was measured with the second version of the naïve rating scale by Klepsch and colleagues [57]. The subscale on *extraneous cognitive load* (ECL; 3 items; used in H2a, H2d and H3a) is aimed at measuring cognitive load that is caused by the design of the learning material and is unproductive for the learning task itself. The subscale *germane cognitive load* (GCL; 3 items; used in H2b, H2e and RQ3b) is aimed at measuring cognitive load that is caused by the learning related cognitive processes of the learners. The subscale

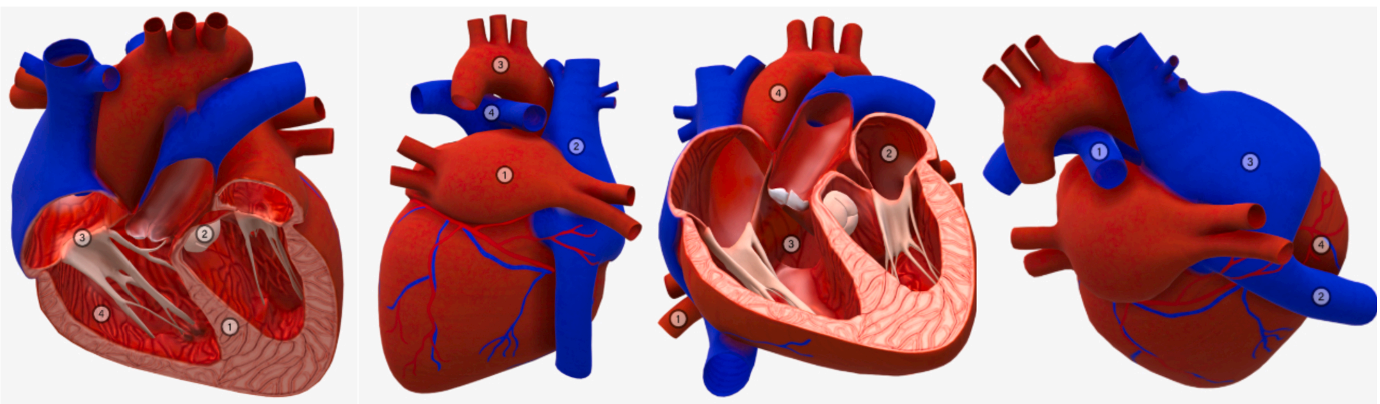


Fig. 3. Pictures used for the component part of the knowledge test.

intrinsic cognitive load (ICL; 2 items; used in H2c) is aimed at measuring cognitive load that is caused by the inherent complexity of the learning material in interaction with the learner's prior knowledge. The questionnaire was used in its original form and participants were told the task they should rate were the activities that were executed with the tablet application. The statements were rated in a seven-point response format, in which 1 was the lowest and 7 the highest agreement with the statement, and mean scores were calculated for each subscale. Internal consistency measured through Cronbach's alpha for the ECL (.73) and GCL (.72) subscales and Spearman-Brown coefficient for the ICL subscale (.69) were acceptable for the scales.

Mental rotation test. The Mental Rotation Test (MRT) by Peters et al. [81] was used to assess the participants' mental rotation abilities. It contained 12 items (half of the original number of items) and was limited to three minutes. The MRT is a test that requires the ability to mentally rotate 3D figures and to assign them to a reference figure. Each item presents a static reference image of a 3D figure and four different static images of the same figure. Two of those images show the figure rotated, the other two show the figure in a rotated and mirrored state. The participants were asked to indicate the two figures that were only rotated but not mirrored, for as many items as possible during the three minutes. Active rotation was not possible with the static pictures, so that rotation had to take place mentally. Since two of the images are rotated for each item, two correct answers can be counted for each item, which are then summed for a score between 0 and 24.

2.4. Procedure

First, the participants were informed about the content, purpose, and procedure of the experiment. They were made aware that all data would be collected and processed anonymously and that they had the possibility to stop the experiment at any time without giving reasons. After the participants had signed the informed consent form, they started the study in the experimental condition that was assigned to them. The experiment started with the mental rotation test, which was limited to three minutes. The participants then answered the Expectancy Value Questionnaire. They were then asked to contact the researcher to receive the learning material consisting of the informational texts on the human heart as well as the tablet with the respective AR application for either the 2D or the 3D condition. The participants read the informational text and looked at the tablet-based 2D or 3D visualization, in which they could find the components of the human heart mentioned in the text. There was no time limit for this. After the participants had studied the learning material, they returned the learning material to the examiner and began to answer the questionnaires on cognitive load and the three characteristics of AR. This was followed by the knowledge tests, where first the spatial: components part, then the general and then the spatial: relations part were administered. At the end of the survey, the participants provided their demographic data, as well as their previous experience with AR and mobile learning applications. Finally, the participants received a debriefing, in which they were informed about the manipulation of spatial representation in the AR application.

3. Results

For all tests described in this section, a significance level of $\alpha = .05$ was applied. When the respective variables were not distributed normally, nonparametric Mann-Whitney U tests [72] were used for testing hypotheses concerning group differences (H1a, H1b, H2a, H2b). When the distribution was normal, t-tests were administered as Welch's t-test [103] by default as suggested by Delacre et al. [27] based on simulations showing that Welch's t-test has a more stable Type I error rate and thus provides better results when the assumption of homogeneity of variance is not met, and most of the time has at least the same power as Student's t-test when the assumption is met. This way, homogeneity of variance does not have to be assumed and thus does not have to be tested, although Levene's tests did support a homogeneity of variance for all variables. Interpretations of the effect size Cohen's d are based on the classifications by Cohen [23]. For correlations between variables, Kendall's τ [54] was used, due to non-normal distributions in most variables. In the mediation (H2d and H2e) and moderation (H3) analyses, a percentile bootstrapping method [32] was used for the calculation of the significance of the effects and standard errors to account for non-normal distribution of the variables: unstandardized effects were computed for each of 10,000 bootstrapped samples, and the 95% confidence interval was computed by determining the effects at the 2.5th and 97.5th percentiles.

For the equivalence hypotheses (H1c, H2c), two one-sided t-test (TOST) equivalence tests were used [64]. The smallest effect size of interest (SESOI; [61]) for the equivalence tests was set at a small effect size of Cohen's $d = +/-0.3$ beforehand, but needs to be corrected to Cohen's $d = +/-0.32$ because this is the smallest detectable effect size with $n = 75$ in each group. We used the tool described by Lakens [63] to calculate this smallest detectable effect size. The equivalence bounds for all equivalence analyses are set based on this. For Cohen's d, 95% confidence intervals are provided in all analyses. For the moderation analyses the MRT test scores are centered for an easier interpretation of the effect estimates. When the 3D condition is compared to the 2D condition, positive effect size values mean that the 3D condition has a higher average score than the 2D condition for the respective variable, and negative effect size values mean that the 2D condition has a higher average score. In the mediation analyses, a dummy coding of the predictor variable dimensionality of visualization was administered with the 2D condition as 0 and the 3D condition as 1. Here again a positive estimated value of the relation means that the score is higher for the 3D than the 2D condition and vice versa for a negative value.

3.1. Sample characteristics

3.1.1. Belief, expectancy and value

To describe the sample and the groups in more detail, self-reported knowledge beliefs, task expectancy, and value were collected before the start of the learning task in a response format from 1 (low) to 5 (high). Equivalence tests for the three variables with equivalence bounds at Cohen's $d = +/-0.32$ detected no equivalence for the groups concerning task expectancy ($M_{3D} = 2.85$, $SD_{3D} = 0.72$; $M_{2D} = 2.89$, $SD_{2D} = 0.67$), lower bound, $t(147.36) = 161$, $p = .055$, upper bound, $t(147.36) = -2.31$, $p = .011$, but also no significant difference, $U = 2722.00$, $p = .729$, $d = -0.06$, 95% CI [−0.38, 0.26]. The perceived value ($M_{3D} = 3.12$, $SD_{3D} = 0.79$; $M_{2D} = 3.32$, $SD_{2D} = 0.65$) was also not equivalent in the groups, lower bound, $t(142.64) = 0.20$, $p = .422$, upper bound, $t(142.64) = -3.72$, $p < .001$, and it was also not significantly different, $t(142.64) = -1.76$, $p = .080$, $d = -0.29$, 95% CI [−0.61, 0.04]. The groups did, however, differ significantly in their knowledge beliefs and thus their self-reported pre-knowledge on the topic. The group which would receive the 2D visualization reported a higher pre-knowledge ($M_{2D} = 2.07$, $SD_{2D} = 0.62$) than the group which would receive the 3D visualization ($M_{3D} = 1.84$, $SD_{3D} = 0.63$), $U = -2142.50$, $p = .023$, $d = -0.38$, 95% CI [−0.70, −0.05]. This difference is opposite to the expected difference after interaction with the learning material. To see how these variables correlate with and may have had an influence on the results in the knowledge test parts, a correlation analysis was executed per group. In Table 1, Kendall's τ correlation coefficients of the scores on the three subscales and the scores on the three types of knowledge are shown. Only the scores of the belief and expectancy subscales correlated significantly with spatial components knowledge in the 3D condition. These results should be considered in the interpretation of the overall results.

Table 1
Kendall's τ correlations of belief, expectancy and value subscales with knowledge test results split by group.

		Belief	Expectancy	Value
Spatial components knowledge	3D (n = 75)	.28*	.28*	.22
	2D (n = 75)	.10	.13	.08
	All (N = 150)	.17*	.20*	.14
Spatial relations knowledge	3D (n = 75)	.07	.01	.10
	2D (n = 75)	.02	.10	−.05
	All (N = 150)	.01	.05	−.01
General knowledge	3D (n = 75)	.07	.04	−.01
	2D (n = 75)	.05	.21	.08
	All (N = 150)	.04	.11	.02

Note. * $p < .05$.

3.1.2. Mental rotation abilities

To ensure that the participants did not differ between conditions in their mental rotation abilities, pre-learning task mental rotation test (MRT) scores were compared. These MRT scores were also used to test H3a and RQ3b. Although an equivalence test with equivalence bounds at Cohen's $d = +/-0.32$ detected no equivalence for the groups concerning the MRT score ($M_{3D} = 12.11$, $SD_{3D} = 5.89$; $M_{2D} = 12.61$, $SD_{2D} = 5.09$), lower bound, $t(144.98) = 1.40$, $p = .082$, upper bound, $t(144.98) = -2.52$, $p = .006$, also no significant difference was detected, $U = 2600.00$, $p = .425$, $d = -0.09$, 95% CI [−0.41, 0.23]. The groups did thus not differ in their mental rotation abilities, although they were also not equivalent in the determined bounds.

3.2. Manipulation check

To check if the manipulation of the dimensionality of the visualization did indeed influence the participants' perception of the application, the ARcis Questionnaire was administered. With dimensionality as part of the AR characteristic spatiality, we expected that spatiality would be perceived as higher for the 3D AR representation than the 2D AR representation, while this would not be the case for contextuality and interactivity. In a one-sided Mann-Whitney U test, we found that spatiality was indeed perceived as higher in the 3D condition ($M_{3D} = 5.02$, $SD_{3D} = 1.17$) than the 2D condition ($M_{2D} = 3.38$, $SD_{2D} = 1.15$), $U = 871.50$, $p < .001$, $d = 1.41$, 95% CI [1.01, 1.80]. The effect size for this difference is very large, meaning that the manipulation of the dimensionality had the expected influence on the perceived spatiality. Significant differences were also found between the groups in two-sided Welch's t-tests concerning the perceived contextuality ($M_{3D} = 4.44$, $SD_{3D} = 1.14$; $M_{2D} = 3.95$, $SD_{2D} = 1.15$), $t(148) = 2.59$, $p = .010$, $d = 0.42$, 95% CI [0.10, 0.75], and the perceived interactivity ($M_{3D} = 3.89$, $SD_{3D} = 1.21$; $M_{2D} = 3.32$, $SD_{2D} = 1.21$), $t(148) = 2.90$, $p = .004$, $d = 0.47$, 95% CI [0.14, 0.80]. The participants in the 3D condition perceived both variables as higher than the participants in the 2D condition. These results were not expected, although the effect sizes were much lower than for the perceived spatiality. The manipulation of the dimensionality did thus not only have an influence on the perceived spatiality, but also the perceived interactivity and contextuality, although the influence on perceived spatiality was highest. This needs to be taken into account when interpreting the results.

3.3. Knowledge

Spatial Components Knowledge. H1a, in which it was proposed that learners using the 3D AR model of the human heart would have more resulting knowledge on the position of the components of the human heart than learners using the 2D AR graphic, was tested using the results from the spatial: components part of the knowledge test. Participants could receive between 0 and 16 points on that part of the test, and participants in the 3D condition ($M_{3D} = 4.56$, $SD_{3D} = 2.17$) had an average score that was descriptively more than 0.5 points higher than for participants in the 2D condition ($M_{2D} = 4.00$, $SD_{2D} = 2.45$). These results can also be seen in Fig. 4(a). In a Shapiro-Wilk test we did not find a normal distribution of the spatial components knowledge variable in either group, 3D ($W = 0.97$, $p = .044$) or 2D ($W = 0.95$, $p = .007$). We thus subsequently used a one-sided Mann-Whitney U test with the dimensionality of the visualization as a grouping and the spatial components knowledge test score as an outcome variable, which showed no significant difference between the groups, $U = 2427.50$, $p = .073$, $d = 0.24$, 95% CI [−0.08, 0.56]. Although descriptively the results were as expected, H1a was not supported: viewing the 3D visualization of the human heart did not lead to a significantly higher knowledge of positions of the heart's components than viewing the 2D visualization.

Spatial Relations Knowledge. To test H1b, in which it was proposed that learners using the 3D AR model of the human heart have more resulting knowledge concerning the spatial relations between components of the human heart than learners using the 2D AR graphic, the results from the spatial: relations part of the knowledge test were consulted. Participants could receive between 0 and 5 points on that part of the test, and participants in the 3D condition ($M_{3D} = 2.69$, $SD_{3D} = 1.10$) had an average score that was descriptively about 0.45 points higher than for participants in the 2D condition ($M_{2D} = 2.25$, $SD_{2D} = 1.21$). These results can also be seen in Fig. 4(b). In a Shapiro-Wilk test we did not find a normal distribution of the spatial relations knowledge variable in either group, 3D ($W = 0.91$, $p < .001$) or 2D ($W = 0.93$, $p < .001$). We thus subsequently used a one-sided Mann-Whitney U test with the dimensionality of the visualization as a grouping and the spatial relations knowledge test score as an outcome variable. The difference between the groups was significant with a small effect size, $U = 2233.00$, $p = .012$, $d = 0.38$, 95% CI [0.05, 0.70]. H1b was thus supported: viewing the 3D visualization of the human heart led to a higher knowledge of the spatial relations between the heart's components than viewing the 2D visualization. Because the difference in H1a was not significant, the subsequent analyses concerning spatial knowledge (H2e and H2d, H3a and RQ3b) will only be executed for the spatial relations knowledge, for which a significant difference was found in H1b.

General Knowledge. To test H1c, in which it was proposed that the two groups with different visualizations would be equal concerning the resulting knowledge on the general part of the test, we executed an equivalence test. Participants could receive between 0 and 5 points on the general part of the test, and participants in the 3D condition ($M_{3D} = 3.00$, $SD_{3D} = 0.99$) descriptively had a slightly higher average score than participants in the 2D condition ($M_{2D} = 2.77$, $SD_{2D} = 1.10$). These results can also be seen in Fig. 4(c). In a Shapiro-Wilk test we did not find a normal distribution of the general knowledge variable in either group, 3D ($W = 0.90$, $p < .001$) or 2D ($W = 0.90$, $p < .001$). Due to the large sample size ($n = 75$ per condition), our sample is likely quite robust to violations of the assumption of normality [69], so we used the two one-sided t-test (TOST) to test for equivalence of the groups. Equivalence bounds at Cohen's $d = +/-0.32$ translated to raw bounds at $+/-0.33$ and thus approximately a difference of one third of a point in the raw scores of the general knowledge test in the two groups. The hypothesis that general knowledge was equivalent in the two conditions was not supported, 90% CI for d [−0.06, 0.49], lower bound, $t(146.34) = 3.29$, $p < .001$, upper bound, $t(146.34) = -0.63$, $p = .265$ (see also Fig. 5). Descriptively, and as seen in Fig. 4(c), the 3D group had a slightly higher knowledge than the 2D group. However, an additional Mann-Whitney U test did not show a significant difference between the groups, $U = 2410.00$, $p = .116$, $d = 0.22$, 95% CI [−0.11, 0.54]. Overall, H1c was thus not supported: viewing the 2D graphic of the human heart did not lead to the same general knowledge as viewing the 3D model within the assumed bounds, although the scores also did not differ significantly. This shows that we do not have enough data to conclude that no effect is present, but also not enough data to conclude that an effect is present [62].

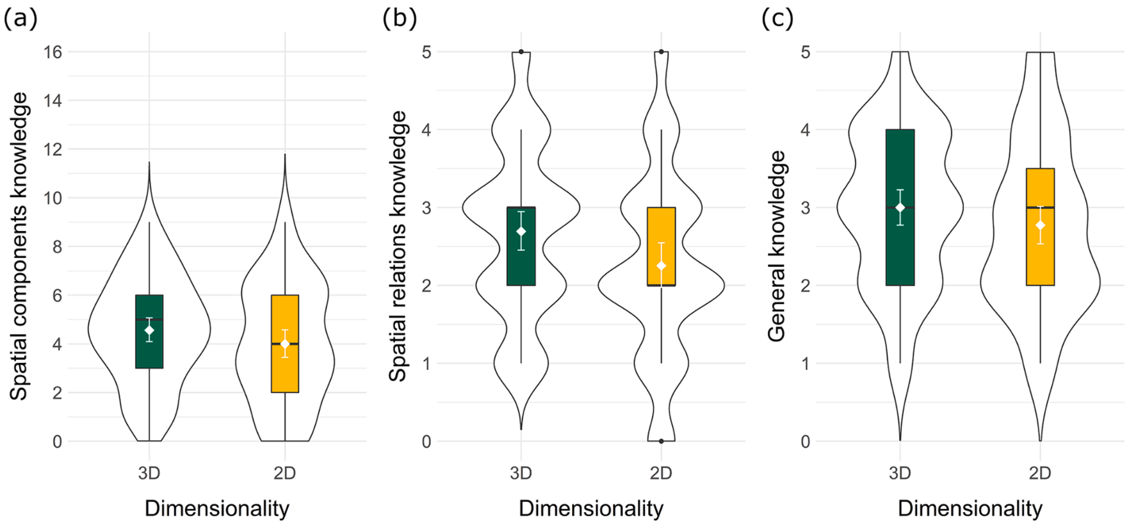


Fig. 4. Distribution of (a) spatial components knowledge, (b) spatial relations knowledge and (c) general knowledge test scores split by group [boxplot with IQR (filled), mean with bootstrapped 95% CI (white), violin plot for distribution (outline)].

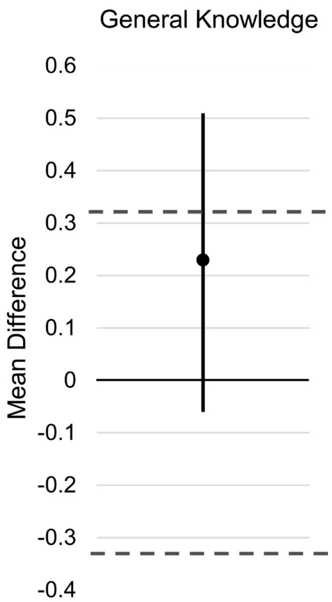


Fig. 5. Result TOST equivalence test for general knowledge showing the 90% CI for raw mean difference [-0.06, 0.51] and raw bounds at +/-0.33. The 90% CI ends above the upper bound, showing no equivalence inside these bounds.

3.4. Cognitive load

Extraneous cognitive load. To test H2a, in which it was proposed that learners using the 3D AR model of the human heart have a lower ECL during the task than learners using the 2D AR graphic, the results from the ECL subscale of the cognitive load questionnaire by Klepsch et al. [57] were consulted. The scores could range from 1 to 7, and participants in the 3D condition ($M_{3D} = 2.64$, $SD_{3D} = 1.08$) had an average score that descriptively was indeed lower than for participants in the 2D condition ($M_{2D} = 2.90$, $SD_{2D} = 1.27$). These results can also be seen in Fig. 7(a). In a Shapiro-Wilk test we did not find a normal distribution of the ECL variable in either group, 3D ($W = 0.95$, $p = .005$) or 2D ($W = 0.95$, $p = .003$). We thus subsequently used a one-sided Mann-Whitney U test with the dimensionality of the visualization as a grouping and the ECL subscale test score as an outcome variable, which showed no significant difference between the groups, $U = 2517.00$, $p = .133$, $d = -0.21$, 95% CI [-0.54, 0.11]. H2a was thus not supported: viewing the 3D visualization of the human heart did not lead to a significantly lower ECL than viewing the 2D visualization.

Germane cognitive load. H2b, in which it was proposed that learners using the 3D AR model of the human heart have a higher GCL during the task than learners using the 2D AR graphic, was tested using results from the GCL subscale of the cognitive load questionnaire. The scores could range from 1 to 7, and participants in the 3D condition ($M_{3D} = 5.40$, $SD_{3D} = 1.10$) indeed descriptively had a higher average score than participants in the 2D condition ($M_{2D} = 5.10$, $SD_{2D} = 1.10$). These results can also be seen in Fig. 7(b). In a Shapiro-Wilk test we did not find a normal distribution of the GCL variable in either group, 3D ($W = 0.95$, $p = .003$) or 2D ($W = 0.94$, $p = .001$). We thus subsequently used a one-sided Mann-Whitney U test with the dimensionality of the visualization as a grouping and the GCL subscale score as an outcome variable. The difference between the groups was significant with a small effect size, $U = 2363.00$, $p = .045$, $d = 0.28$, 95% CI [-0.05, 0.60]. H2b was thus supported: viewing the 3D visualization of the human heart led to a significantly higher GCL than viewing the 2D visualization.

Intrinsic cognitive load. In H2c it was proposed that learners using the 3D AR model of the human heart would be equal in ICL during the task as learners using the 2D AR graphic. This was tested using results from the ICL subscale of the cognitive load questionnaire. The scores could range from 1 to 7, and participants in the 3D condition ($M_{3D} = 3.63$, $SD_{3D} = 1.31$) indeed descriptively had nearly the same average score as participants in the 2D condition ($M_{2D} = 3.68$, $SD_{2D} = 1.22$) with a mean difference of only 0.05 points. These results can also be seen in Fig. 7(c). In a Shapiro-Wilk test we found a normal distribution of the ICL variable in both groups, 3D ($W = 0.97$, $p = .073$) and 2D ($W = 0.97$, $p = .071$). To test for equivalence of the groups, again a two one-sided t-tests (TOST) equivalence test was executed. Equivalence bounds at Cohen's $d = +/-0.32$ translated to raw bounds at $+/-0.41$ and thus a difference of a bit more than one third of a point in the raw scores of the ICL subscale in the two groups. The hypothesis that ICL was equivalent in the two conditions was supported, 90% CI for d [-0.40, 0.29], lower bound, $t(147.31) = 1.70$, $p = .045$, upper bound, $t(147.31) = -2.22$, $p = .014$ (see also Fig. 6). A one-sided t-test also showed no significant difference between the groups, $t(147.31) = -0.26$, $p = .797$, $d = -0.04$, 95% CI [-0.36, 0.28]. H2c was thus supported: viewing the 3D model of the human heart led to a similar ICL as viewing the 2D graphic within the assumed bounds.

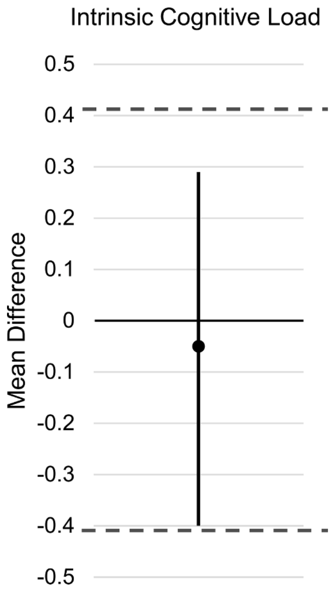


Fig. 6. Result TOST equivalence test for ICL showing the 90% CI for raw mean difference [-0.40, 0.29] and raw bounds at +/-0.41. The 90% CI ends below the upper bound and above the lower bound, showing equivalence inside these bounds.

3.4.1. Mediation of the relationship between dimensionality and spatial relations knowledge

Mediation analyses were used to test H2d and H2e, in which mediation of the influence of the dimensionality on the spatial knowledge was proposed. Specifically, H2d suggested that the 3D visualization would lead to lower ECL, which would in turn lead to a higher spatial knowledge. For the analysis, the dimensionality of the visualization was used as a predictor, while the score on the ECL subscale was used as a mediator and the score on the spatial relations knowledge test part was used as an outcome variable in the model. A summary of the results for the model including completely standardized effect sizes (β) are shown in Fig. 8(a). While path b (ECL on spatial relations knowledge) showed a significant regression coefficient, $b = -0.24$, 95% CI [−0.40, −0.08], $SE = 0.08$, $\beta = -.24$, $z = -2.89$, $p = .004$, for path a (dimensionality on ECL) no significant regression was found, $b = -0.25$, 95% CI [−0.64, 0.11], $SE = 0.19$, $\beta = -.11$, $z = -1.31$, $p = .189$. The indirect effect of the dimensionality of the visualization over ECL on spatial relations knowledge was also not significant, $b = 0.06$, 95% CI [−0.03, 0.18], $SE = 0.05$, $\beta = .03$, $z = 1.15$, $p = .250$. Although descriptively the data point into the right direction with a negative effect of the 3D visualization on ECL and in turn a negative effect on spatial relations knowledge, the mediation of the relationship by ECL and thus H2d is not supported.

Additionally, the other mediation hypothesis H2e suggested that the 3D visualization would lead to higher GCL, which would in turn lead to higher spatial knowledge. For the analysis, the dimensionality of the visualization was again used as a predictor, while the score on the GCL subscale was used as a mediator and the score on the spatial relations knowledge test part was used as an outcome variable in the model. A summary of the results for the model including completely standardized effect sizes (β) are shown in Fig. 8(b). While path b (GCL on spatial relations knowledge) showed a significant regression coefficient, $b = 0.40$, 95% CI [0.25, 0.55], $SE = 0.08$, $\beta = .38$, $z = 5.26$, $p < .001$, for path a (dimensionality on GCL) no significant regression was found, $b = 0.31$, 95% CI [−0.05, 0.66], $SE = 0.18$, $\beta = .14$, $z = 1.71$, $p = .087$. This differs from the results concerning the difference between the groups in GCL (H2b), because that hypothesis was tested with a one-sided test, while

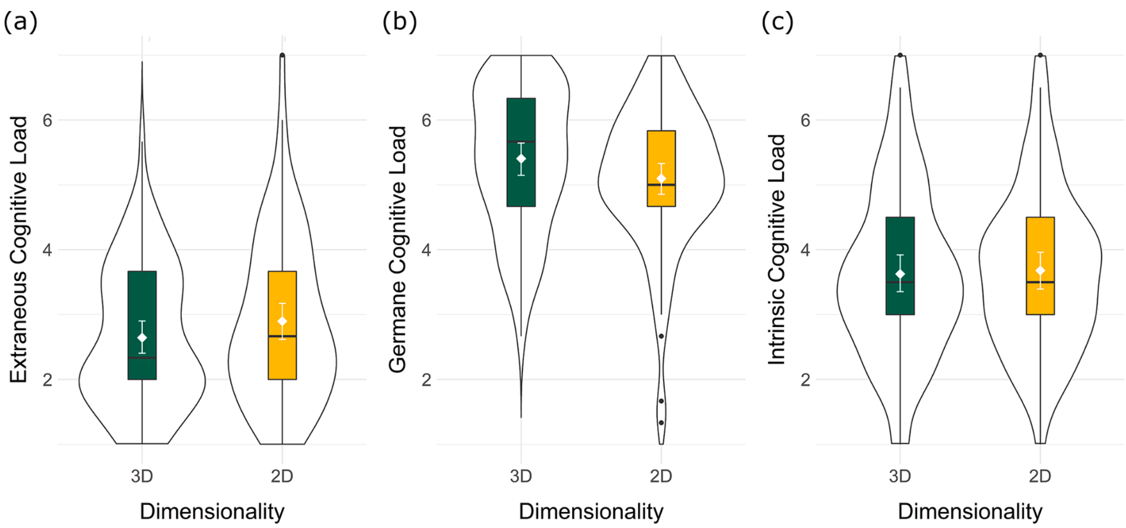


Fig. 7. Distribution of (a) ECL, (b) GCL and (c) ICL subscale scores split by group [boxplot with IQR (filled), mean with bootstrapped 95% CI (white), violin plot for distribution (outline)].

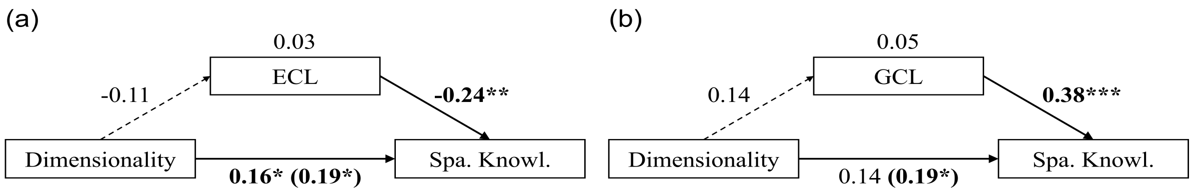


Fig. 8. Meditation model for (a) H2d and (b) H2e including completely standardized effect sizes (β) and significance levels for all effects, including the indirect effect. Significance levels: * $p < .05$, ** $p < .01$, *** $p < .001$.

this is not the case for the regression in this mediation model. The indirect effect of the dimensionality of the visualization over GCL on spatial relations knowledge was also not significant, $b = 0.12$, 95% CI [−0.02, 0.29], $SE = 0.08$, $\beta = .05$, $z = 1.58$, $p = .114$. Although descriptively the data point into the right direction with a positive effect of the 3D visualization on GCL and in turn a positive effect on spatial relations knowledge, the mediation of the relationship by GCL and thus H2e is not supported.

3.5. Mental rotation abilities as moderator

To measure mental rotation abilities, the score on the mental rotation test (MRT) by Peters et al. [81] was used. Scores could range from 0 to 24 and the mean score for the whole sample in the study was 12.36 with a standard deviation of 5.49. In H3a and RQ3b it was proposed that the path from the predictor (the dimensionality of the visualization) to the mediator (the two different cognitive load types) specified in the mediation models in H2d and H2e would be moderated by mental rotation abilities. Specifically, it was suggested that learners with lower mental rotation abilities would profit more from the 3D visualization, leading to a bigger decrease in ECL and a bigger increase in GCL and in turn to higher spatial knowledge.

To first explore how mental rotation abilities may generally moderate the influence of the dimensionality of the visualization on the spatial relations knowledge found in H1b without including the mediation, a moderation model was applied to the data with dimensionality of the visualization as the predictor, spatial relations knowledge score as the outcome, and MRT score as the moderator variable. The MRT score was centered around the grand mean so that the results can be interpreted more easily. In Fig. 9(a), the interaction effect of the dimensionality and the MRT score on spatial relations knowledge can be seen, which is significant, $F(1, 146) = 4.64$, $p = .033$, $\omega^2 = 0.02$. In a subsequent simple slope analysis, this interaction was further specified. In Fig. 9(b) it can be seen that participants with a higher MRT score (mean + 1SD) profit most from 3D in comparison to 2D ($b = 0.85$, 95% CI [0.32, 1.37], $SE = 0.27$, $t(146) = 3.19$, $p = .002$) with a mean increase of 0.85 points in the knowledge test, while participants with a lower MRT score (mean – 1SD) do not profit from it at all ($b = 0.03$, 95% CI [−0.50, 0.56], $SE = 0.27$, $t(146) = 0.11$, $p = .912$). Participants with an average MRT score also profit from 3D in comparison to 2D ($b = 0.44$, 95% CI [0.07, 0.81], $SE = 0.19$, $t(146) = 2.34$, $p = .020$), although not as much as the more skilled participants with a mean increase of 0.44 points in the knowledge test. This shows that learners with higher mental rotation abilities benefitted more from the 3D visualization in comparison to the 2D visualization. When looking at H3a and RQ3b, this relationship displays the opposite of the effect that was expected, which was that learners with lower mental rotation abilities would benefit more from the 3D visualization.

As seen in the mediation analyses (H2d and H2e), the estimated effects of the a-paths in the models (influence of dimensionality on ECL and GCL) did not differ significantly from zero. Adding mental rotation abilities as a moderator to that path, as suggested in H3a and RQ3b, may further clarify the relationship. To test H3a, a moderated mediation model was specified based on the mediation model in H2d with a conditional indirect effect as a function of mental rotation abilities which was added as a moderator to the effect of dimensionality on ECL. The interaction of dimensionality and MRT on ECL (moderation of path a) can be seen in Fig. 10(a). Here it can be seen that in the 3D condition ECL decreases when mental rotation abilities increase, while in the 2D condition it stays around the same level. This interaction is not significant, $F(1, 146) = 1.31$, $p = .254$, $\omega^2 < 0.01$. Although the effect is not significant, we further descriptively explore the mediation models on the different levels. Due to the nature of the moderated mediation model, the mediation models for each level of MRT (Mean - 1SD, Mean, Mean + 1SD) only differ on the moderated path a and thus the indirect effect from the already established model in H2d. In Table 2 (a), the standardized effects for the three levels on path a and the indirect effect and their p-values are shown. Although the moderation is not significant, a direction of change from lower to higher MRT score can be seen in the descriptive values. The regression coefficient of path a is negative for all three MRT score levels, showing that participants in the 3D condition scores lower than those in the 2D condition on the ECL subscale on all levels. This difference grows with a higher MRT score, showing that the score on the ECL subscale when using the 3D visualization in comparison to the 2D visualization decreases even more when the MRT score is higher. The indirect effect and thus the mediation over ECL, in contrast, is positive for all levels and increases with higher mental rotation abilities. Descriptively, students with higher mental rotation abilities thus

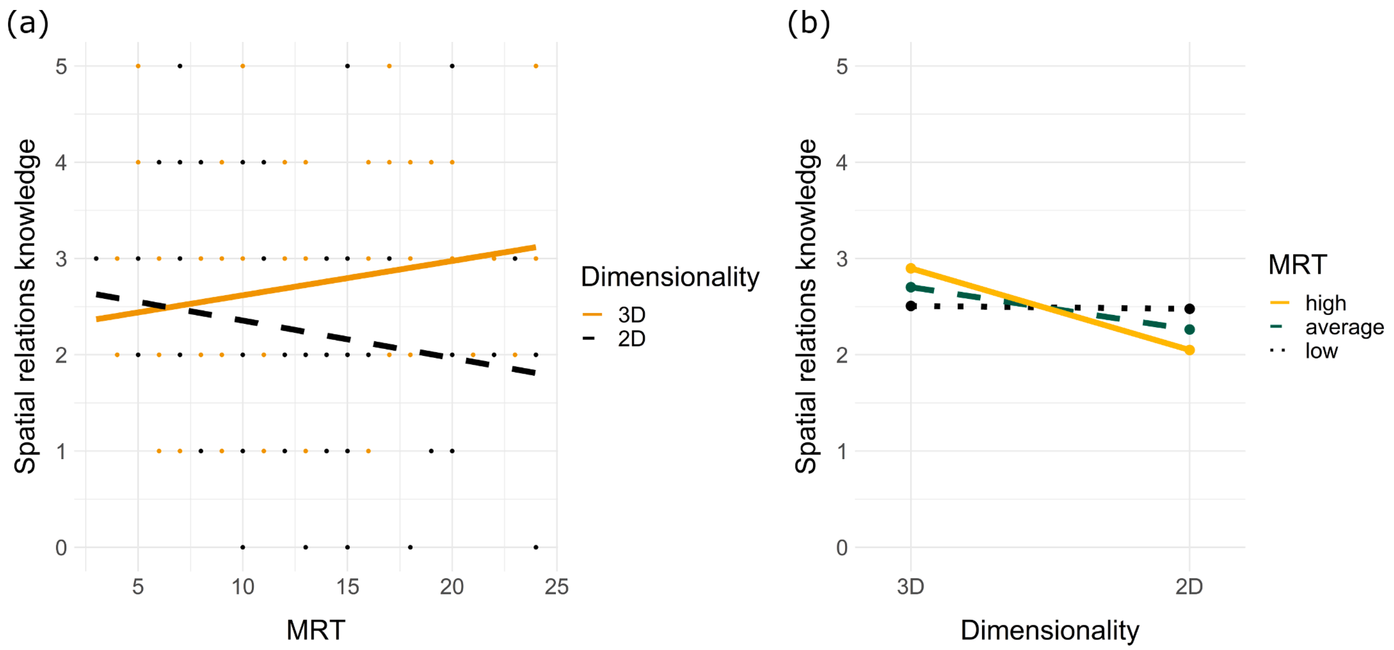


Fig. 9. (a) Interaction effect between dimensionality of visualization and MRT score on spatial relations knowledge, and (b) simple slopes of the MRT score levels concerning the effect of the visualization on spatial relations knowledge.

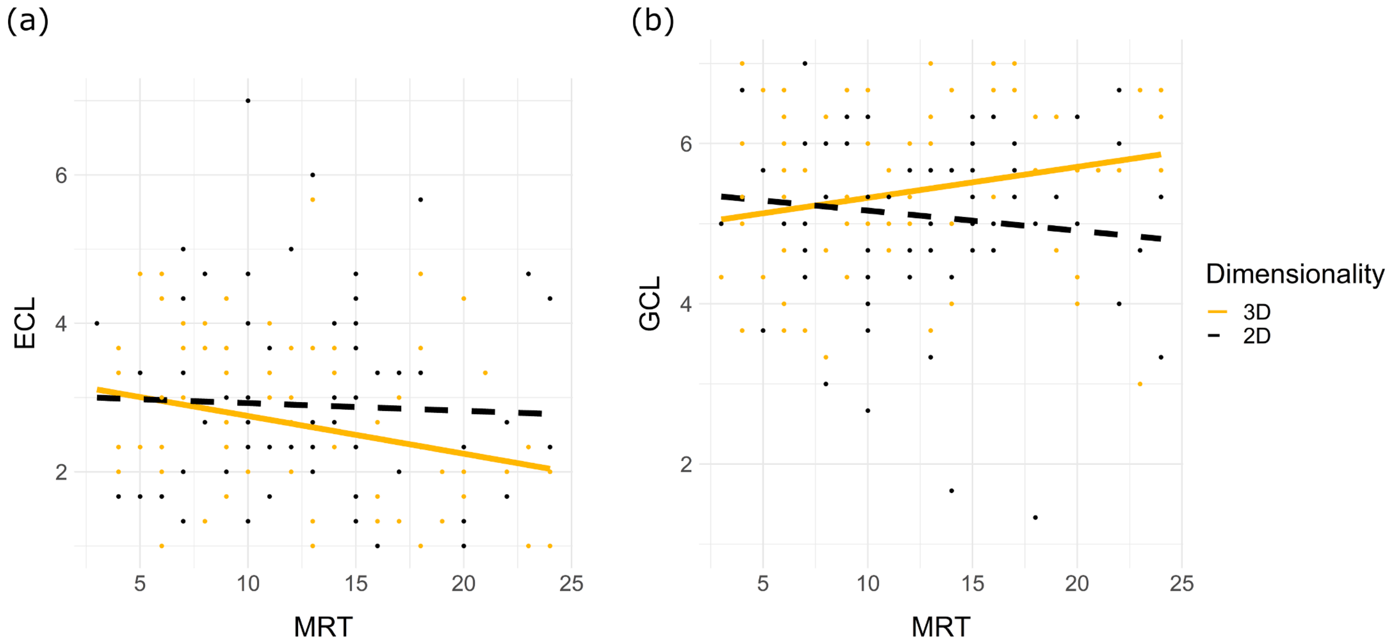


Fig. 10. Scatterplot including interaction of MRT and visualization on (a) ECL and (b) GCL.

Table 2
Completely standardized effect sizes (β) and *p*-values for the relevant effects (computed with bootstrap percentiles method with 10,000 samples) in the (a) ECL and (b) GCL mediation model for different levels of MRT score.

	Mean - 1 SD		**Mean**		**Mean + 1 SD**	
	β	*p*-value	β	*p*-value	β	*p*-value
(a) ECL model:						
Path a	−.02	.862	−.11	.154	−.21	.067
Indirect Effect	−.00	.863	.03	.196	.05	.114
(b) GCL model:						
Path a	−.02	.878	.14	.079	.30	.012*
Indirect Effect	−.01	.880	.05	.108	.11	.036*

benefit from the 3D visualization in a way that it decreases their ECL and in turn increases their spatial relations knowledge. This moderation is not significant, and it is also contrary to the moderation that was proposed in H3a, were it was suggested that especially learners with low mental rotation abilities would profit from using the 3D visualization. H3a was thus not supported by the data.

To explore through RQ3b whether the proposed moderated mediation is present in the same model but with GCL instead of ECL as a mediator variable, another moderated mediation model was tested. The interaction of dimensionality and MRT on GCL (moderation of path a) can be seen in Fig. 10(b). There, in the 3D condition GCL increases when mental rotation abilities increase, while in the 2D condition it decreases. This interaction is not significant, $F(1, 146) = 3.74$, $p = .055$, $\omega^2 = 0.02$. Although the effect is not significant, we further descriptively explore the mediation models on the different levels. In Table 2 (b) the standardized effects for the three levels on path a and the indirect effect and their *p*-values are shown again. Although the moderation is not significant, again a direction of change from lower to higher MRT score can be seen in the descriptive values. The regression coefficients of path a and the indirect effect are negative for the low MRT level and positive for the average and high level, showing that the 3D condition scores higher than the 2D condition on the GCL subscale when MRT is average or higher. This difference grows with a higher MRT score. Descriptively, students with higher mental rotation abilities thus benefit from the 3D visualization in a way that it increases their GCL and in turn increases their spatial relations knowledge. Although the moderation is not significant overall, the indirect effect of the mediation is significant when the MRT score is high, showing that GCL partially mediates the effect of dimensionality on spatial relations knowledge for this level of ability. Because this moderation is not significant and the descriptive values are again contrary to the moderation that was proposed in RQ3b, were it was suggested that especially learners with low mental rotation abilities would benefit from using the 3D visualization, RQ3b was thus also not supported by the data.

4. Discussion

The goal of the study presented in this paper is to shed light on how and under which conditions a 3D presentation in AR may be better suited for learning about spatial aspects of an object than a 2D presentation in AR, by answering the research questions "How does the dimensionality of the visualization of a 3D object in AR influence cognitive load and learning outcomes, and which role do spatial abilities play in this relationship?". Concerning the first research question, it was shown that, as expected, learners who received the 3D AR presentation had higher resulting spatial relations knowledge about the object, but not higher resulting general knowledge about the topic. While we did find that at least GCL during the learning task was influenced by the form of presentation and both ECL and GCL during the task had an influence on spatial relations knowledge, we did not find that self-reported ECL or GCL generally mediated the influence of the presentation on spatial relations knowledge. For the second research question concerning mental rotation abilities as a specific form of spatial abilities, we found that learners with average and high mental rotation abilities showed higher resulting knowledge when receiving the 3D but not the 2D presentation, while for learners with low mental rotation abilities this was not the case. This moderation was furthermore, at least partly, found in the influence of the form of presentation on cognitive load arising from GCL, where for learners with high mental rotation abilities the effect of the presentation on spatial relations knowledge was mediated by GCL cognitive load, while this was not the case for learners with average and low mental rotation abilities. 3D AR visualizations may thus be especially valuable for learning spatial structures of objects, although higher mental rotation abilities may be necessary to process the information so that it leads to better learning outcomes. In the following sections, those results and their implications are discussed in more detail.

4.1. Knowledge

Concerning the resulting knowledge of the learners using the 3D or 2D presentation of the human heart in the study, it was hypothesized that spatial components knowledge (H1a) and spatial relations knowledge (H1b) would be higher for the learners working with the 3D presentation, while general knowledge (H1c) would be the same for both groups. The hypothesis concerning spatial components knowledge was not supported, as this type of knowledge was only descriptively but not significantly higher in the 3D group. We found support for the hypothesis concerning spatial relations knowledge, showing that this type of knowledge was indeed higher for learners in the 3D group, although only with a small effect size. General knowledge was not equivalent in the two conditions with a score descriptively higher in the 3D group, so that this hypothesis is not supported, but also no significant difference was found between the groups. These outcomes show that only the result in the spatial relations part of the knowledge test was influenced by the dimensionality of the representation of the human heart. Neither spatial components knowledge nor general knowledge was significantly influenced, although it seems that the 3D group may have had at least a small advantage as seen in the descriptive results.

An explanation for why the spatial components knowledge did not differ significantly between the groups could be that its relation to the dimensionality of the presentation was not as extensive as we expected. The score on this test may have been more dependent on the recall of the exact terms than the spatial structure, although we did try to avoid that by coding all written terms that somehow resembled the correct term as correct. Furthermore, in comparison to spatial relations knowledge, less interacting information needs to be kept in memory, so that the support that the 3D visualization offers for building a correct 3D mental model of the human heart may not be as necessary. General knowledge probably was not the same in the two groups because although the questions aimed at topics explained in the text, some of them might have also had a connection to the dimensionality of the visual representation. In general, these results show that using a 3D presentation of an object in AR can lead to higher knowledge especially about more complex spatial aspects of the object but not necessarily to higher general knowledge about the topic. This supports the general notion that viewing 3D content is a main feature of AR, that AR can support spatial learning and that AR should be used in especially spatial areas (see, e.g., [20,85,105]), but further connects these ideas, showing that the virtual components in the AR material should be 3D rather than 2D to leverage the advantages concerning spatial learning. Furthermore, the distinction between different kinds of knowledge provides more detailed insights, showing that specifically learning of spatial structure of objects can be supported with 3D visualizations in AR, while learning about more general aspects is not necessarily improved.

4.2. Cognitive load

The hypotheses we proposed concerning differences between the groups in cognitive load stated that ECL would be lower (H2a) and GCL higher (H2b) with the 3D than with the 2D visualization, while ICL would be the same in the two groups (H2c). The results of the study did not support the hypothesis concerning ECL, showing that this type of load was not influenced significantly by the dimensionality of the visualization. The hypothesis concerning GCL was supported with this type of load being significantly higher for learners receiving the 3D AR visualization than for learners receiving the 2D AR visualization. These results suggest that although extraneous cognitive processing caused by the dimensionality of the representation is not significantly reduced, germane processing is indeed increased. In combination with the results concerning the increased spatial relations knowledge in the 3D group, this suggests that through the 3D AR visualization, learners could process the object more completely and more correctly into a mental model.

A potential explanation for why ECL was not lower when using the 3D in comparison to the 2D AR representation although the findings in the literature suggested it would have been, is that the learning task was different from learning material in other studies using 3D visualization. When looking at the material in the study by Stull and Hegarty [92], for example, which found a significant effect of the visualization on learning and attributed this to reduced cognitive load, it is clear that the transformation from 2D to 3D representations of molecules is the focus of the study. For the present study, no transformation of the 2D picture of the human heart into a 3D mental model was necessary as part of the learning task, so that the learners in the 2D condition may not have built a 3D mental model of the object at all. This would explain why ECL, which we hypothesized would come from a mental transformation of the dimensionality of the visualization, did not differ. It would also explain why GCL, which we hypothesized would come from the building of a 3D mental model of the object, did differ. It is also in accordance with the results showing that spatial relations knowledge was higher for the group with the 3D visualization, because the building of the 3D mental model prompted by the 3D AR visualization may have provided the learners with more correct and complete spatial relations knowledge.

Another possible explanation why ECL did not differ was that in general the load elicited by the task may not have been high enough to show a difference between the two conditions. In another study that showed a lower cognitive load measured both subjectively and objectively, the representations were used to execute a paper-folding task [26]. The learners thus had to physically interact with material while watching the representations, so that cognitive load may have in general been higher and the usage of the 3D representations more relevant than in the present study.

The assumption of equivalent ICL in the two groups was supported. This suggests that the amount of cognitive processing depending on the content of the material was very similar with the 3D and the 2D visualization and that differences in cognitive load may indeed be attributed to the presentation of the material and not a difference in complexity. We thus found different results concerning the three kinds of cognitive load, namely equivalence for ICL, descriptively higher ECL in 2D, and significantly higher GCL in 3D, which shows that differentiating between them was important for the present study, because the differences in ECL may have canceled out the differences in GCL when only measuring cognitive load in general.

Concerning the relationship of dimensionality of visualizations, cognitive load, and knowledge, two mediation hypotheses were formulated and tested. We proposed that extraneous (H2d) and germane cognitive processing (H2e) would mediate the effect of the dimensionality on spatial relations knowledge. For both hypotheses, no significant mediation effect was found. Still, as expected, the direction of the effect of dimensionality on extraneous cognitive processing was descriptively negative, while the effect of extraneous cognitive processing on spatial relations knowledge had a significant negative relation. In the second mediation, as expected, the direction of the effect of the dimensionality on germane cognitive processing was descriptively positive, while the relation of germane cognitive processing to spatial relations knowledge also showed a significant positive effect. The significant relations show that the different types of cognitive processing had the expected influence on the learning outcomes concerning spatial relations knowledge.

While cognitive load has been assumed an important aspect in learning about 3D objects and learning in AR, its specific examination in the context of AR-based education is still scarce. In a mapping review, Buchner and colleagues [13] found 64 studies that looked at cognitive load in AR, but only one study used the tripartite differentiation of intrinsic, extraneous, and germane cognitive load proposed by cognitive load theory [97], which was found to be important in the present study. Furthermore, a high proportion (73%) of the studies in the review were media comparison studies. Lee [65], for example, found a decrease in both mental effort and mental load after learners had trained with 3D AR models instead of only 2D drawings, but these results may have been confounded by different factors, for example the different interactive possibilities in the AR condition, the novelty of AR in comparison to just drawings. In the present study, we tried to limit all potential confounding factors, so that the increase in germane cognitive load and the descriptive decrease in extraneous cognitive load can be attributed to the dimensionality of the visualization.

4.3. Mental rotation abilities

The moderation analyses that were executed concerning the learners' mental rotation abilities can be used at least partially to explain why no general mediation effect of cognitive load was found. In a preliminary general moderation analysis, we found that the influence of the dimensionality of visualization on spatial relations knowledge was moderated by mental rotation abilities: the higher the abilities, the more the learners profited from using the 3D representation, in comparison to the 2D representation. Instead of supporting the ability-as-compensator hypothesis, as was hypothesized, the results rather supported the ability-as-enhancer hypothesis. It is thus in accordance with the results from the study by Huk [49] and not with the results from the meta-review by Höffler [46]. Learning with 3D AR may be a special case of spatial learning, due to the additional spatial information through perspective changing and reference to the real world, so that the results from studies with non-AR 3D material may not be completely transferable to AR. The present study thus adds insights on the role of spatial abilities and specifically mental rotation abilities in the specific case of 3D AR visualizations.

Specifically, we proposed that the effect of the visualization on extraneous cognitive processing would be moderated by the mental rotation abilities of the learners (H3a). This was not supported by the data, which showed no significant interaction effect and descriptive values opposite of what we expected. We also wanted to explore if the effect of the visualization on germane cognitive processing was moderated by learners' mental rotation abilities (RQ3b). This was not the case, with again no significant moderation effect and descriptive values opposite of what we expected. Descriptively, we found that for learners with higher mental rotation abilities, the 3D AR visualization decreased the extraneous cognitive processing, while it also descriptively increased the germane processing, which both led to increased knowledge concerning spatial aspects of the human heart. As part of the moderated mediation in RQ3b, a significant mediation of the influence of the dimensionality on the spatial relations knowledge through germane cognitive processing was found for learners with higher mental rotation abilities.

The results can be explained by the fact that 3D AR visualizations convey more information than 2D AR visualizations, namely additional spatial information. We expected that this additional information would support learners with low mental rotation abilities to more easily build an accurate mental model. Instead, the results suggest that learners with high mental rotation abilities had the tools to build more correct and comprehensive mental models from the 3D visualizations, but low abilities learners did not. For learners with high mental rotation abilities learning with the 3D visualization we found that germane processing was increased, and it mediated the effect of visualization on spatial relations knowledge. These results suggest that learners with high mental rotation abilities could use their resources to build a mental model of the human heart and thus learn more from 3D AR, while learners with low mental rotation abilities could not. It is possible that the lack of a significant mediation through ECL was due to not needing to transform the 2D AR visualization into a 3D mental model, and this could also explain why in the 2D group no advantage arose from higher mental rotation abilities – they just did not play a role in the processing of the 2D AR visualization.

Although we found in the general moderation model that mental rotation abilities moderated the influence of the dimensionality of the visualization on spatial relations knowledge, cognitive load could not completely be established as a variable to mediate the effect, so it is important to also look for other potential mediating factors. A study that looked at the usage of 3D AR visualizations of molecules to solve tasks concerning chemistry, for example, found sex difference showing that while men profited from the 3D visualizations, women profited from 2D visualizations [43]. Further analyses excluded the possibility that this was due to differences in spatial abilities, so that this was an additional factor that could play a role when learning with 3D AR visualizations, for example due to increased familiarity and experience with 3D objects (e.g., because of more experience with 3D video games on average) in men. This factor of familiarity is important to be inspected in future studies on learning with and about 3D objects.

4.4. Implications

In general, the present study used a research approach that differs from most other studies on AR-based learning, which often use a media comparison approach, comparing an AR application to, for example, a traditional medium like a book or a less traditional medium like a non-AR simulation. Media comparison studies have been widely criticized due to some challenges, for example, the uncontrollability of confounding variables and the missing knowledge about the media and learner attributes that make a medium effective [94]. Alternative studies that have been proposed by Surry and Ensminger are intra-medium comparison studies and aptitude-treatment-interaction studies. In the present study, we used both approaches, manipulating a specific attribute of an AR-based learning experience, namely dimensionality, and taking a closer look at how this attribute has an effect on learners with different characteristics, namely mental rotation abilities. By focusing on this specific attribute, other confounding variables were limited and the effect of dimensionality on knowledge and cognitive load was established, dependent on the learners' mental rotation abilities. Investigating mental rotation abilities as a potentially moderating variable provided a more nuanced picture, showing that the effects we found for knowledge and cognitive load were not present for learners with lower mental rotation abilities.

The results concerning the knowledge test show that learning about spatial aspects like the spatial relations between components of an object may be supported by using 3D instead of 2D AR visualizations. This should be taken into account when designing AR applications with the goal of supporting the forming of mental models of spatial objects or structures, but also in general when choosing whether to transport information through 2D or 3D visualizations.

The results concerning the differences in cognitive load between the conditions suggest that germane cognitive processing seems to be encouraged but extraneous cognitive processing may not necessarily be decreased when a 3D AR visualization in comparison to a 2D AR visualization is used. Perhaps 3D representations encourage the building of a 3D mental model even when the task is not directly related to the transformation of a 2D visualization into a (mental) 3D visualization, so that resulting knowledge about the object is more complete and correct. For practical applications this implies that when spatial aspects of an object are important for a subsequent task (like the knowledge test in the case of this study), using a 3D (AR) visualization can lead to a first building of a 3D mental model, but this might not occur when using a 2D (AR) visualization.

Concerning mental rotation abilities, the results support the idea that it is important to take into account the cognitive abilities of learners when implementing 3D (AR) visualizations, because not everybody may learn from them in the same way. A parallel can be seen in research on the usage of static and dynamic visualizations, where it has also been found that people with lower spatial abilities benefit from dynamic visualizations, while people with higher spatial abilities do not [8,46]. Also, the benefit of a combination of spoken words and animations has been found to depend on spatial abilities, showing more benefits for students with higher spatial abilities [76]. Second, the present study specifically supports an ability-as-enhancer hypothesis. Although this is contrary to what we expected based on the meta-review by Höffler [46], the present study provides new insights for the specific case of 3D AR material, which may be perceived differently from non-AR material due to the reference to the real world and the potential to move around it. The results expand the findings of another study on the role of spatial abilities in learning with 3D AR visualizations, which showed support for the ability-as-compensator hypothesis in the context of 3D spatial visualization abilities and learning task results, while showing support for the ability-as-enhancer hypothesis in the context of 2D spatial memory abilities and spatial knowledge test outcomes [3]. While 2D spatial memory abilities were not included in the present study, 3D spatial visualization abilities were measured with a different mental rotation test than in the referenced study. Both hypotheses seem to have a place in AR, but the specific abilities and the role they play for which learning processes and outcomes need to be further disentangled in future research. It is important to know which learners benefit from which sort of visualization. For this case, the potential gap between the learners can be closed, because it is possible to train learners' spatial abilities [79]. New technologies that can visualize virtual 3D models like AR, VR, and desktop applications have the potential to help with this training (e.g., [22,31,74]).

4.5. Limitations and future studies

There were some limitations in the design and execution of the study. First, although the goal was to only manipulate dimensionality as a part of the spatiality of an AR visualization, the manipulation check showed that both interactivity and contextuality were also perceived as more pronounced in the 3D condition than in the 2D condition. This may on one hand imply that the manipulation of the dimensionality did not work as intended. On the other hand, this may show that dimensionality is related to not only spatiality, but that it is also interwoven with the other characteristics, so that it is difficult to measure them separately. Moving around a virtual 3D object might, for example, be perceived as different from moving around a virtual 2D graphic, and the context might receive a different meaning when an object floats above a scene than when it lies on a surface. We did not track or observe how the participants interacted with the applications, so we do not know if the interaction was indeed different for the two conditions. In future studies it would be interesting to see how exactly participants use the interactive potential of the applications and if different kinds of visualizations lead to different interaction. Furthermore, tracked or observed data concerning interactive activities could be reviewed with a qualitative approach, providing deeper insights into the specific learning processes that take place and going beyond the systematic comparison of the experimental approach. This furthermore shows that it is important to differentiate between the technological implementation of an AR visualization and how learners use and experience it. We tried to change the users' experience of the spatiality by manipulating one spatial factor, namely the dimensionality of the visualization. This led to not only a difference in the perceived spatiality, but also the perceived interactivity and contextuality, which shows that the manipulation of one technological factor in AR can influence the whole psychological experience of the users. We plan to further develop and validate the ARcis questionnaire on the three characteristics of AR, so that in the future it may, for example, be used as a manipulation check in studies when the goal is to manipulate one of the characteristics, or to compare learners' experience of two different AR applications concerning the three characteristics.

While we focused on a spatial anatomy-related topic in the present study, the virtual object we used had some limitations. Because it was ultimately a three-dimensional cross-section of the human heart, the 3D representation was not as necessary as it would have been with an object that would be viewed from the outside. In biology education, 2D images of cross-sections of anatomical structures are often used, because they include most of the necessary information, as did the 2D AR representation that was used here. In the present study, an added value by the 3D visualization was still found, which may have only had to do with the encouragement of the building of a 3D mental model but not with the ease of understanding or processing of the representation. For an object for which the outer structure without a cross-section is the focus of a learning task (e.g., the structure of the modules of the international space station; the structure of buildings in a city), it may be even more important to visualize it in 3D. Furthermore, the complexity of the object may play a role here. The heart as an organ and its different components are of course very complex, but the content we used in the present study came from grade 5/6 schoolbooks, so that it was probably not that difficult for our participants to learn. As the complexity of a spatial structure increases, either because it is viewed on a deeper level or because the structure is so big that many different components are included, it may be even more useful to use a 3D AR visualization. A variation in the complexity of the spatiality of the structure and a replication of the results from this study with similarly complex material will be considered for future studies.

Cognitive load was measured through a questionnaire based on subjective retrospective self-reports. This may be a problem, because learners may not always be able to monitor and have insights into their cognitive state. Especially learners with lower prior knowledge may not necessarily rate cognitive load as expected and may, for example, confuse ICL and ECL [110]. Furthermore, the state of cognitive load may change over time during the learning task, which cannot be tracked with just one post-task measurement. The questionnaire by Klepsch et al. [57] was chosen for the present study, because it differentiates between extraneous, germane, and intrinsic cognitive load, which was important for the hypotheses. Objective measures may be able to give a less biased picture of the cognitive load because no introspection is necessary, but the mapping of different objective measures and the three types of cognitive load is not straightforward. There have already been attempts to map physiological, objective measures like pupillary and eye-movement data captured with eye-tracking technologies to the individual types of cognitive load, but a lot more research is necessary in this area [53,109]. Still, collecting those data to confirm the general tendency of the self-reports may be interesting for future studies and especially the possibility to view the development of load over time may be an informative addition here. In general, the present study has shown that splitting cognitive load into the different types can be beneficial for gaining more detailed insights, which we believe will also be the case in other studies, so that self-report questionnaires may be a necessary tool but should be enriched through additional objective measurements.

Although the participants in the study sample were randomly distributed to the two conditions except for a matching of gender between the two groups, the self-reported pre-knowledge concerning the learning topic differed between the groups. The other pre-measured variables (task value, task expectancy, mental rotation abilities) were also not equivalent in the two groups, although they did not differ significantly. The group who afterwards received the 2D AR visualization rated their prior knowledge as higher than the group who afterwards received the 3D version. This could have had an impact on both knowledge and cognitive load measures, especially content-related load. However, the knowledge test showed a higher resulting knowledge in the 3D than the 2D condition, so that the direction of the difference was even reversed from subjective prior to measured posterior knowledge. In future studies, this problem should be approached beforehand, so that the samples in the conditions are matched concerning their prior knowledge and potentially other important variables. An objective knowledge test could also be administered instead of a self-rating. Here we decided against a test so that we do not prime participants to focus on the spatial aspects because that was one potential effect of the 3D representation which we did not want to diminish.

In the present study, we focused on mental rotation abilities as one form of spatial abilities. There are, of course, also other spatial abilities that may play a role when processing a 3D visualization. It is quite certain that those other abilities that we did not measure also played a role for the learners in the learning task in our study. In another study on learning of spatial structures, Krüger and Bodemer (2021) found that different kinds of spatial abilities had different moderating influence on learning tasks and outcomes. Future studies should also examine different spatial abilities including the roles they may play in the processing of 3D AR visualizations and if the results of the present study can be generalized to other spatial abilities.

The focus of the present study was on the manipulation of one specific factor of the visualization of an object in AR, namely the dimensionality, which we propose to be part of the spatiality of AR visualizations. AR, of course, has much more potential than just adding a third dimension to a virtual element. Virtual 3D elements could also be dynamic (e.g., [19,52,89]), interactive (e.g., [21,33,34]), and embedded into a relevant context [7,50,86]. The interaction of these different factors that can play a role when visualizing virtual elements in AR, should be examined more closely, as they are often not used independently but concurrently in AR applications. The chosen approach of intra-medium comparison and aptitude-treatment-interaction analysis provided some important advantages for the interpretation of the data. For example, confounding variables can be ruled out and the difference between effects in higher and lower mental rotation abilities learners could be detected. We suggest executing more studies including these approaches in the future.

5. Conclusion

All in all, the study provides some interesting findings concerning learning with 2D and 3D AR visualizations, its relation to cognitive load and the potential influence of mental rotation abilities. The 3D visualization of objects in AR can thus have a positive influence, increasing germane cognitive load and learning of spatial structures, although an adequate level of mental rotation abilities may be necessary for effective processing of the information. The study focuses on the dimensionality of the visualization as an isolated factor, while trying to keep all other potentially influential factors as similar as possible between the conditions. Additionally, learners' mental rotation abilities are taken into account. The results show that AR can be used to visualize a 3D representation in a way that it is quite easy to receive by learners and can also lead to improved learning outcomes, especially for learners with higher mental rotation abilities. To establish a further empirical basis for the design and implementation of AR learning environments with a focus on spatial objects, more systematic and empirical research focusing on the dimensionality of representation as an important part of the AR characteristic spatiality and its interaction with the other characteristics contextuality and interactivity is still necessary. Other learner characteristics and skills should also be considered for research, to inform the design for specific target groups of AR-based learning applications. AR seems to have a bright future for education – if it is implemented adequately for specific learning goals.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interest

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

References

[1] Akçayır M, Akçayır G. Advantages and challenges associated with augmented reality for education: a systematic review of the literature. Educ Res Rev 2017;20:1–11. https://doi.org/10.1016/j.edurev.2016.11.002.
[2] Krüger JM, Buchholz A, Bodemer D. Augmented reality in education: three unique characteristics from a user's perspective. In: Chang M, So H-J, Wong L-H, Yu F-Y, Shih JL, editors. Proceedings of the 27th international conference on computers in education. Taiwan: Asia-Pacific Society for Computers in Education; 2019. p. 412–22.
[3] Krüger JM, Bodemer D. Space, a central frontier—the role of spatial abilities when learning the structure of 3D AR objects. IEEE; 2021. p. 258–65. https://doi.org/10.23919/iLRN52045.2021.9459365.
[4] Azer SA, Azer S. 3D anatomy models and impact on learning: a review of the quality of the literature. Health Prof Educ 2016;2(2):80–98. https://doi.org/10.1016/j.hpe.2016.05.002.
[5] Azuma R. A survey of augmented reality. Presence Teleoper Virtual Environ 1997;6(4):355–85. https://doi.org/10.1162/pres.1997.6.4.355.
[6] Bacca J, Baldiris S, Fabregat R, Graf Kinshuk S. Augmented reality trends in education: a systematic review of research and applications. Educ Technol Soc 2014;17(4):133–49.
[7] Barmaki R, Yu K, Pearlman R, Shingles R, Bork F, Osgood GM, Navab N. Enhancement of anatomical education using augmented reality: an empirical study of body painting. Anat Sci Educ 2019;12(6):599–609. https://doi.org/10.1002/ase.1858.
[8] Berney S, Bétrancourt M, Molinari G, Hoyek N. How spatial abilities and dynamic visualizations interplay when learning functional anatomy with 3D anatomical models. Anat Sci Educ 2015;8(5):452–62. https://doi.org/10.1002/ase.1524.
[9] Billinghurst M, Dünser A. Augmented reality in the classroom. Computer 2012;45(7):56–63. https://doi.org/10.1109/MC.2012.111.
[10] Billinghurst M, Kato H, Poupyrev I. The MagicBook: a transitional AR interface. Comput Graph 2001;25(5):745–53. https://doi.org/10.1016/S0097-8493(01)00117-0.
[11] Bower M, Howe C, McCredie N, Robinson A, Grover D. Augmented reality in education—cases, places and potentials. Educ Media Int 2014;51(1):1–15. https://doi.org/10.1080/09523987.2014.889400.
[12] Brown D. The Orion constellation as an installation: an innovative three-dimensional teaching and learning environment. Phys Teach 2013;51(3):160–2. https://doi.org/10.1119/1.4792013.
[13] Buchner J, Buntins K, Kerres M. A systematic map of research characteristics in studies on augmented reality and cognitive load. Comput Educ Open 2021;2:100036. https://doi.org/10.1016/j.caeo.2021.100036.
[14] Campbell N.A., Reece J.B., Urry L.A., Cain M.L., Wasserman S.A., Minorsky P.V., et al. Kreislauf und Gasaustausch. In: Heinisch J.J., Paululat A., editors. Campbell Biologie. 10th ed. Pearson Studium: 2016. p. 1229–1268.
[15] Carroll JB. Human cognitive abilities: a survey of factor-analytic studies. Human cognitive abilities: a survey of factor-analytic studies. Cambridge University Press; 1993.
[16] Castro-Alonso J.C., Jansen P. Sex differences in visuospatial processing. In: Castro-Alonso J.C., editor. Visuospatial processing for education in health and natural sciences. Springer International Publishing; 2019. p. 81–110. doi:10.1007/978-3-030-20969-8_4.
[17] Chen CH, Yang JC, Shen S, Jeng MC. A desktop virtual reality earth motion system in astronomy education. J Educ Technol Soc 2007;10(3):289–304.
[18] Chen P., Liu X., Cheng W., Huang R. A review of using augmented reality in education from 2011 to 2016. In: Popescu E., Kinshuk, Khribi M.K., Huang R., Jemni M., Chen N.-S., Sampson D.G., editors. Innovations in smart learning, lecture notes in educational technology. Singapore: Springer; 2017. p. 13–18.
[19] Chen SC, Hsiao MS, She HC. The effects of static versus dynamic 3D representations on 10th grade students' atomic orbital mental model construction: evidence from eye movement behaviors. Comput Hum Behav 2015;53:169–80. https://doi.org/10.1016/j.chb.2015.07.003.
[20] Cheng KH, Tsai CC. Affordances of augmented reality in science learning: suggestions for future research. J Sci Educ Technol 2013;22(4):449–62. https://doi.org/10.1007/s10956-012-9405-9.
[21] Chien CH, Chen CH, Jeng TS. An interactive augmented reality system for learning anatomy structure. In: Proceedings of the international multiconference of engineers and computer scientists (IMECS 2010). Vol. 1. Hong Kong: IAENG; 2010. p. 17–9. Vol. 1, IMECS 2010.
[22] Cohen CA, Hegarty M. Visualizing cross sections: training spatial thinking using interactive animations and virtual objects. Learn Individ Differ 2014;33:63–71. https://doi.org/10.1016/j.lindif.2014.04.002.
[23] Cohen J. Statistical power analysis for the behavioral sciences. 2nd. L. Erlbaum Associates; 1988.
[24] Copolo CE, Hounshell PB. Using three-dimensional models to teach molecular structures in high school chemistry. J Sci Educ Technol 1995;4(4):295–305. https://doi.org/10.1007/BF02211261.
[25] Craig A.B. Chapter 2-Augmented reality concepts. In: Craig A.B, editor. Understanding augmented reality. Morgan Kaufmann; 2013. p. 39–67. doi:10.1016/B978-0-240-82408-6.00002-3.
[26] Dan A, Reiner M. Reduced mental load in learning a motor visual task with virtual 3D method. J Comput Assist Learn 2018;34(1):84–93. https://doi.org/10.1111/jcal.12216.
[27] Delacre M, Lakens D, Leys C. Why psychologists should by default use Welch's *t*-test instead of Student's *t*-test. Int Rev Soc Psychol 2017;30(1):92–101. https://doi.org/10.5334/irsp.82.

[28] Donaghy KJ, Saxton KJ. Connecting geometry and chemistry: a three-step approach to three-dimensional thinking. J Chem Educ 2012;89(7):917–20. https://doi.org/10.1021/ed200345w.
[29] Dori YJ, Barak M. Virtual and physical molecular modeling: fostering model perception and spatial understanding. J Educ Technol Soc 2001;4(1):61–74.
[30] Dunleavy M., Dede C. Augmented reality teaching and learning. In: Spector J.M., Merrill M.D., Elen J., Bishop M.J., editors. Handbook of research on educational communications and technology. 4th ed. New York: Springer; 2014. p. 735–745.
[31] Dünser A, Steinbügl K, Kaufmann H, Glück J. Virtual and augmented reality as spatial ability training tools. In: Proceedings of the 7th ACM SIGCHI New Zealand chapter's international conference on computer-human interaction: design centered HCI (CHINZ '06). ACM; 2006. p. 125–32.
[32] Efron B, Tibshirani RJ. Confidence intervals based on bootstrap percentiles. An introduction to the bootstrap. US: Springer; 1993. p. 168–77. https://doi.org/10.1007/978-1-4899-4541-9.
[33] Erolin C. Interactive 3D digital models for anatomy and medical education. In: Rea P.M., editor. Biomedical visualisation, 2. Springer International Publishing; 2019. p. 116. doi:10.1007/978-3-030-14227-8_1.
[34] Fatemah A, Rasool S, Habib U. Interactive 3D visualization of chemical structure diagrams embedded in text to aid spatial learning process of students. J Chem Educ 2020;97(4):992–1000. https://doi.org/10.1021/acs.jchemed.9b00690.
[35] Ferrara F, Mammana MF. Seeing in space is difficult: an approach to 3D geometry through a DGE. In: Proceedings of the joint meeting of PME 38 and PME-NA 36, 3; 2014. p. 57–64.
[36] Fleck S, Simon G. An augmented reality environment for astronomy learning in elementary grades: an exploratory study. In: Proceedings of the 25th ICME conference francophone on l'interaction homme-machine - IHM '13; 2013. p. 14–22. https://doi.org/10.1145/2534903.2534907.
[37] Foo JL, Martinez-Escobar M, Juhnke B, Cassidy K, Hisley K, Lobe T, Winer E. Evaluating mental workload of two-dimensional and three-dimensional visualization for anatomical structure localization. J Laparoendosc Adv Surg Tech 2013;23(1):65–70. https://doi.org/10.1089/lap.2012.0150.
[38] Garg AX, Norman G, Sperotable L. How medical students learn spatial anatomy. Lancet N Am Ed 2001;357(9253):363–4. https://doi.org/10.1016/S0140-6736(00)03649-7.
[39] Garzón J, Acevedo J. Meta-analysis of the impact of Augmented Reality on students' learning gains. Educ Res Rev 2019;27:244–60. https://doi.org/10.1016/j.edurev.2019.04.001.
[40] Garzón J, Pavón J, Baldiris S. Systematic review and meta-analysis of augmented reality in educational settings. Virtual Real 2019;23(4):447–59. https://doi.org/10.1007/s10055-019-00379-9.
[41] Gottlieb M, Haala G, Beyer I. Gesund und fit. Natura: Biologie für Gymnasien 1. Ernst Klett Verlag; 2009. p. 105–45.
[42] Guedes KB, de Sá Guimarães M, de S, Méxas JG. Virtual reality using stereoscopic vision for teaching/learning of descriptive geometry. In: Proceedings of the fourth international conference on mobile, hybrid, and on-line learning, ELmL. IARIA; 2012. p. 24–30.
[43] Habig S. Who can benefit from augmented reality in chemistry? Sex differences in solving stereochemistry problems using augmented reality. Br J Educ Technol 2020;51(3):629–44. https://doi.org/10.1111/bjet.12891.
[44] Hackett M, Proctor M. Three-dimensional display technologies for anatomical education: a literature review. J Sci Educ Technol 2016;25(4):641–54. https://doi.org/10.1007/s10956-016-9619-3.
[45] Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. Advances in psychology, 52. Elsevier; 1988. p. 139–83. https://doi.org/10.1016/S0166-4115(08)62386-9.
[46] Höffler TN. Spatial ability: its influence on learning with visualizations—a meta-analytic review. Educ Psychol Rev 2010;22(3):245–69. https://doi.org/10.1007/s10648-010-9126-7.
[47] Hollender N, Hofmann C, Deneke M, Schmitz B. Integrating cognitive load theory and concepts of human-computer interaction. Comput Hum Behav 2010;26(6):1278–88. https://doi.org/10.1016/j.chb.2010.05.031.
[48] Holmes CA, Newcombe NS, Shipley TF. Move to learn: integrating spatial information from multiple viewpoints. Cognition 2018;178:7–25. https://doi.org/10.1016/j.cognition.2018.05.003.
[49] Huk T. Who benefits from learning with 3D models? The case of spatial ability. J Comput Assist Learn 2006;22(6):392–404. https://doi.org/10.1111/j.1365-2729.2006.00180.x.
[50] Hwang W-Y, Zhao L, Shadiev R, Lin L-K, Shih TK, Chen H-R. Exploring the effects of ubiquitous geometry learning in real situations. Educ Technol Res Dev 2020;68(3):1121–47. https://doi.org/10.1007/s11423-019-09730-y.
[51] Isik-Ercan Z., Kim B., Nowak J. 3D visualization in elementary education astronomy: teaching urban second graders about the sun, earth, and moon. In: Lytras M.D., Ordonez De Pablos P., Ziderman A., Roulstone A., Maurer H., Imber J.B., editors. Knowledge management, information systems, e-learning, and sustainability research, 111. Berlin Heidelberg: Springer; 2010. p. 500–505. doi:10.1007/978-3-642-16318-0_64.
[52] Kamat VR, Martinez JC. Dynamic 3D visualization of articulated construction equipment. J Comput Civ Eng 2005;19(4):356–68. https://doi.org/10.1061/(ASCE)0887-3801(2005)19:4(356).
[53] Kastaun M, Meier M, Küchemann S, Kuhn J. Validation of cognitive load during inquiry-based learning with multimedia scaffolds using subjective measurement and eye movements. Front Psychol 2021;12:3660. https://doi.org/10.3389/fpsyg.2021.703857.
[54] Kendall MG. A new measure of rank correlation. Biometrika 1938;30(1–2):81–93. https://doi.org/10.1093/biomet/30.1-2.81.
[55] Kerawalla L, Luckin R, Seljeflot S, Woolard A. "Making it real": exploring the potential of augmented reality for teaching primary school science. Virtual Real 2006;10(3–4):163–74. https://doi.org/10.1007/s10055-006-0036-4.
[56] Klein MI, Lio CH, Grant R, Carswell CM, Strup S. A mental workload study on the 2d and 3d viewing conditions of the da Vinci surgical robot. In: Proceedings of the human factors and ergonomics society annual meeting. 53(18). HFES; 2009. p. 1186–90.
[57] Klepsch M, Schmitz F, Seufert T. Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. Front Psychol 2017;8. https://doi.org/10.3389/fpsyg.2017.01997.
[58] Kösa T, Karakuş F. Using dynamic geometry software Cabri 3D for teaching analytic geometry. Proced Soc Behav Sci 2010;2(2):1385–9. https://doi.org/10.1016/j.sbspro.2010.03.204.
[59] Küçük S, Kapakin S, Göktaş Y. Learning anatomy via mobile augmented reality: effects on achievement and cognitive load. Anat Sci Educ 2016;9(5):411–21. https://doi.org/10.1002/ase.1603.
[60] Lai A-F, Chen C-H, Lee G-Y. An augmented reality-based learning approach to enhancing students' science reading performances from the perspective of the cognitive load theory. Br J Educ Technol 2019;50(1):232–47. https://doi.org/10.1111/bjet.12716.
[61] Lakens D. Performing high-powered studies efficiently with sequential analyses. Eur J Soc Psychol 2014;44(7):701–10. https://doi.org/10.1002/ejsp.2023.
[62] Lakens D. The 20% statistician: absence of evidence is not evidence of absence: testing for equivalence. The 20% Statistician; 2016, May 20. Retrieved from http://daniellakens.blogspot.com/2016/05/absence-of-evidence-is-not-evidence-of.html.
[63] Lakens D. The 20% Statistician: how a power analysis implicitly reveals the smallest effect size you care about. The 20% Statistician; 2017, May 11. Retrieved from http://daniellakens.blogspot.com/2017/05/how-power-analysis-implicitly-reveals.html.
[64] Lakens D, Scheel AM, Isager PM. Equivalence testing for psychological research: a tutorial. Adv Methods Pract Psychol Sci 2018;1(2):259–69. https://doi.org/10.1177/2515245918770963.
[65] Lee I-J. Using augmented reality to train students to visualize three-dimensional drawings of mortise–tenon joints in furniture carpentry. Interact Learn Environ 2019:1–15. https://doi.org/10.1080/10494820.2019.1572629.
[66] Lindgren R, Moshell JM. Supporting children's learning with body-based metaphors in a mixed reality environment. In: Proceedings of the 10th international conference on interaction design and children - IDC '11; 2011. p. 177–80. https://doi.org/10.1145/1999030.1999055.
[67] Liou H-H, Yang SJH, Chen SY, Tarng W. The influences of the 2D image-based augmented reality and virtual reality on student learning. J Educ Technol Soc 2017;20(3):110–21.
[68] Livingston M.A., Dey A., Sandor C., Thomas B.H. Pursuit of "X-Ray Vision" for augmented reality. In: Huang W., Alem L., Livingston M.A., editors. Human factors in augmented reality environments. New York: Springer; 2013. p. 67–107. doi:10.1007/978-1-4614-4205-9_4.
[69] Lumley T, Diehr P, Emerson S, Chen L. The importance of the normality assumption in large public health data sets. Annu Rev Public Health 2002;23(1):151–69. https://doi.org/10.1146/annurev.publhealth.23.100901.140546.
[70] Luursema JM, Verwey WB, Kommers PAM, Geelkerken RH, Vos HJ. Optimizing conditions for computer-assisted anatomical learning. Interact Comput 2006;18(5):1123–38. https://doi.org/10.1016/j.intcom.2006.01.005.
[71] Macchiarella ND, Liu D, Gangadharan SN, Vincenzi DA, Majoros AE. Augmented reality as a training medium for aviation/aerospace application. In: Proceedings of the human factors and ergonomics society annual meeting. 49; 2005. p. 2174–8. https://doi.org/10.1177/154193120504902512.
[72] Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. Ann Math Stat 1947;18(1):50–60. https://doi.org/10.1214/aoms/1177730491.
[73] Marks SC. The role of three-dimensional information in health care and medical education: the implications for anatomy and dissection. Clin Anat 2000;13(6):448–52. https://doi.org/10.1002/1098-2353(2000)13:6<448::AID−CA10>3.0.CO;2-U. 10.1002/1098-2353(2000)13:6<448::AID−CA10>3.0.CO;2-U.
[74] Martín-Gutiérrez J, Contero M, Alcañiz M. Augmented reality to training spatial skills. Proced Comput Sci 2015;77:33–9. https://doi.org/10.1016/j.procs.2015.12.356.
[75] Martín-Gutiérrez J, Luís Saorín J, Contero M, Alcañiz M, Pérez-López DC, Ortega M. Design and validation of an augmented book for spatial abilities development in engineering students. Comput Graph 2010;34(1):77–91. https://doi.org/10.1016/j.cag.2009.11.003.
[76] Mayer RE, Sims VK. For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. J Educ Psychol 1994;86(3):389–401. https://doi.org/10.1037/0022-0663.86.3.389.
[77] Microsoft. Remix 3D. Microsoft; 2016. www.remix3d.com.
[78] Bickel H., Spieß C. Gesund und fit. Natura: Biologie 5/6, G9-Ausgabe, Nordrhein-Westfalen. Ernst Klett Verlag; 2019. p. 145–190.
[79] Newcombe NS, Stieff M. Six myths about spatial thinking. Int J Sci Educ 2012;34(6):955–71. https://doi.org/10.1080/09500693.2011.588728.
[80] Núñez M, Quirós R, Núñez I, Carda JB, Camahort E. Collaborative augmented reality for inorganic chemistry education. In: Proceedings of the 5th WSEAS/IASME international conference on engineering education. 5; 2008. p. 271–7.
[81] Peters M, Laeng B, Latham K, Jackson M, Zaiyouna R, Richardson C. A redrawn Vandenberg and Kuse mental rotations test—different versions and factors that affect performance. Brain Cogn 1995;28(1):39–58. https://doi.org/10.1006/brcg.1995.1032.

[82] Preece D, Williams SB, Lam R, Weller R. "Let's Get Physical": advantages of a physical model over 3D computer models and textbooks in learning imaging anatomy. Anat Sci Educ 2013;6(4):216–24. https://doi.org/10.1002/ase.1345.
[83] Prokýšek M, Rambousek V, Wildová R. Research into spatial intelligence and the efficiency of the application of spatial visualization in instruction. Proced Soc Behav Sci 2013;84:855–9. https://doi.org/10.1016/j.sbspro.2013.06.661.
[84] PTC Inc. Vuforia Augmented Reality SDK (Version 7.5). PTC Inc; 2018.
[85] Radu I. Augmented reality in education: a meta-review and cross-media analysis. Pers Ubiquitous Comput 2014;18(6):1533–43. https://doi.org/10.1007/s00779-013-0747-y.
[86] Radu I, Schneider B. What can we learn from augmented reality (AR)?. In: Proceedings of the CHI conference on human factors in computing systems - CHI '19; 2019. p. 1–12. https://doi.org/10.1145/3290605.3300774.
[87] Renkl A, Atkinson RK. Structuring the transition from example study to problem solving in cognitive skill acquisition: a cognitive load perspective. Educ Psychol 2003;38(1):15–22. https://doi.org/10.1207/S15326985EP3801_3.
[88] Shelton B.E., Stevens R.R. Using coordination classes to interpret conceptual change in astronomical thinking. In: Kafai Y.B., Sandoval W.A., Enyedy N., Nixon A.S., Herrera F., editors. Proceedings of the 6th international conference on learning sciences. Santa Monica, CA: Lawrence Erlbaum Associates; 2004. p. 634.
[89] Shin D, Park S. 3D learning spaces and activities fostering users' learning, acceptance, and creativity. J Comput High Educ 2019;31(1):210–28. https://doi.org/10.1007/s12528-019-09205-2.
[90] Sommerauer P, Müller O. Augmented reality for teaching and learning—a literature review on theoretical and empirical foundations. In: Proceedings of the twenty-sixth European conference on information systems (ECIS2018). AISeL; 2018. https://aisel.aisnet.org/ecis2018_rp/31.
[91] Song KS, Lee WY. A virtual reality application for geometry classes. J Comput Assist Learn 2002;18(2):149–56. https://doi.org/10.1046/j.0266-4909.2001.00222.x.
[92] Stull AT, Hegarty M. Model manipulation and learning: fostering representational competence with virtual and concrete models. J Educ Psychol 2016;108(4):509–27. https://doi.org/10.1037/edu0000077.
[93] Sun K-T, Lin C-L, Wang S-M. A 3-D virtual reality model of the sun and the moon for e-learning at elementary schools. Int J Sci Math Educ 2010;8(4):689–710. https://doi.org/10.1007/s10763-009-9181-z.
[94] Surry DW, Ensminger D. What's wrong with media comparison studies? Educ Technol 2001;41(4):32–5.
[95] Sweller J. Cognitive load theory. In: Mestre J.P., Ross B.H., editors. Psychology of Learning and Motivation, 55. Elsevier; 2011. p. 37–76. doi: 10.1016/B978-0-12-387691-1.00002-8.
[96] Sweller J, van Merriënboer JJG, Paas FGWC. Cognitive architecture and instructional design: 20 years later. Educ Psychol Rev 2019;31(2):261–92. https://doi.org/10.1007/s10648-019-09465-5.
[97] Sweller J, van Merrienboer JJG, Paas FGWC. Cognitive architecture and instructional design. Educ Psychol Rev 1998;10(3):251–96. https://doi.org/10.1023/A:1022193728205.
[98] Triepels CPR, Smeets CFA, Notten KJB, Kruitwagen RFPM, Futterer JJ, Vergeldt TFM, Van Kuijk SMJ. Does three-dimensional anatomy improve student understanding? Clin Anat 2020;33(1):25–33. https://doi.org/10.1002/ca.23405.
[99] Unity Technologies. Unity (Version 2018.2.11f1). Unity Technologies; 2018.
[100] Valimont RB, Gangadharan SN, Vincenzi DA, Majoros AE. The effectiveness of augmented reality as a facilitator of information acquisition in aviation maintenance applications. J Aviat/Aerosp Educ Res 2007. https://doi.org/10.15394/jaaer.2007.1478.
[101] van Merriënboer JJG, Kester L, Paas FGWC. Teaching complex rather than simple tasks: balancing intrinsic and germane load to enhance transfer of learning. Appl Cogn Psychol 2006;20(3):343–52. https://doi.org/10.1002/acp.1250.
[102] Vincenzi DA, Valimont RB, Macchiarella N, Opalenik C, Gangadharan SN, Majoros AE. The effectiveness of cognitive elaboration using augmented reality as a training and learning paradigm. In: Proceedings of the human factors and ergonomics society annual meeting. 47. HFES; 2003. p. 2054–8. https://doi.org/10.1177/154193120304701909.
[103] Welch BL. The generalization of `Student's' problem when several different population variances are involved. Biometrika 1947;34(1/2):28–35. https://doi.org/10.2307/2332510.
[104] Wigfield A, Eccles JS. Expectancy–value theory of achievement motivation. Contemp Educ Psychol 2000;25(1):68–81. https://doi.org/10.1006/ceps.1999.1015.
[105] Wu HK, Lee SW-Y, Chang H-Y, Liang J-C. Current status, opportunities and challenges of augmented reality in education. Comput Educ 2013;62:41–9. https://doi.org/10.1016/j.compedu.2012.10.024.
[106] Wu H-K, Shah P. Exploring visuospatial thinking in chemistry learning. Sci Educ 2004;88(3):465–92. https://doi.org/10.1002/sce.10126.
[107] Yuen SCY, Yaoyuneyong G, Johnson E. Augmented reality: an overview and five directions for AR in education. J Educ Technol Dev Exch 2011;4(1). https://doi.org/10.18785/jetde.0401.10.
[108] Zinchenko YP, Khoroshikh PP, Sergievich AA, Smirnov AS, Tumyalis AV, Kovalev AI, Gutnikov SA, Golokhvast KS. Virtual reality is more efficient in learning human heart anatomy especially for subjects with low baseline knowledge. New Ideas Psychol 2020;59:100786. https://doi.org/10.1016/j.newideapsych.2020.100786.
[109] Zu T, Hutson J, Loschky LC, Rebello NS. Using eye movements to measure intrinsic, extraneous, and germane load in a multimedia learning environment. J Educ Psychol 2019. https://doi.org/10.1037/edu0000441.
[110] Zu T, Munsell J, Rebello NS. Subjective measure of cognitive load depends on participants' content knowledge level. Front Educ 2021;6. https://doi.org/10.3389/feduc.2021.647097.

8.4 Paper 4 – Krüger & Bodemer, subm.

Krüger, J. M., & Bodemer, D. (subm.). *Positioning augmented reality information for learning in nature: An exploratory pilot study* [Manuscript submitted for publication].

Positioning augmented reality information for learning in nature: An exploratory pilot study

Jule M. Krüger and Daniel Bodemer

Abstract

Background: Augmented reality (AR) is an innovative way of visualizing instructional information combining virtual and physical elements. One promising function concerns the placement of virtual information at contextually relevant points in physical natural environments.

Method: We describe an exploratory pilot study examining the influence of this positioning on learning behavior, processes, and outcomes, using a tablet-based learning experience on local plants in nature. In a between-subjects design, information about local plants is either anchored to the corresponding plant or positioned in its vicinity but separated. We expect that the connection of virtual information and plants has a positive influence on learning behavior, immersion, motivation, germane cognitive load and learning outcomes.

Findings: In interviews with 18 participants, we found that when learners received the learning material closer to the plants, they described to be more focused on them, feel more surrounded by the material, more motivated and showed different engagement with the learning material. The close availability of the plants was used to compare them to the virtual material.

Contributions: The study provides first insights into potentially positive effects of positioning virtual information in AR close to corresponding physical objects. Future studies can build on this exploratory pilot study.

1 Spatial integration in AR learning environments

In multimedia learning, the placement of content including the spatial integration of multiple external representations is of general interest. This is, for example, described in the spatial contiguity principle (Mayer, 2020) and the split-attention effect (Ayres & Sweller, 2014), which both describe the necessity to place corresponding information close to each other. While in conventional multimedia design and research this usually focuses on a combination of text and pictures, in augmented reality (AR) a layer of information is added through the combination of physical and virtual elements. The integration of information may thus be especially relevant due to a potentially rich natural context provided by the physical environment. In AR, virtual content can be positioned at a specific place inside the natural physical world, which can be leveraged for location-based instructional opportunities. Information can be placed in thematically relevant or related physical environments, for example placing information about a tree within a forest, and it can be placed in relation to corresponding physical objects, for example anchoring virtual information about a tree to that specific tree in nature. This way, a thematic connection of the surrounding physical environment to virtual elements can be leveraged to support coherent knowledge construction. This unique aspect of AR has been described as part of the AR-specific characteristic contextuality (Krüger et al., 2019). For an educational AR experience this means that instructional information can be displayed in a usually non-educational physical real-world context at a thematically relevant place and time. Contextual representations have been identified as a feature of AR that can be beneficial for conceptual learning, and the placement of information in real-world settings has been identified as beneficial for learners' affective reactions (Schneider & Radu, 2022).

AR applications have been characterized as rather place-dependent or place-independent (Dunleavy & Dede, 2014), with possible instances in between. Wetzel et al. (2011) use three categories when classifying mobile AR games concerning their dependence on the semantical location context from low to high (1. independent, 2. loosely coupled, 3. dependent) and Reid et al. (2005) similarly describe three levels of meaningfulness of the surrounding place in mediascapes (1. arbitrary linkage, 2. physicality, 3. particular location). This differentiation between an application that can be used anywhere and an application that is coupled to a certain kind of context (e.g., any tree or forest), but not an exact place (e.g., that specific tree or forest), may be helpful for the design and evaluation of educational AR experiences. Karapanos and colleagues (2012), for example, placed an access point to a narrative about a city either at a location independent of the narrative, at a location with a matching atmosphere, or at the original location with specific physical cues matching the narrative. Placing virtual information close to corresponding physical elements may especially play a role when learning material is connected to physical objects and places that cannot be found or moved somewhere else. This includes objects and places in nature which cannot themselves be designed or changed to display additional information, but which can be enriched through instructional information in an AR application. This way, educational information can be provided without intervening in the natural setting itself.

Concerning the contextual link of real world and virtual elements in location-aware applications, the importance of contextually close coupling has been emphasized in the literature on narration-based experiences (Georgiou & Kyza, 2021; Karapanos et al., 2012). Here, mainly the potential for immersion within the environment and the narrative have been highlighted. While Georgiou and Kyza (2021) also found positive effects of close coupling on learning gains in the form of reasoning, their study also focused on a narrative experience supported by physical locations. The spatial integration of virtual and physical elements can, however, also be leveraged in non-narrative educational settings. As described above, through AR, educational information can be situated in a relevant context (Dunleavy & Dede, 2014), facilitating authentic and contextualized learning experience and outcomes (Bower et al., 2014). Immersive interfaces can enhance learning through situated settings, which is in accordance with the idea that meaningful learning should take place in the context in which it will be used (Dede, 2009), as suggested on the basis of situated learning theory (see Brown et al., 1989). Placing virtual information in an authentic, relevant context may improve learning and understanding, increase enjoyment, and lead to more positive attitudes towards a topic (Harley et al., 2016; Kamarainen et al., 2013). It may also provide learners with more information in decisions concerning socio-scientific issues and increase their senses of immediacy, presence and immersion (Chang et al., 2013). Accordingly, situated learning has been identified as an important theoretical framework for AR (Bower et al., 2014; Dunleavy & Dede, 2014; Sommerauer & Müller, 2018).

1.1 Immersion and motivation in contextualized AR learning environments

While many constructs can play a role in immersive learning environments, the feelings of immersion and motivation have been identified as two central variables for learning with immersive media (see, for example, models by Dengel & Mägdefrau, 2018; Makransky & Petersen, 2021). These two variables can be influenced by the closeness of information positioning in an AR-based learning environment, and can in turn influence learning outcomes, as will be described in the following.

Immersion has been described as a system's characteristic (Slater & Wilbur, 1997) or a person's experience (Witmer & Singer, 1998) in virtual environments. To examine learning processes, the human experience is important, defining immersion as the subjective sense of being inside an environment, including virtual or AR-based environments (Georgiou & Kyza, 2017b). Kim (2013) describes AR-specific context immersion as an experience of awareness of the real context in its interplay with the AR-based elements. Three dimensions of context immersion described here include the time and location-based context, the object-based context, and the user-based context. When looking at how virtual, contextually relevant information can be added in AR as described above, the time and location-based context describes the influence of the general surrounding environment (e.g., the forest) and the object-based context describes the influence of the physical anchor object (e.g., a specific tree). Georgiou and Kyza (2017a) define immersion in location-based AR as a multi-level construct and propose three levels of immersion: 1) engagement, including interest, time investment and usability; 2) engrossment, including emotional attachment and focus of attention; 3) total immersion, including presence and flow. In location-based AR environments, learners with high levels of immersion displayed different learning behaviors (Georgiou & Kyza, 2017a) and had higher learning outcomes (Georgiou & Kyza, 2018). In narrative-based learning experiences, sense of presence and reasoning were increased with more closely coupled virtual and real-world elements (Georgiou & Kyza, 2021), and closer coupling increased immersion in the story and mental imagery (Karapanos et al., 2012). In general, this thus shows a potential influence of placement of AR-based information in relation to physical objects on learners' immersion and related experience.

Motivation has also been defined as a relevant variable in AR, with empirical findings usually supporting the idea that AR in education increases learner motivation (Akçayır & Akçayır, 2017; Garzón et al., 2019; Radu, 2014). Due to the often design-based nature of studies on AR in education, a model that is often employed in this context is the ARCS model by Keller (2010). The model describes the four aspects attention, relevance, confidence, and satisfaction as levels building upon each other and influenced by the design of the material. Looking at the influence of the context in AR on motivation, a recent study contextualized vocabulary learning with physical objects through an AR-based implementation (Weerasinghe et al., 2022). A real-time visualization of vocabularies embedded in the physical environment through a head-mounted display increased motivation and recall in comparison to a tablet-based implementation with a photo instead of a view of the physical context.

Furthermore, in multiple studies, a connection between immersion and motivation was found. In game-based learning, immersion had an influence on enjoyment (Liu et al., 2014), and enjoyment was positively correlated with sense of presence of AR objects and in VR environments (Sylaiou et al., 2010). Enjoyment is an important aspect of intrinsic motivation and can support persistence in task execution (Reeve, 1989). The potential influence of contextual coupling on immersion as described in section 1.1.2 might thus in turn positively influence motivation and enjoyment. In total, these insights on the influence of context and immersion on motivation suggest that placing virtual information in a relevant physical context can have a positive influence on motivation.

Increased motivation is an important factor especially in increasing engagement and effort of learners in multimedia-based learning (Mayer, 2014; Paas et al., 2005). The effort that learners invest into engaging with learning material has been connected to an active component of cognitive load, which stands in contrast to a passive component that the instructional material elicits (Klepsch & Seufert, 2021). This active component was shown to be correlated with germane cognitive load, the component of cognitive load describing processes relevant for deeper learning and schema construction (see Sweller et al., 1998 for the description of the three types of cognitive load in the cognitive load theory). Due to the suggested influence of motivation on effort that learners put into their engagement with the learning material, the increased motivation through the contextually relevant close placement of virtual information should also increase effort and thus germane cognitive load, which in turn should improve learning outcomes.

1.2 Goal and research questions

The current pilot study has an exploratory and qualitative focus, describing the implementation and reception of an AR-based application in nature which places virtual information either directly at corresponding physical objects or with a little distance to them. The study's goal is to examine the influence of positioning of contextually relevant virtual information in an AR-based experience on learners' experience and behavior, including exploratory analyses of cognitive and motivational variables. The study is aimed at providing insights into learning processes and the question of how learning with AR works though a value-added study design as suggested by Buchner and Kerres (2022), while taking place in an authentic, natural context. The research question is: "How does the closeness of placement of thematically relevant learning material in a physical context in AR-based learning environments influence learning behavior, processes and outcomes?". Based on the literature review, we propose that learners who learn with material that is placed closer to corresponding objects in the surrounding physical world are more involved with and motivated to learn about the material, which may lead to more effort and a better learning outcome. Due to the exploratory nature, no specific hypotheses but expectations are formulated and explained for different facets of the learning behavior and experience in the following.

We expect that the combination of physical and virtual elements in the AR-based learning experience has an impact on learners' behavior. When virtual information is placed more closely to corresponding physical objects, we would expect that learners more specifically connect these elements to each other and take a closer look at the physical objects due to the apparent relevance. We will explore if there are any other differences in learning behavior due to the closeness of the information. Furthermore, looking at the interaction with the application, there are textual virtual elements that provide additional descriptive information, and graphical virtual elements that either show more detailed depictive representations or show a virtual version of the physical object. We will explore if learners access the virtual information differently when the context is positioned either close to or far from them. One expectation we have is that learners who receive the virtual information directly in front of the physical object look at the virtual version of that object less often than learners who are further away from the physical objects.

The literature review showed that close coupling of virtual and physical information in AR may have a positive influence on immersion. In the current study, we thus expect that the closeness of the placement of virtual information to relevant physical objects has a positive influence on immersion. Looking at the different subfactors of AR immersion described above, we would expect that the placement especially has a positive influence on interest as part of the level of engagement, with the more specific connection to physical objects in the environment offering an increase in interest for the learning material. Further, we expect that on the level of engrossment, emotional attachment to the material is easier when the learner is closer to the physical objects, and that attention can be focused more easily on the material. We also expect that on a total immersion level, sense of presence is increased due to being more physically embedded within the material when closer to the physical objects.

Based on the literature review, we would further expect that the closeness of the placement of virtual information to relevant physical objects has a positive influence on motivation. Concerning the subfactors of motivation based on the ARCS framework described above, we would expect the placement of virtual material directly at the relevant physical objects to increase the attention due to an easier focus on the relevant material when close to the physical objects. The perceived relevance is expected to be increased due to the material being embedded directly into a relevant context, so that the field of application is apparent. Confidence is expected to be increased when virtual information is closer to the physical objects due to the possibility to compare information with reality. Satisfaction is expected to be higher for closer placement due to expected feelings of enjoyment as described above.

As described above, motivation is expected to have an influence on learners' effort and thus deeper learning processes. Due to the suggested positive effect on motivation, we also expect that the closeness of the placement has a positive influence on cognitive processes and learning outcomes.

2 Materials and Methods

2.1 Design and participants

In a between-subjects design with two groups, $N = 19$ participants took part. The study took place outside in a natural setting and the participants were recruited through the study course's recruitment channels. The manipulated variable was the spatial closeness to the context, with one group receiving virtual information directly at the relevant physical objects (near contextual placement, Figure 1 left) and the other group receiving virtual information in the general environment including the relevant physical objects, but not directly at the objects (far contextual placement, Figure 1 right). Dependent variables are behavior during the learning task, immersion, motivation, and learning outcomes. The participants were not distributed randomly into the two conditions, but study slots were matched between the two groups as equally as possible based on weekdays, start times, weather, and temperature. The N = 19 participants were aged between 18 and 33 (data from n = 13, M = 22.62, SD = 4.52), with 3 male and 16 female. All of them were students in the study courses Applied Cognitive and Media Science or Psychology at the University of Duisburg-Essen, which are unrelated to the topic of the learning material. For all 19 participants, tracking data for the usage of the application are available, and interview data are available for n = 18 of the participants. Due to data loss as a result of a cyberattack at the university, questionnaire and test data are only available for n = 8 of the participants (4 data sets for each condition) and will be considered completely exploratorily, only complementary to the other data, and will not be interpreted further in the discussion section. Sample description based on these n = 8 participants shows low prior knowledge in a test for naming the ten plants that are part of the learning material (M = 1.21, SD = 0.19; out of 10 points). Knowledge beliefs on a scale from 1 (low) to 5 (high) were also quite low (M = 1.46, SD = 0.43), with medium task expectancies (M = 2.44, SD = 0.62), and medium perceived value (M = 2.65, SD = 0.51). Answering in a five-point response format "never" (1), "rarely" (2), "now and then" (3), "often" (4), "regularly" (5), all participants indicated to have used general mobile applications regularly (M = 5.00, SD = 0.00, never: 0%), and mobile learning applications had been used quite often (M = 3.50, SD = 1.20, never: 13%). The participants had not used AR applications (M = 1.75, SD = 0.71, never: 38%) or specifically AR learning applications (M = 1.63, SD = 0.74, never: 50%) very often. This study with the ID psychmeth_2022_AR24_18 was conducted in accordance with ethical guidelines and the protocol was approved by the ethics committee of the Department of Computer Science and Applied Cognitive Science at the University of Duisburg-Essen (ethics vote ID: 2208PFKJ7141). All participants provided their informed consent by signing a consent form before starting their participation.

Figure 1. Scanning markers in the study in the near condition at the trees (left) and in the far condition on the stones (right).

2.2 *Materials*

2.2.1 *AR Application and interaction tracking*

The tablet-based AR application used was created for this study. With this application on a 10.5-inch tablet, learners could scan ten different AR markers which were pictures of plants like trees or bushes. When scanning the pictures, additional virtual information was added as an overlay, including four smaller pictures of the bark/trunk, fruit, blossom, and leaves of that plant, and two additional pictures relevant for the plant (see Figure 2 top left). When clicking on the pictures, textual information sometimes with additional pictures for more detail was shown next to the respective overlay picture (see Figure 2 top right and bottom left). Only one text panel was visible at a time. Learners could switch between this picture-and-text view and a 3D model of the respective plant through a checkmark in the top left corner (see Figure 2 bottom left). In the texts about the plants, different botanical terms or expectedly unfamiliar terms were used. All underlined/italic words could be clicked on in the text panels, showing a glossary entry about that specific word or area of information (see Figure 3 left). The glossary panels could also be accessed through a word list that could be opened in the top right corner (see Figure 3 right) and were the same for all plants, offering some more general information about plants.

Figure 2. Interface of AR application when scanning a marker picture with images view activated (top left), with one image clicked showing additional text (top right; bottom left), and with 3D model view activated (bottom right).

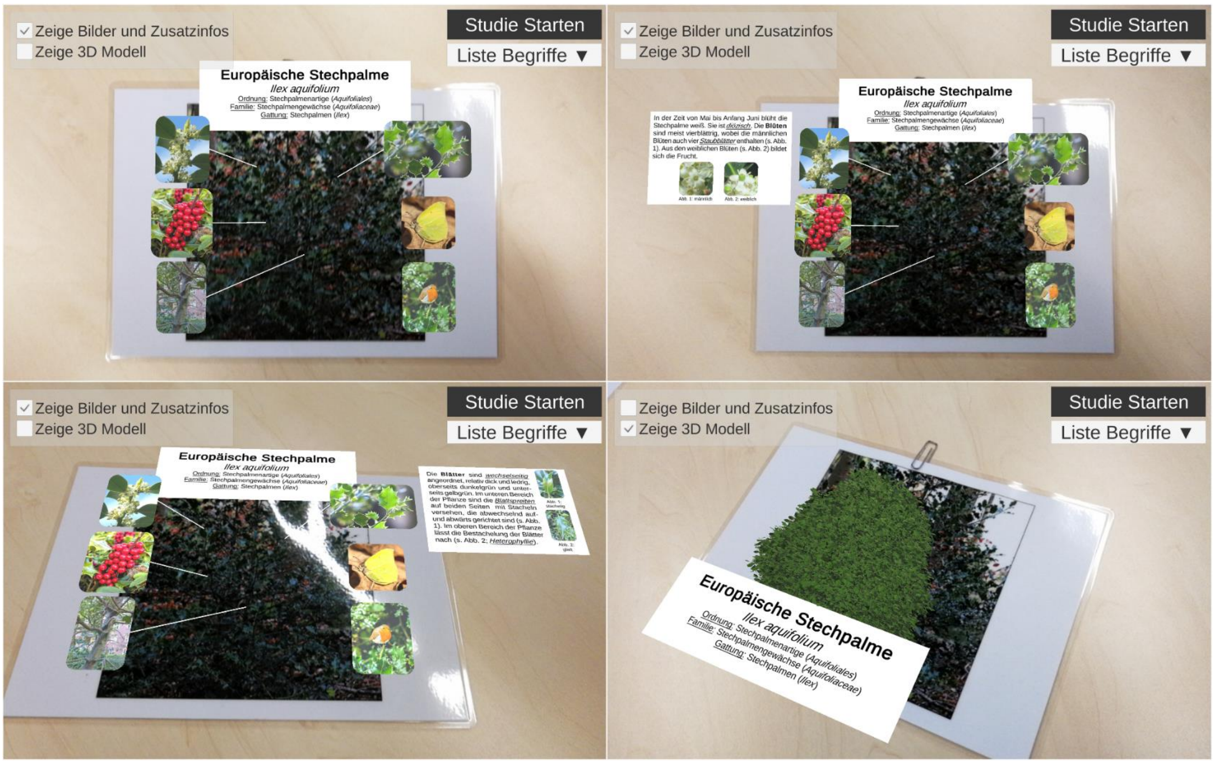


Figure 3. Glossary entries in AR application with exemplary entry on the structure of the blossom (left) and the list of all glossary entries to choose from (right).

In both conditions, the tablet-based application was exactly the same, but the set-up of the markers around the lawn differed. The markers were either pinned to the respective plants growing around the lawn (see Figure 4; near condition) or they were pinned to rocks laying in front of the lawn in the same order as the respective plants grew around the lawn (see Figure 5; far condition). Learners in the near condition were told that the AR markers were pinned to the corresponding plants, while learners in the far condition were told that the corresponding plants also grew around the lawn, without information about their exact placement. In both conditions, the learners started at the right side of the area with European holly, which was used by the investigator to explain the application. To the left, the AR markers were in this order attached to the plants common hazel, silver lime, European ivy, common oak, snowberry, sycamore maple, common hornbeam, field maple, and European yew, and this order was also followed for the placement of the AR markers on the stones.

Figure 4. Examples of placement of markers directly at trees in the near condition with markers circled in blue in the bottom two pictures.

Figure 5. Examples of placement of markers away from trees on stones in the far condition with markers circled in blue in the bottom two pictures.

All interaction with the application was automatically tracked during the learning phase, including the scanning of the AR markers showing the pictures, the activation and deactivation of the text panels, the activation and deactivation of the 3D model, and the activation and deactivation of glossary panels. The tracking data was used to analyze learners' behavior during the learning phase.

2.2.2 Interview questions

At the end of the study, participants answered open questions in a short interview. The answers were written down in a summarized form by the researcher. To gain insights into the participants' experience of their learning behavior, they were asked what influence they thought the setup (i.e. standing in front of the plants in the near condition and standing in general proximity to the plants in the far condition) had on their learning behavior. Further, they were asked what they think they would have done differently if they had been in the other setup (i.e. not standing in front of the plants and standing in front of the plants respectively). They were asked if they would then have felt more or less surrounded by the material (immersion), if they would have dealt with the material more or less intensively (engagement, effort), and if they would have been more or less motivated to engage with the material (motivation). They were also asked for further comments after that. If the researcher thought that an answer could be more specific, they asked for further elaboration.

2.2.3 Questionnaires

For the exploratory quantitative evaluations, different questionnaires were applied after the learning phase. As mentioned before, only eight datasets of the questionnaire data are available due to data loss through a cyberattack at the university. Concerning immersion, interest (four items), usability (four items), emotional attachment (three items), focus of attention (three items), presence (four items) and flow (three items) were measured with self-translated subscales of the ARI questionnaire by Georgiou and Kyza (2017a). The seven-point response format ranged from "totally disagree" (1) to "totally agree" (7) and means were calculated per subscale. McDonald's omega was excellent for interest ($\omega = 0.92$), good for flow ($\omega = 0.83$), acceptable for presence ($\omega = 0.72$) but questionable for usability ($\omega = 0.69$), emotional attachment ($\omega = 0.64$), and focus of attention ($\omega = 0.68$). Data from these three subscales should be interpreted with care.

Motivation was measured through a self-translated version of the reduced instructional materials motivation survey (RIMMS; Loorbach et al., 2015), including the subscales attention, relevance, confidence, and satisfaction based on the ARCS model (Keller, 2010) with three items each. The five-point response format included the answer options "not true" (1), "slightly true" (2), "moderately true" (3), "mostly true" (4) and "very true" (5). Means were calculated per subscale. McDonald's omega was excellent for attention ($\omega = 0.94$), confidence ($\omega = 0.95$), and satisfaction ($\omega = 0.95$), and good for relevance ($\omega = 0.85$).

To measure cognitive elaboration processes quantitatively, the germane cognitive load subscale (three items) of the cognitive load questionnaire by Klepsch et al. (2017) was used. The seven-point response format ranged from "not at all true" (1) to "completely true" (7) and means were calculated for the subscale. McDonald's omega was good for the germane cognitive load subscale ($\omega = 0.88$).

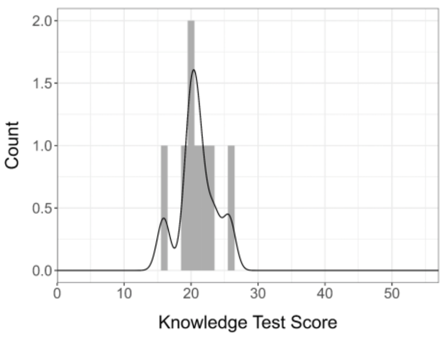
A self-developed questionnaire was administered to ask participants about where they had looked during the learning phase. Firstly, they were asked about the proportion of time they had looked at the tablet in comparison to the environment. Secondly, focusing on the time they had looked at the tablet, they were asked about the proportion of time they had looked at the virtual elements in comparison to the camera view of the physical world. Thirdly, focusing on the time they had looked at the camera view of the tablet, they were asked about the proportion of time they had looked at the marker picture in comparison to the surrounding environment. All proportions were determined on slider scales from 0 to 100 and participants were asked to elaborate on each of their answers in open text fields.

2.2.4 Prior knowledge and knowledge tests

Prior knowledge was tested concerning the specific plants included in the learning material. The pictures of the plants used as AR markers in the learning phase and the picture of the leaves used in the AR application were shown for each of the ten plants included in the learning material. The participants were asked to choose the plants' names from a drop-down list of 27 names of middle-European trees and bushes. This way recognition of the trees without knowing their specific spelling was possible. A total of 10 points could be achieved.

Learning outcomes were measured through a knowledge test (30 questions) which included textual and pictorial elements in the questions. Questions were formulated systematically on the different information provided in the AR application, with questions on the individual plants, questions on more overarching concepts and questions on the information from glossary entries. To reach a comprehensive picture of the learners' knowledge, different types of questions were used. Nine true/false statement questions, three multiple choice questions, nine multiple response questions, eight open questions with between two and ten open text fields (in total 60 open text fields) and one ordering question were included in the knowledge test. All open text fields needed only a one-word answer, naming for example a plant or plant component for the possibility of an unambiguous scoring of the responses. The questions included simple recognition questions in most multiple choice and response questions, recall questions in most open questions, and even transfer of the general information about the structure of plant parts on pictures of real plants. 20 items included a textual question and textual answer options, three items included a textual question and pictorial answer options, and seven items included a picture-based question and textual answer options. Scoring differed for the question types, with a total of 57 achievable points. In the part of the sample for which data on the knowledge test were available (n = 8), the knowledge test scores show that the test was in general difficult, with maximum of 25.68 and a minimum of 15.90 of 57 possible points scored in the sample (M = 20.92; SD = 2.88). See Figure 6 for the distribution of knowledge test scores of these participants.

Figure 6. Distribution of knowledge test scores for n = 8 participants.



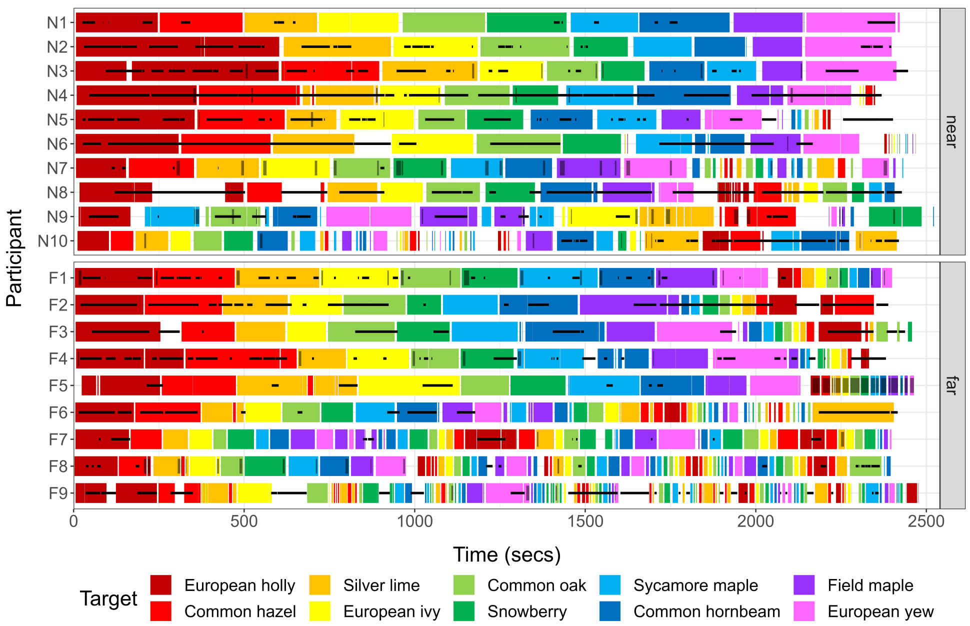
2.3 Procedure

The study took place outside on a lawn on the university campus grounds where the ten trees that were part of the learning material were growing. First, the researcher welcomed the participants to the study, briefed them on the study content and asked for their consent to participate and save their data. Then, the participants answered the prior knowledge test and determined their knowledge belief, task expectancy and subjective value concerning knowledge about Middle-European trees and bushes. Afterwards, the learning phase with the AR application on the tablet took place. The researcher gave the participants an introduction to the functions of the application and instructed them to look at the different plants, comparing them and looking for similarities and differences as a preparation for the knowledge test. The markers were set up for either the near or the far condition and participants had 40 minutes to walk from marker to marker and learn about all ten plants. All interaction with the application was tracked. Afterwards, the participants answered the questionnaires on cognitive load, motivation, and immersion. Due to the temperature below 10 °C and to have a short break before the knowledge test, afterwards the researcher and the participants moved to a room inside where the questions on the use of the application (which required open text answers), the knowledge test and the concluding interview were administered. In the end, demographic data were requested, and the participants were debriefed.

3 Results

The results focus on a first exploratory, qualitative pilot evaluation of how the positioning of contextually relevant, corresponding information has an influence on learning behavior, immersion, motivation, and learning outcomes. The qualitative and quantified data of the application usage tracked by the application (N = 19) was used to analyze learning behavior. Responses from the interviews (n = 18; one interview could not be administered due to technical issues) were used for a qualitative analysis of learning behavior, immersion, motivation and learning, and responses from the different questionnaires and tests (n = 8) were used for further descriptive evaluations. The results section is split by variable, integrating quantitative and qualitative descriptive results per concept. In Figure 7, an overview of the behavior of all 19 participants can be seen, which will be referenced in the following sections. In the figure, participants are sorted by the number of marker scanning events, from few to many, and participant IDs are assigned based on that order. Participant IDs in the near condition start with an N, participant IDs in the far condition start with an F.

Figure 7. Order and duration of scanning AR markers per participant, color per plant. Grey overlays show when the 3D model view is activated, and black lines show when glossary panels are activated. Split by group and sorted by number of target scan events (few to many) to determine participant IDs (IDs start with N in the near condition and with F in the far condition).



3.1 Learning behavior

The learning behavior is analyzed based on three types of data: interview responses, application interaction tracking, and questionnaire data. Table 1 lists all topics mentioned in the interviews concerning learning behavior. In the near condition, four participants mentioned that they compared the virtual information from the tablet and the physical plants, while in the far condition the environment was described to have no influence on more than half of the participants' behavior. Only three participants described that they looked for the plants in the environment but were not able to find (all of) them. Three participants in the far condition would have looked at the physical plants closer if positioned more closely, and another three would have compared or related the physical and virtual elements. This is thus in accordance with our expectations that closer physical objects lead learners to connect virtual and corresponding physical elements. Two participants in the near condition described that they touched the plants, while three participants in the far condition would have touched the plants if closer to them. While one person in the near condition described that less walking would have been necessary when the information would not have been placed at the plants, one person in the far condition would have walked around the plant if closer to it. In general, this shows the influence that the spatial closeness of virtual information to corresponding physical objects has on learning behavior, describing how learners pay more attention to the physical objects, even experience them with more senses like touch, and compare them to the virtual elements. As part of the questionnaire, learners were asked how much they looked at which aspects of the learning experience and material. The available data of the $n = 8$ participants show a big descriptive difference between the conditions in how much they state to have looked at the tablet in comparison to their surroundings. Participants in the near condition on average looked at the tablet 83% of the time ($M_n = 83.00$, $SD_n = 10.92$), with a minimum of 67% and a maximum of 90%. Participants in the far condition, however, looked at the tablet more often, with an average of 98% ($M_f = 98.00$, $SD_f = 3.37$), with a minimum of 93% and a maximum of 100%. In the open questions asking to elaborate their responses, the participants in the near condition describe that they mainly looked at the tablet (which can be seen in the high mean value), but all four participants also describe that they sometimes looked at the plants in the environment, for example to compare the information from the tablet to the real plant. One participant even described touching the plants, which is in accordance with the responses in the interview. The participants in the far condition, on the other hand, describe that they (nearly) only looked at the tablet (which can be seen in the very high mean value), although one participant stated that this may have been different if the plants would have been easier to recognize in the environment, and if more time was available. This supports the statements from the interview, describing that close information reception brought attentional focus to the environment, which was not observed very often when information was received further away from the plants.

Table 1. Learning behavior described in interview responses with participant IDs split by condition.

Topics mentioned in interview concerning learning behavior	
near	**far**
▪ Compared virtual information from tablet and physical plants (N1, N3, N5, N10) ▪ Touched plants (N6, N9) ▪ Also focused on looking at plants because they are there (N5) ▪ Focused on reading information, not so much looking at physical plants, but looked around to see where the plants are (N8) ▪ Read all glossary entries at first plant (N8) *If material further away from plants:* ▪ Would have looked at virtual pictures longer (N6), read virtual texts in more detail (N5) ▪ Less walking necessary (N8) ▪ Would have searched for plants (N4)	▪ No influence of environment on behavior, only looking at tablet (F2, F3, F5, F6, F7, F8) ▪ Looking for plants in environment but did not find them (F1, F4), tried to match plants to information (F9) *If material closer to plants:* ▪ Would have looked at physical plant closer (F1, F2, F4), more than looking at marker pictures (F3) ▪ Would have compared/related physical and virtual elements (F1, F2, F8) ▪ Would have touched (F1, F4, F6) ▪ Would have walked around the plant (F6) ▪ No influence on behavior (F7)

In the AR application, all data of learners interacting with the AR markers and the virtual interface were tracked. To see if there were systematic differences in behavior between the near and the far condition, the order and duration of interacting with the different AR markers is analyzed, which is visualized in Figure 7. Furthermore, patterns of individual participants are described and connected to their interview responses in Table 1. In general, all participants looked at all plants at least once, except for one participant in the far condition (F2) who did not look at the European yew and one participant in the near condition (N2) who did not look at the common hazel. In the interview, F2 stated that they only looked at the tablet, which may have caused them to miss the last plant. N2 took part in 2 °C temperature, which may have had an influence on their behavior, although this cannot be confirmed because their interview could not take place due to technical issues. Most participants follow the same order in the beginning, starting from European holly, which was also used by the investigator to explain the application, continuing on to common hazel, silver lime, European ivy, common oak, snowberry, sycamore maple, common hornbeam, field maple, and European yew, which is the order in which the plants were growing and thus the AR markers in both conditions were setup from right to left. F8 and F9 are the two participants with the most marker-scan events and are in the far condition. As can be seen in Figure 7, F8 with 70 marker scans follows a more systematic pattern from right to left and back, while F9 with 127 marker scans seems to start in that order, but then jumps from one plan to another in no apparent order. In the interviews, F8 stated that they only looked at the tablet, with no influence of the environment on their behavior, which may explain why they were guided by the spatial setup. F9, on the other hand, described that they tried to match the surrounding plants to the information, so that their behavior may have been influenced by the material and their perception of the plants instead of the spatial setup alone. The only participant that clearly follows a different, individual order than proposed by the setup from the beginning is N9 with 51 and thus the most marker scans in the near condition. N9 also stated in the interview that they touched the plants, which suggests that they were more guided by the potentials of what they could do than the setup of the experiment. After the first block of looking at all the plants once, different interaction patterns can be seen. Participants N1, N2, N3, N4 (mostly) finished after looking at everything once. All these participants were in the near condition and participants N1 and N3 also stated in the interview that they compared virtual information from tablet and physical plants, so that they may have stayed at the individual plants as long as possible. Participants N6, N8, F1, F5, and F8 (first half) repeated the same or a similar order after looking at all plants once. While N6 stated in the interview that they touched plants, N8 stated that they focused on reading the information without looking at the plants so much. F5 and F8 described that the environment had no influence and they only looked at the tablet. Participants N5, F2, F3, F4, F6, and F7, went to all plants in the opposite order after the first round. N5 in the near condition described to compare tablet and physical plants, and F4 in the far condition stated that they looked for the surrounding plants. In contrast, F3, F6, F7 and F8 (second half), which were all in the far condition, stated that the environment had no influence on their behavior, with all of them going back and forth along the plants multiple times and thus fully following the spatial setup. Finally, N7, N10, and F9 jumped between different trees without an apparent order after viewing all plants in order once. N10 stated in the interview that they compared virtual and physical information. There thus seem to be different processes for looking at the plants, with the clearest differences between groups when it comes to looking at every plant only once, happening only in the near condition, and going back and forth instead of starting from the beginning again mainly in the far condition. We also examined the quantified tracking data, including the number, the total duration and the average duration of events tracked in the application (see Table 2 for an overview of data concerning some variables). Looking at the activation of targets, descriptively targets were scanned on average about 22 times less often in the near than the far condition (M_n = 29.70, SD_n = 14.98; M_f = 52.00, SD_f = 32.68), although the total duration of target activation was quite similar in the two conditions with only a 7 second difference (M_n = 2074.23, SD_n = 202.38; M_f = 2081.11, SD_f = 206.59), making the average duration of target activation descriptively about 35 seconds longer for the near than the far condition (M_n = 89.25, SD_n = 48.52; M_f = 54.68, SD_f = 30.51). When looking more specifically at the activation of the text panels attached to the targets, which delivers most of the information about the plants, it can be seen that although descriptively about 45 less activations of text panels took place in the near than the far condition (M_n = 391.20, SD_n = 187.47; M_f = 436.33, SD_f = 202.34), participants in the near condition had text panels activated for about 88 seconds longer than participants in the far condition (M_n = 2074.23, SD_n = 202.38; M_f = 2081.11, SD_f = 206.59). This describes a higher number of switches and rescans of the AR markers in the far condition, with more clicks to view text panels, but less time spent with open text panels that could be read.

When examining the interaction with the virtual elements, there have also been different courses of action for the participants. Concerning the 3D models, there are six participants who did not look at any models, three in the near condition (N1, N2, N8) and three in the far condition (F2, F3, F6).

Participant F6 from this group stated in the interview that they would have walked around the physical plant if closer to it, but they seemingly did not use the opportunity to view the 3D model. Eight participants only looked at some 3D models. Only three participants which are all in the far condition, looked at all 3D models (F1, F5, F8). These three participants had very different patterns of looking at the 3D models, as can be seen in Figure 7. F1 activated the 3D model at the beginning of looking at each plant, and also stated in the interview that they looked for the plants in the environment. They may have tried to support the search process with viewing the 3D models. F8, on the other hand, activated the 3D model view at the end of looking at each plant and stated in the interview that they only looked at the tablet but would have compared physical and virtual elements. F5 very differently looked at all plants for some time following the order from right to left, and afterwards again looked at all plants in the same order with the 3D models view activated. Concerning the 3D model, we again examined the quantified tracking data shown in Table 2. We expected that the 3D model would be viewed less often in the near condition, because the plants are easily viewable in the physical environment. In comparison, descriptively the 3D model is viewed only about one time less often in the near than the far condition (M_n = 6.90, SD_n = 7.69; M_f = 8.22, SD_f = 8.76), but the duration of viewing is on average about 16 seconds shorter for the near than the far condition (M_n = 35.84, SD_n = 51.51; M_f = 51.78, SD_f = 73.08), thus supporting the expectations.

We also examined the interaction with the purely virtual elements that are not attached to the AR marker at all, namely the glossary entries. In Figure 7, it can be seen that most participants had the glossary entries activated while scanning the AR markers, but a few participants used breaks between scanning AR markers to look at the glossary entries, for example N8 and F9. N8 stated in the interview that they read all glossary entries at the first plant, which is clearly visible in Figure 7. It is also visible that most participants only opened the entries sporadically, but a few participants had the glossary entries open for a long time, for example N6 and N4. From a quantitative perspective (see also Table 2), descriptively glossary panels were activated about 3 times more often in the near condition than the far condition (M_n = 29.00, SD_n = 18.25; M_f = 26.22, SD_f = 13.23), were activated for a total of 321 seconds (> 5 minutes) longer on average (M_n = 934.65, SD_n = 644.69; M_f = 613.67, SD_f = 395.79), and the average activation duration was also about 8 seconds longer (M_n = 34.78, SD_n = 29.53; M_f = 26.52, SD_f = 16.27). Again, the interaction with the virtual material was very different for participants, with the 3D model activated more consistently and for a longer time in the far condition, as expected, and the glossary panels activated for a longer time in the near condition.

Table 2. Descriptives of count, total duration, and average duration of events as part of the tracked interaction with the AR application split by condition.

Tracked events	observed range	near $n = 10$	far $n = 9$	Mean diff.	Cohen's d
		M (SD) [a]	M (SD) [a]		
Target scanned					
Count	13 - 127	29.70 (14.97)	**52.00 (32.68)**	−22.30	−0.89
Total duration (sec)	1597 - 2331	2074.23 (202.38)	**2081.11 (206.59)**	−6.88	−0.03
Average dur. (sec)	13 - 179	**89.25 (48.52)**	54.68 (30.51)	34.57	0.84
Text panel activated					
Count	201 - 789	391.20 (187.47)	**436.33 (202.34)**	−45.13	−0.23
Total duration (sec)	4609 - 6932	**5962.04 (636.17)**	5873.79 (557.98)	88.25	0.15
Average dur. (sec)	6 - 32	**18.58 (8.58)**	16.50 (7.76)	2.08	0.25
3D model activated					
Count	0 - 24	6.90 (7.69)	**8.22 (8.76)**	−1.32	−0.16
Total duration (sec)	0 - 220	35.84 (51.51)	**51.78 (73.08)**	−15.94	−0.25
Average dur. (sec)	3 - 13	4.27 (1.56)	**5.57 (3.70)**	−1.30	−0.47
Glossary panels activated					
Count	6 - 61	**29.00 (18.25)**	26.22 (13.23)	2.78	0.17
Total duration (sec)	37 - 2097	**934.65 (644.69)**	613.67 (395.79)	320.98	0.58
Average dur. (sec)	6 - 109	**34.78 (29.53)**	26.52 (16.27)	8.26	0.34

[a.] Higher mean per subscale in **bold**

3.2 *Immersion*

Immersion is analyzed based on two types of data, interview responses and questionnaire items. Table 3 lists all topics mentioned in the interviews concerning immersion. In the interview, participants in the near condition described that they felt surrounded by the material through direct contact with a plant and felt real in the context, which can be attributed to feelings of presence. Furthermore, one person who said that they are usually distracted very easily by their surroundings found themselves not being distracted and very concentrated during the learning phase, which shows a potential support of a focus of attention towards the material. One participant stated that they would have felt less surrounded if not that close to the plants. However, two participants also said that it would not have made that much of a difference, with one participant saying that they mainly looked at the tablet anyways. In the far condition, one participant stated that they felt more within the situation in which the knowledge would have to be applied due to receiving the information in the general vicinity of the plants, although they also would have felt more surrounded by the environment if closer to the plants, as did seven other participants. They mentioned a closer connection to the physical plants in real life as a potential reason for this. One participant said that there would not have been that much of a difference. Concerning the subconstruct of usability, the participants described issues with the application in both conditions, like difficulties with reading the material due to its attachment to the AR markers and the inability to zoom into the virtual material. Some usability issues mentioned also had to do with the outside placement of the study, including the movement of the AR marker in the wind and the placement of markers close to the ground (e.g., lying in a field of ivy) mentioned in the near condition, and the cold temperature mentioned by a participant in the far condition.

Table 3. Immersion described in interview responses with participant IDs split by condition.

Topics mentioned in interview concerning immersion	
near	**far**
▪ Feeling more surrounded (N3) through direct contact with plant (N8) ▪ Feeling real (N5) ▪ Other people did not distract (N8), very concentrated during learning phase (N4) ▪ Usability: material hard to look at for reading (N9), no zooming possible (N6, N9), movement marker due to wind (N8, N9), squatting for marker scanning (N8) *If material further away from plants:* ▪ Would have felt less surrounded (N5) ▪ Would not have been different (N6), mainly looked at tablet anyways (N3)	▪ Felt more within the application situation (F4) ▪ Pictures small (F9) ▪ Usability: material hard to look at for reading (F4, F1, F9), no zooming possible (F6), cold temperature (F3) *If material closer to plants:* ▪ Would have felt more surrounded by the material (F2, F3, F6, F5, F1, F8, F9), if embedded in nature (F4) ▪ Closer connection to physical plant in real life (F8, F9) ▪ Not so big of a difference (F7)

For further exploration of quantitative data, we examined the questionnaire data for the variable immersion for n = 8 participants. It can be seen in Table 6 that the ARI constructs with possible ranges from 1 to 7 show a descriptively higher score for the near condition for interest (M_n = 4.88, SD_n = 1.01; M_f = 3.69, SD_f = 1.59), emotional attachment (M_n = 3.58, SD_n = 1.17; M_f = 2.33, SD_f = 0.67), and presence (M_n = 3.25, SD_n = 1.40; M_f = 2.19, SD_f = 1.09). Usability (M_n = 5.94, SD_n = 0.55; M_f = 5.88, SD_f = 1.09) was very similar in the two conditions, and focus of attention (M_n = 4.25, SD_n = 1.29; M_f = 4.25, SD_f = 0.88) was the same. Flow was descriptively even lower in the near than the far condition (M_n = 3.50, SD_n = 1.73; M_f = 4.42, SD_f = 2.04). The results on presence are in accordance with the interview responses, showing that participants felt or would have expected to feel more surrounded by and present in the natural environment when closer to the plants. The scores on focus of attention are not in accordance with the interview responses, but the participants were not asked for a statement on this and only two participants in the near condition freely stated that they were not distracted by the environment, including people walking by. Usability issues have been reported in both conditions, which is supported by the similar quantitative results. In total, the results of the interview responses suggest a higher immersion when learning with material closer to the corresponding plants, although subconstructs may show different patterns.

3.3 Motivation

Motivation is also analyzed based on two types of data, interview responses and questionnaire items. Table 4 lists all topics mentioned in the interviews concerning motivation. One participant in the near condition described that they were motivated by directly identifying plants and thus confirming learning success, which can be connected to the factor of confidence, and another stated that they were motivated through seeing the plants in reality, which may have to do with perceived relevance. Participants in the far condition said that they would have been more motivated if closer to the plants and that the situation would have been more interesting and exciting. This was supported by participants in the near condition, describing that they would have been less motivated if further away from the plant, with one specifically mentioning that they would have been less motivated to look for the corresponding plant. Still, individual participants said that their motivation would not have change. In both conditions, participants described that the learning experience was fun, thus showing satisfaction in both conditions.

Table 4. Motivation described in interview responses with participant IDs split by condition.

Topics mentioned in interview concerning motivation	
near	**far**
▪ Was fun, interesting (N4) ▪ Motivating by directly identifying plants, seeing success (N8) ▪ Motivation when seen in reality (N3) *If material further away from plants:* ▪ Would have been less motivated (N1, N6, N8, N7, N9, N10) to look for the plant (N10) ▪ No difference (N5)	▪ It was fun (F2, F4) ▪ Quite motivated through the tablet application and walking around (F9) *If material closer to plants:* ▪ Would have been more motivated (F3, F6, F5, F1) ▪ Would have been more interesting (F2) and exciting (F8) ▪ No (big) difference (F4, F7)

For further exploration of quantitative data, we looked at the questionnaire data for the different levels of motivation based on the ARCS model for $n = 8$ participants. It can be seen in Table 6 that the ARCS constructs with possible ranges from 1 to 5 show a descriptively higher score for the near condition for attention ($M_n = 3.42$, $SD_n = 1.42$; $M_f = 2.83$, $SD_f = 1.17$), and relevance ($M_n = 3.17$, $SD_n = 0.79$; $M_f = 2.42$, $SD_f = 1.10$). Confidence was very similar between the two conditions ($M_n = 2.75$, $SD_n = 1.00$; $M_f = 2.83$, $SD_f = 1.00$) and satisfaction ($M_n = 2.58$, $SD_n = 1.10$; $M_f = 2.75$, $SD_f = 1.37$), on the other hand, was descriptively lower in the near than the far condition. While the result on relevance is in accordance with what was mentioned in the interview responses concerning the higher motivation of looking at the plants in reality, the results on confidence and satisfaction do not necessarily support these. Attention was not specifically mentioned here, but the responses mentioned in the section on immersion describing that the environment did not distract from the learning material in the near condition are supported by the descriptive scores here. In total, the results of the interview responses suggest a higher motivation when learning with material that is received directly at the corresponding plants, although the different sublevels may not show a completely clear picture.

3.4 Cognitive processes and knowledge

Cognitive learning processes and outcomes, including elaboration processes and knowledge, are analyzed based on interview responses, questionnaires, and the knowledge test. Table 5 lists all topics mentioned in the interviews concerning cognitive processes and outcomes. Concerning intensity of engaging with the learning material, participants in the near condition would have engaged more intensively with the virtual material and less intensively with the physical plants if further away from the plants. One participant mentioned that less senses would have been involved because they would not have touched the plants. Many participants from the far condition supported this, saying that they would have engaged more intensively with the physical material, but less intensively with the virtual material, although one participant also stated that they would have engaged more intensively with the virtual material if closer to the corresponding plants. In contrast, three participants in the near and one participant in the far condition said that they would not have engaged more or less intensely with the learning material. Concerning knowledge construction processes, some participants described the information that they received through the learning experience. In the near condition, participants described that the physical plants were important to see as they provided additional information about the plants' looks and made comparison to the virtual elements easier. They mentioned that it was easier to connect virtual and physical elements this way and one participant even expected that they would be able to transfer the information more easily onto other real-world situations. On the other hand, the spaced-out placement of the materials stuck to the plants was described to make comparisons between the individual plants harder. Even when the virtual information was only placed in the general vicinity of the plants in the far condition, one participant described that the material was more tangible through being placed in the real world. Looking that the virtual material, it was also highlighted that it can offer additional pictorial information independent of the current state of the corresponding physical object. In the current setting, not all plants had leaves, and none was currently blossoming, so that the virtual material could be used complementarily. Participants in the near condition suggested that it would probably have been harder to learn if the information was not placed directly at the plants, while participants in the far condition supported this claim, suggesting that it would probably have been easier to learn when directly at the plants. Participants in the far condition mentioned the potentials of a closer placement in terms of easier remembering of virtual aspects, easier visual perception, and physical pictorial representations that could have replaced the virtual ones if the season permits. However, two participants in the near condition stated that learning for the knowledge test itself would probably have been easier without being at the specific place, although one qualified their statement saying that long-term learning may be increased by the placement in the environment.

Table 5. Cognitive processes and knowledge described in interview responses with participant IDs split by condition.

Topics mentioned in interview concerning information processing and knowledge	
near	**far**
▪ Real plants important to see, as they provide additional information about looks (N6), movement in the wind (N3), make comparison easier (N1, N4, N7, N10) ▪ Easier to connect virtual and physical elements (N1, N3) and to transfer onto other situations in reality (N3) ▪ Focus also on physical plants, because they were there (N5) ▪ Virtual information can show different states of plants, not just what is available at that time (N6, N9), e.g., due to seasonal changes (N4) ▪ Overwhelmed by learning material, a lot of information (N6) *If material further away from plants:* ▪ No direct comparison of physical plants and virtual elements (N1, N7, N10) ▪ Easier comparison between different plants when information is not placed at plants but can be moved and viewed directly next to each other (N9) ▪ Would have remembered detailed information less (N4) ▪ Would have looked at details more to remember them (N10) ▪ More information through virtual material (N7) ▪ Better / more efficient learning for test when not in location (N8, N9), but maybe less long-term learning (N9) ▪ Would have engaged more intensively with virtual material, because physical plants not available to look at (N6, N5, N3, N7, N10) ▪ Would have engaged less intensively with physical plants (N3), e.g., less senses due to not touching (N9) ▪ Would have put in less effort (N7) ▪ No difference in intensity of engaging (N1, N8, N4)	▪ Material more tangible when plants in surrounding environment (F9) ▪ Was very engaged and put a lot of effort into learning (F9) ▪ Surprised how little remembered for knowledge test (F2) *If material closer to plants:* ▪ Would have helped more with learning with information directly at relevant location (F9), easier to remember visual aspects (F3) ▪ Relation to physical world would have been easier (F8), would have known which plant and which picture were related (F7) ▪ Perception would have been easier due to better visual quality (F8) ▪ More pictorial representation (F2), would have looked at physical leaves, blossoms, etc., instead, if season permits (F3) ▪ Would have engaged more intensively with physical material (F4, F6, F5, F7, F8), e.g., touching (F4), direct comparison instead of 3D model (F8) ▪ Would have engaged less intensively with virtual material, because physical plants available to look at (F3, F6, F1) ▪ Would have engaged more intensively with virtual material (F5) ▪ No difference in intensity of engaging (F9)

For further exploration of quantitative data on cognitive learning processes, we examined the questionnaire data for the variable germane cognitive load for $n = 8$ participants. It can be seen in Table 6 that germane cognitive load, with a possible range of 1 to 7, is descriptively higher for the near than the far condition ($M_n = 5.08$, $SD_n = 1.87$; $M_f = 4.25$, $SD_f = 1.10$). While the results of the interview seem to suggest that learners would have balanced out how intensively they would have engaged with the different parts of the learning material based on its placement, the results concerning the germane cognitive load questionnaire suggest that engagement was more intense when receiving the material closer to the plants. This is also in accordance with the results from immersion and motivation, which seem to have been increased by closeness to the corresponding physical objects. For further exploration of quantitative data on cognitive learning outcomes, we looked at the knowledge test data for n = 8 participants. It can be seen in Table 6 that knowledge with a possible range of 0 to 57, is quite similar in both conditions (M_n = 20.95, SD_n = 1.76; M_f = 20.88, SD_f = 4.03). It is apparent, that the spread of data is wider for the far than the near condition. When looking at different question types more closely, we see open questions being more often answered correctly by learners in the near condition (M_n = 6.12, SD_n = 1.01; M_f = 5.52, SD_f = 2.66), while questions with answer options provided were more often answered correctly by learners in the far condition (M_n = 14.84, SD_n = 1.56; M_f = 15.36, SD_f = 1.73). Another distinction can be made between questions that can be answered based on the material from the individual plants or based on the material from the glossary panels. While questions about individual plants were answered more correctly by learners in the near condition (M_n = 17.28, SD_n = 2.22; M_f = 16.46, SD_f = 3.98), questions about the glossary entries were answered more correctly by learners in the far condition (M_n = 3.68, SD_n = 1.44; M_f = 4.42, SD_f = 0.82). These results are only descriptive and from a small subset of participants but suggest that learners receiving material directly at the plants may profit especially when it comes to open questions and questions about the individual plants themselves. This supports the interview responses stating that the focus was guided more towards the plants through the closer placement.

Table 6. Descriptives of variables measured as part of the post-learning phase questionnaire and test.

Measured variables	**possible range**	**near** ***n* = 4**	**far** ***n* = 4**	**Mean difference**	**Cohen's *d***
		***M* (*SD*)** [a]	***M* (*SD*)** [a]		
Immersion					
Interest	1 - 7	**4.88 (1.01)**	3.69 (1.59)	1.19	0.89
Usability	1 - 7	**5.94 (0.55)**	5.88 (1.09)	0.06	0.07
Emotional attachment	1 - 7	**3.58 (1.17)**	2.33 (0.67)	1.25	1.32
Focus of attention	1 - 7	4.25 (1.29)	4.25 (0.88)	0.00	0.00
Presence	1 - 7	**3.25 (1.40)**	2.19 (1.09)	1.06	0.85
Flow	1 - 7	3.50 (1.73)	**4.42 (2.04)**	−0.92	−0.48
Motivation					
Attention	1 - 5	**3.42 (1.42)**	2.83 (1.17)	0.58	0.45
Relevance	1 - 5	**3.17 (0.79)**	2.42 (1.10)	0.75	0.78
Confidence	1 - 5	2.75 (1.00)	**2.83 (1.00)**	−0.08	−0.08
Satisfaction	1 - 5	2.58 (1.10)	**2.75 (1.37)**	−0.17	−0.13
Germane cognitive load	1 - 7	**5.08 (1.87)**	4.25 (1.10)	0.83	0.54
Knowledge	0 - 57	**20.95 (1.76)**	20.88 (4.03)	0.08	0.02
Open questions	0 - 31	**6.12 (1.01)**	5.52 (2.66)	0.59	0.30
Closed questions	0 - 26	14.84 (1.56)	**15.36 (1.73)**	−0.52	−0.31
Plant questions	0 - 41	**17.28 (2.22)**	16.46 (3.98)	0.81	0.25
Glossary questions	0 - 16	3.68 (1.44)	**4.42 (0.82)**	−0.74	−0.63

[a.] Higher mean per subscale in **bold**

4 Discussion

The goal of the current exploratory pilot study is to examine the influence of placement of contextually relevant virtual information in an AR-based experience on learners' experience and behavior, including exploratory analyses of cognitive and motivational variables. In order to answer the research question "How does the closeness of placement of thematically relevant learning material in a physical context in AR-based learning environments influence learning behavior, processes and outcomes?", we implemented an AR-based application in nature, placing virtual information either directly at the relevant physical objects or with a little distance to them. To examine if learners who learn with material that is placed closer to corresponding objects in the surrounding physical world are more involved with and motivated to learn about the material, leading to more effort and a better learning outcome, we analyzed tracking data from N = 19, interview responses from n = 18 and questionnaire and test data from n = 8 participants.

The first type of variable we explored is learning behavior. We formulated the expectation that learners would more specifically connect virtual and physical elements to each other and take a closer look at the physical objects when they are placed closer together. Indeed, the interview data show that learners who received the material close to the physical plants also placed more attention on these plants and their connection to the virtual materials. Learners who received the material further away were also given the information that the plants could be found in their surroundings, but most did not describe to have made the effort to look for them. The ones who said that they tried to look for the plants stated that it was hard to recognize them because not all of them had leaves, so that they stopped trying. In the questionnaire data, this is also apparent in the learners' statements on how much they looked at the tablet in comparison to the environment, showing that learners further away from the plants did not look at their environment very much or at all. Both groups received the information that the plants in the material were available in their environment, but this information only impacted the learning behavior of the group in which the AR markers were placed directly at the corresponding plants. Placing the information directly at the plants made it easier for learners to recognize the corresponding plants. Furthermore, their focus was moved to those physical plants and their connection with virtual elements. While participants were instructed to look for similarities and differences between the different plants, they were not instructed to compare the virtual and physical information, which they were thus only implicitly guided to when standing in front of the plants. This suggests that the closeness to the plants provided the possibility of a bottom-up or stimulus-driven guidance of attention (see Egeth & Yantis, 1997) and thus a direct perception of an external representation with a potential to automatically use it in further processes of mental integration (see Zhang, 1997).

The interview data suggest that participants in the far condition paid less attention to the physical plants, which may have had different reasons. They may have not seen the physical plants as important parts of the learning material, so they did not take the effort to identify them. To examine if the learners in the far condition only did not look for the surrounding plants because they had struggle to identify them, it may be good to repeat the study with a setup in which it is clear which AR marker belongs to which plant, although they are placed far from each other. Just showing spatially integrated information does not automatically lead to mental integration, and a task to externally integrate elements can improve this (e.g., Bodemer et al., 2004, 2005). In the current study, it is not clear if learners indeed integrated mental representations of the physical plants and the virtual textual and pictures, although the interview responses suggest that they combined the materials. However, there was no knowledge test asking for this integration. It might be the case that the contextualized learning experience had a positive influence on learners' perception of integration without actually having a positive effect on their integrated learning of physical and virtual elements. As the instructions focused on the comparison between the different plants, this should be examined in a future study focusing on the integration of virtual and physical information as a learning objective.

In general, it seems that in their learning path participants were (first) led by the spatial setup of the plants and thus the AR markers and not by a drive to compare specific plants with each other. Only one participant had a fully individual path. After the first or more rounds of looking at all of the plants in the given order, a few participants also changed their path, although from the current data it is not clear how they determined this order. This might be interesting to explore in future studies looking at which learning paths learners choose.

The tracking data showed that targets were activated less often in the near than the far condition, with more switches between and rescans of the AR markers when walking from stone to stone instead of from plant to plant. This may have had to do with the physical setup in the available area, which involved a bigger distance between the markers at the plants than on the stones. When looking at the learning behavior in handling the virtual part of the application, we expected that learners who received the virtual information directly in front of the physical object would look at the virtual 3D model included in the application less often than learners who received the information further away. Descriptively, we indeed found this result in the quantified tracking data, which showed that participants in the near condition looked at the 3D model for a shorter time period. As already described above, the participants placed more attention on the physical plants in the near condition, some leveraging the specific physical characteristics by touching the plants. One participant in the far condition described that they would walk around plants if they received the material directly in front of the plants, which shows that a 3D model may be less necessary for that group. Concerning the interaction with the other fully virtual aspects in the AR application, it was found that participants in the near condition descriptively had the glossary entries activated more often than participants in the far condition. Maybe the closeness to the plants made the relevance to look at the more general botanical concepts clearer, although it needs to be further examined which specific mechanism may have been at work here.

Learners' feeling of being surrounded by the environment, as an approximation of immersion, seems to have been higher for learners receiving the material directly at the corresponding plants than for learners receiving the material further away. This is in accordance with empirical results showing that sense of presence is increased in learning experiences with more closely coupled virtual and physical elements (Georgiou & Kyza, 2021), and increased immersion when context was coupled with the material more closely (Karapanos et al., 2012). As AR-specific immersion has been identified as a relevant variable in AR-based learning environments (e.g., context immersion, Kim, 2013), it is important to take a closer look at how exactly AR has an influence on immersion. Through the definition by Georgiou and Kyza (2017a) including the three levels of immersion with the six subconstructs interest, usability, emotional attachment, focus of attention, presence, and flow, a more differentiated examination of immersion in AR-based learning environments is possible. In future studies, the analysis of these subconstructs should be extended beyond the small subset of participants in this study, so that a more detailed picture of the specific mechanisms that play a role in increasing immersion in closely contextualized AR-based learning settings can be gained.

Looking at learners' motivation, the interview responses suggest increased motivation when learners received instructional material directly at the corresponding plants. This is in accordance with the results by Weerasinghe et al. (2022) who found increased motivation in learners who learned vocabularies embedded in instead of separated from a physical context. In total, motivation was found to be increased in AR-based educational experiences (Akçayır & Akçayır, 2017; Garzón et al., 2019; Radu, 2014), although it is not completely clear which mechanisms are at work here. The connection to a real-world context that may especially increase perceived relevance may play an important role. However, from the results of the current study, no specific mechanisms of how motivation is influenced could be identified. In following studies, this needs to be assessed more closely, for example through a more detailed distinction between different subconstructs of motivation based on the ARCS framework (Keller, 2010). In the current study, this distinction was only made in a small subset of participants for which questionnaire data were available, with very limited insights. Further studies might want to confirm a potential pattern for different subfactors of motivation in closely contextualized AR-based learning experiences.

Concerning learners' cognitive processes, mental engagement with the learning material was described in the interview responses. Participants stated that they engaged or would have engaged more intensely with the information from the physical plants when they received the information directly at the plants, but engaged or would have engaged less intensely with the virtual material. When they received the information further away from the plants, they engaged or would have engaged more intensely with the virtual material. This supports the patterns shown for the learning behavior, focusing more on the physical plants when available. The question is, if virtual material can replace physical material, or if the information they offer are too different, which we discuss in the next section. Concerning participants' knowledge, the results are not that conclusive because the knowledge test could not be analyzed for all participants. From the interview responses, there is less a focus on how much knowledge the participants gained, but how much information was received through the AR-based learning experience. Participants with virtual information close to the plants described that the virtual and physical objects complemented each other. In the DeFT framework by Ainsworth (2006), it is described that multiple external representations can have different functions. The physical plants have the potential to offer complementary information, for example through different sensory input like touch and smell. In the current study, learners indeed touched the plants, testing the information given in the text that a leaf is soft. Furthermore, the information of how those plants grow in a real-world environment cannot necessarily be transported well through virtual representations but through looking at them in a physical natural environment. On the other hand, the virtual graphics are timeless and unchanging, so that they can be used as complementary information to the physical plants. Other functions described in the DeFT framework focus on the construction of deeper understanding through the processes of abstraction, extension, and relation. In the current setting, learners may be supported in abstracting the information they receive in the application onto the physical plant, in extending what they know from the physical plants they may have already seen before to the virtual instructional information, and in relating the virtual and the physical plants. Abstraction may be supported because the physical plants offer a perspective on the plants that is additional to the virtual text and pictures in the application. Indeed, one participant stated that they had remembered to have seen the blossoms of one of the plants in another season (extension) and one participant described that they compared and connected the virtual and physical elements (relation). It has been suggested that physical and virtual representations should be purposefully combined in instructional design (Rau, 2020), and the characteristics of AR, including contextualizing virtual learning material within physical, authentic real-world environments through specific spatial placement, can enable this.

4.1 Limitations and future studies

The study has some limitations that may limit the generalizability of the results to other situations. In general, the participants stated that the usability of the material attached to the AR markers was bad because of the placement in hard-to-reach positions (e.g., on the ground due to the European ivy growing there) and the impossibility to zoom into the pictures and text (although movement towards the AR marker was possible to view things more closely). This may have led to frustration in some participants. One participant even stated in the interview that they thought that the placement of the markers on the ground was part of the study to observe how motivated people are to look at the information. A limitation of executing a study outside is that real-world conditions may be disrupting potential study results. Because the study took place in fall, it was windy during some but not all learning phases, which means that the AR markers were moved by the wind differently for different participants. As the wind mostly influenced movement of the AR markers placed directly at the plants, the near condition may have been systematically more influenced by this. Another source of discomfort may have been the cold temperatures below 10 °C. In addition to this, executing the study in fall meant that not all the trees still had leaves and that none were blossoming. Some participants highlighted that the virtual material could add information that is independent of the season, which can help with this, but it was mainly mentioned as a negative point of participants who were not able to identify the plants or would have liked to see them with leaves.

In addition, the fact that this was a study with specific conditions may have had an influence on participants' behavior. Multiple participants mentioned that time pressure led them to not focus too much on their environment in order to get through all the information in the AR application. Especially in the far condition, this may have had an influence on their focus on the plants because they may have thought that the environment was only meant to contextualize the learning task without playing a role for the learning content. In the near condition, on the other hand, the plants may have been perceived to be part of the material that needs to be learned for the knowledge test and thus attended more. In future studies it would also be interesting to examine the learning behavior of learners without a time pressure.

In the interaction with fully virtual aspects in the AR application, it was found that most participants had the glossary entries open while scanning AR markers. This seems to show that the glossary panels did not necessarily disturb the view through the tablet camera and may have been used for explaining terms, which was their intended application. Still, some participants seem to have not realized the importance of accessing the glossary entries for more general information and did not look at them that often. In the instructions it was explained that the learners should look at similarities and differences between the different plants to learn for the knowledge test. They may have found it less relevant to look at the glossary entries, although those would have helped with comparing the different plants based on more general botanical topics like the blossom components and leaf structure.

This study was a pilot study focusing on a first exploratory evaluation of the placement of virtual information in relation to corresponding physical objects in a natural setting. The results from this study allow for the formulation of more specific hypotheses concerning immersion, motivation, cognitive processes, and knowledge. In following steps, it is important to fix some usability issues with the AR application mentioned by participants in the interviews, so that results can be fully attributed to differences in the closeness of placement in the two conditions. Furthermore, in a following study the sample size needs to be increased so that a meaningful amount of quantitative data can be collected and analyzed to answer the hypotheses.

4.2 Implications and conclusion

In general, the most important insight of this study is that a placement of virtual information close to corresponding physical objects guides learners' attention towards these objects. This may be an obvious advantage, but it is good to see it confirmed by people's descriptions of their learning behavior and perception of the relevant information in an interview after learning in an AR experience. When a learning objective of an experience includes the integration of physical and virtual elements, the virtual elements should thus be presented close to the corresponding physical elements, if possible. Furthermore, both immersion and motivation may be impacted positively by moving closer to the corresponding physical material, so that this is another potentially effect for learning experiences that include physical elements. It needs to be examined in future studies how exactly these mechanisms operate, but the current study provides interesting exploratory results upon which further research can be built.

5 References

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, *16*(3), 183–198. https://doi.org/10.1016/j.learninstruc.2006.03.001

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. https://doi.org/10.1016/j.edurev.2016.11.002

Ayres, P., & Sweller, J. (2014). The split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 206–226). Cambridge University Press.

Bodemer, D., Ploetzner, R., Bruchmüller, K., & Häcker, S. (2005). Supporting learning with interactive multimedia through active integration of representations. *Instructional Science*, *33*(1), 73–95. https://doi.org/10.1007/s11251-004-7685-z

Bodemer, D., Ploetzner, R., Feuerlein, I., & Spada, H. (2004). The active integration of information during learning with dynamic and interactive visualisations. *Learning and Instruction*, *14*(3), 325–341. https://doi.org/10.1016/j.learninstruc.2004.06.006

Bower, M., Howe, C., McCredie, N., Robinson, A., & Grover, D. (2014). Augmented reality in education—Cases, places and potentials. *Educational Media International*, *51*(1), 1–15. https://doi.org/10.1080/09523987.2014.889400

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated Cognition and the Culture of Learning. *Educational Researcher*, *18*(1), 32–42. https://doi.org/10.3102/0013189X018001032

Buchner, J., & Kerres, M. (2022). Media comparison studies dominate comparative research on augmented reality in education. *Computers & Education*, 104711. https://doi.org/10.1016/j.compedu.2022.104711

Chang, H. Y., Wu, H. K., & Hsu, Y. S. (2013). Integrating a mobile augmented reality activity to contextualize student learning of a socioscientific issue. *British Journal of Educational Technology*, *44*(3), 95–99. https://doi.org/10.1111/j.1467-8535.2012.01379.x

Dede, C. (2009). Immersive Interfaces for Engagement and Learning. *Science*, *323*(5910), 66–69. https://doi.org/10.1126/science.1167311

Dengel, A., & Mägdefrau, J. (2018). Immersive Learning Explored: Subjective and Objective Factors Influencing Learning Outcomes in Immersive Educational Virtual Environments. *2018 IEEE*

International Conference on Teaching, Assessment, and Learning for Engineering (TALE), 608–615. https://doi.org/10.1109/TALE.2018.8615281

Dunleavy, M., & Dede, C. (2014). Augmented reality teaching and learning. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of Research on Educational Communications and Technology* (4th ed., pp. 735–745). Springer New York.

Egeth, H. E., & Yantis, S. (1997). VISUAL ATTENTION: Control, Representation, and Time Course. *Annual Review of Psychology*, *48*(1), 269–297. https://doi.org/10.1146/annurev.psych.48.1.269

Garzón, J., Pavón, J., & Baldiris, S. (2019). Systematic review and meta-analysis of augmented reality in educational settings. *Virtual Reality*, *23*(4), 447–459. https://doi.org/10.1007/s10055-019-00379-9

Georgiou, Y., & Kyza, E. A. (2017a). Investigating Immersion in Relation to Students' Learning During a Collaborative Location-Based Augmented Reality Activity. In B. K. Smith, M. Borge, E. Mercier, & K. Y. Lim (Eds.), *12th International Conference on Computer Supported Collaborative Learning (CSCL) 2017, Volume 1* (p. 8). International Society of the Learning Sciences. https://doi.org/10.22318/cscl2017.57

Georgiou, Y., & Kyza, E. A. (2017b). The development and validation of the ARI questionnaire: An instrument for measuring immersion in location-based augmented reality settings. *International Journal of Human Computer Studies*, *98*(September 2016), 24–37. https://doi.org/10.1016/j.ijhcs.2016.09.014

Georgiou, Y., & Kyza, E. A. (2018). Relations between student motivation, immersion and learning outcomes in location-based augmented reality settings. *Computers in Human Behavior*, *89*, 173–181. https://doi.org/10.1016/j.chb.2018.08.011

Georgiou, Y., & Kyza, E. A. (2021). Bridging narrative and locality in mobile-based augmented reality educational activities: Effects of semantic coupling on students' immersion and learning gains. *International Journal of Human-Computer Studies*, *145*, 102546. https://doi.org/10.1016/j.ijhcs.2020.102546

Harley, J. M., Poitras, E. G., Jarrell, A., Duffy, M. C., & Lajoie, S. P. (2016). Comparing virtual and location-based augmented reality mobile learning: Emotions and learning outcomes. *Educational Technology Research and Development*, *64*(3), 359–388. https://doi.org/10.1007/s11423-015-9420-7

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 and Education*, *68*, 545–556. https://doi.org/10.1016/j.compedu.2013.02.018

Karapanos, E., Barreto, M., Nisi, V., & Niforatos, E. (2012). Does locality make a difference? Assessing the effectiveness of location-aware narratives. *Interacting with Computers*, *24*(4), 273–279. https://doi.org/10.1016/j.intcom.2012.03.005

Keller, J. M. (2010). *Motivational Design for Learning and Performance*. Springer US. https://doi.org/10.1007/978-1-4419-1250-3

Kim, M. J. (2013). A framework for context immersion in mobile augmented reality. *Automation in Construction*, *33*, 79–85. https://doi.org/10.1016/j.autcon.2012.10.020

Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01997

Klepsch, M., & Seufert, T. (2021). Making an Effort Versus Experiencing Load. *Frontiers in Education*, *6*, Article 645284. https://doi.org/10.3389/feduc.2021.645284

Krüger, J. M., Buchholz, A., & Bodemer, D. (2019). Augmented reality in education: Three unique characteristics from a user's perspective. In M. Chang, H.-J. So, L.-H. Wong, F.-Y. Yu, & J. L. Shih (Eds.), *Proceedings of the 27th International Conference on Computers in Education* (pp. 412–422). Asia-Pacific Society for Computers in Education.

Liu, L., Ip, R., Shum, A., & Wagner, C. (2014). Learning Effects of Virtual Game Worlds: An Empirical Investigation of Immersion, Enjoyment and Performance. *AMCIS 2014 Proceedings*. https://aisel.aisnet.org/amcis2014/VirtualCommunities/GeneralPresentations/3

Loorbach, N., Peters, O., Karreman, J., & Steehouder, M. (2015). Validation of the Instructional Materials Motivation Survey (IMMS) in a self-directed instructional setting aimed at working with technology: Validation of the IMMS. *British Journal of Educational Technology*, *46*(1), 204–218. https://doi.org/10.1111/bjet.12138

Makransky, G., & Petersen, G. B. (2021). The Cognitive Affective Model of Immersive Learning (CAMIL): A Theoretical Research-Based Model of Learning in Immersive Virtual Reality. *Educational Psychology Review*, *33*(3), 937–958. https://doi.org/10.1007/s10648-020-09586-2

Mayer, R. E. (2014). Incorporating motivation into multimedia learning. *Learning and Instruction*, *29*, 171–173. https://doi.org/10.1016/j.learninstruc.2013.04.003

Mayer, R. E. (2020). 9 Spatial contiguity principle. In *Multimedia Learning* (3rd ed., pp. 207–226). Cambridge University Press. https://doi.org/10.1017/9781316941355.012

Paas, F., Tuovinen, J. E., van Merriënboer, J. J. G., & Aubteen Darabi, A. (2005). A motivational perspective on the relation between mental effort and performance: Optimizing learner involvement in instruction. *Educational Technology Research and Development*, *53*(3), 25–34. https://doi.org/10.1007/BF02504795

Radu, I. (2014). Augmented reality in education: A meta-review and cross-media analysis. *Personal and Ubiquitous Computing*, *18*(6), 1533–1543. https://doi.org/10.1007/s00779-013-0747-y

Rau, M. A. (2020). Comparing Multiple Theories about Learning with Physical and Virtual Representations: Conflicting or Complementary Effects? *Educational Psychology Review*, *32*(2), 297–325. https://doi.org/10.1007/s10648-020-09517-1

Reeve, J. (1989). The interest-enjoyment distinction in intrinsic motivation. *Motivation and Emotion*, *13*(2), 83–103. https://doi.org/10.1007/BF00992956

Reid, J., Cater, K., Fleuriot, C., & Hull, R. (2005). *Experience Design Guidelines for Creating Situated Mediascapes* (HPL-2005-181; p. 71). Mobile and Media Systems Laboratory, HP Laboratories.

Schneider, B., & Radu, I. (2022). Augmented Reality in the Learning Sciences. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (3rd ed., pp. 340–361). Cambridge University Press. https://doi.org/10.1017/9781108888295.021

Slater, M., & Wilbur, S. (1997). A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, *6*(6), 603–616. https://doi.org/10.1162/pres.1997.6.6.603

Sommerauer, P., & Müller, O. (2018). Augmented reality for teaching and learning—A literature review on theoretical and empirical foundations. *Twenty-Sixth European Conference on Information Systems (ECIS2018)*. https://aisel.aisnet.org/ecis2018_rp/31

Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251–296. https://doi.org/10.1023/A:1022193728205

Sylaiou, S., Mania, K., Karoulis, A., & White, M. (2010). Exploring the relationship between presence and enjoyment in a virtual museum. *International Journal of Human Computer Studies*, *68*(5), 243–253. https://doi.org/10.1016/j.ijhcs.2009.11.002

Weerasinghe, M., Biener, V., Grubert, J., Quigley, A., Toniolo, A., Pucihar, K. Č., & Kljun, M. (2022). VocabulARy: Learning Vocabulary in AR Supported by Keyword Visualisations. *IEEE Transactions on Visualization and Computer Graphics*, *28*(11), 3748–3758. https://doi.org/10.1109/TVCG.2022.3203116

Wetzel, R., Blum, L., Broll, W., & Oppermann, L. (2011). Designing Mobile Augmented Reality Games. In B. Furht (Ed.), *Handbook of Augmented Reality* (pp. 513–539). Springer New York. https://doi.org/10.1007/978-1-4614-0064-6_25

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, *7*(3), 225–240. https://doi.org/10.1162/105474698565686

Zhang, J. (1997). The Nature of External Representations in Problem Solving. *Cognitive Science*, *21*(2), 179–217. https://doi.org/10.1207/s15516709cog2102_3

8.5 Paper 5 – Krüger & Bodemer, 2022

Krüger, J. M., & Bodemer, D. (2022). Application and investigation of multimedia design principles in augmented reality learning environments. *Information, 13*(2), Article 74. https://doi.org/10.3390/info13020074

information

Article

Application and Investigation of Multimedia Design Principles in Augmented Reality Learning Environments

Jule M. Krüger * and Daniel Bodemer

Research Methods in Psychology—Media-Based Knowledge Construction, University of Duisburg-Essen, 47057 Duisburg, Germany; bodemer@uni-due.de
* Correspondence: jule.krueger@uni-due.de

Abstract: Digital media have changed the way educational instructions are designed. Learning environments addressing different presentation modes, sensory modalities and realities have evolved, with augmented reality (AR) as one of the latest developments in which multiple aspects of all three dimensions can be united. Multimedia learning principles can generally be applied to AR scenarios that combine physical environments and virtual elements, but their AR-specific effectiveness is unclear so far. In the current paper, we describe two studies examining AR-specific occurrences of two basic multimedia learning principles: (1) the spatial contiguity principle with visual learning material, leveraging AR-specific spatiality potentials, and (2) the coherence principle with audiovisual learning material, leveraging AR-specific contextuality potentials. Both studies use video-based implementations of AR experiences combining textual and pictorial representation modes as well as virtual and physical visuals. We examine the effects of integrated and separated visual presentations of virtual and physical elements (study one, $N = 80$) in addition to the effects of the omission of or the addition of matching or non-matching sounds (study two, $N = 130$) on cognitive load, task load and knowledge. We find only few significant effects and interesting descriptive results. We discuss the results and the implementations based on theory and make suggestions for future research.

Keywords: augmented reality; education; instructional design; multimedia design principles; multimodality; technology-enhanced learning

Citation: Krüger, J.M.; Bodemer, D. Application and Investigation of Multimedia Design Principles in Augmented Reality Learning Environments. *Information* **2022**, *13*, 74. https://doi.org/10.3390/info13020074

Academic Editors: Ramon Fabregat, Jorge Bacca-Acosta and N. D. Duque-Mendez

Received: 5 January 2022
Accepted: 2 February 2022
Published: 4 February 2022

1. Introduction

The possibilities for instructional materials have changed a lot over the past decades due to technological developments in the field of digital media. While pictorial presentations can provide additional information, complementing textual elements in traditional media such as books, the addition of auditory narrations and sounds can provide further sensory input. Richard E. Mayer's cognitive theory of multimedia learning (CTML) is based on the suggestion that humans can use two processing channels when learning, the auditory–verbal and the visual–pictorial channels, which both have capacity limitations and are used in the active processing of information [1]. In more recent years, another dimension was added to the mix: combining virtual elements with physical, real-world elements through the possibilities of augmented reality (AR) technologies (e.g., [2–4]). Instructional materials can thus include different representation modes, such as text and graphics, different sensory modalities, such as visual and auditive, and different realities, such as physical and virtual. Due to these vast possibilities it is important to examine how combinations of representations can be used most effectively and efficiently to support learning. The multimedia design principles that were described by Mayer [5], based on CTML, are one framework that has often been used in this regard. The different principles describe how textual and pictorial representations should be combined in order to best facilitate learning. In the current paper, we take a closer look at how the different features of instructional design can be combined based on multimedia principles in the specific case of AR-based learning environments and how this might play a role in supporting learning processes and outcomes. We report two studies that implement and investigate two basic multimedia principles in AR scenarios: (1) the spatial contiguity principle, concerning spatially integrated physical and virtual elements, and (2) the coherence principle, with combined visual and auditory representations of contextually integrated physical and virtual elements.

1.1. Multimedia Learning

One of the most influential theories of multimedia-based learning and instruction is the cognitive theory of multimedia learning (CTML) by Mayer [1]. The theory describes cognitive information processing when learning with multi-modal, multi-representational learning material and is based on three assumptions: dual channels, limited capacity and active processing. The dual-channel assumption describes that the learning process includes two paths for processing information through sensory, working and long-term memory [1]. One path describes the auditory–verbal channel and the other the visual–pictorial channel. These channels are based on an integration of two dual-channel memory theories, namely Baddeley's model of working memory, with the distinction between visual sketchpad and phonological loop (e.g., [6]), and Paivio's dual-coding theory, with the distinction between the verbal system and the non-verbal, imagery system (e.g., [7]). The limited-capacity assumption describes that there is a limit to how much information can be processed simultaneously within one channel, such that instructional material should be designed in a way as to not overload the channels [1]. This is closely related to assumptions in cognitive load theory (CLT), which also postulates duration and capacity limitations of working memory [8,9]. The active-processing assumption describes that learning is not a passive process of information absorption but that learners need to actively process and make sense of perceived information [1]. This includes the selection, organization and integration of information, and should be supported by the instructional design. Three kinds of cognitive processing in learning postulated by CTML are extraneous, essential and generative processing. Extraneous processing describes cognitive processing that is not inherent to the learning goal and not directly induced by the content of the learning material but by its design. Essential processing describes the cognitive processing and representation of the learning content in working memory, including the active process of the selection of relevant information. Generative processing describes deeper learning processes, including the organization and integration of the learning content and prior knowledge into a coherent mental representation, which is proposed to be dependent on learners' motivation [1]. These types of processing can be mapped to the three types of cognitive load in CLT: extraneous, intrinsic and germane cognitive load. In the following sections, based on the assumptions of CTML, we will describe potential instructional differences concerning sensory modalities, the application of multimedia principles and the connection to AR characteristics and implementations.

1.2. Sensory Modalities

The sensory modality perspective in CTML distinguishes between material presented visually and material presented auditorily [1]. Based on Baddeley's model of working memory (e.g., [6]) it is described that visual and auditory materials are received through different sensory input channels and processed differently in the visuo-spatial sketchpad and the phonological loop. CTML proposes that the two channels should be leveraged in multimedia instruction, such that the independent limited resources in each channel can be used to process more information in total. The modality principle, for example, describes that learning is increased when spoken words and pictures are combined than when the same words in written form are combined with pictures [10]. Due to the digital nature of most AR-capable systems it is possible to visualize pictorial and textual information, but it is also possible to add auditory sounds and narrations. In the field of AR soundscapes, for example, spatial audio is recorded and reproduced with the goal of designing a naturalistic and authentic sound experience for AR environments, with virtual sounds adapting to listeners' movements and real-world sounds [11]. Concerning AR experiences, adding this factor of different realities to the factor of modalities leads to four potential information origins: real visual elements, virtual visual elements, real auditory elements and virtual auditory elements. This may provide more complex information, such that a systematic analysis of the learning processes and outcomes under consideration of these different forms of informational elements may be necessary. In the current paper we describe two studies concerning two potential implementations of AR experiences. Concerning modalities, the first application focuses on the combination of real and virtual visual representations without auditory elements and the study's manipulation involves visuo-spatial integration. In the second application, all four kinds of elements are implemented, with the addition of virtual auditory elements being the focus of the study's manipulation.

1.3. Multimedia Design Principles

The representation mode perspective in CTML distinguishes between material presented verbally and material presented non-verbally [1]. Based on Paivio's dual-coding theory (e.g., [7]) it is described that information is processed differently in the verbal system and the non-verbal, imagery system. The general idea behind multimedia learning is that learning from a combination of words and pictures leads to better results than learning from words alone. In order to not overload one or both information-processing channels within their limited capacity, different multimedia principles, which should be considered when designing instructions, were formulated based on many years of empirical research [1]. Two principles for decreasing extraneous cognitive processing and thus increasing resources available for essential and generative processing are the spatial contiguity principle and the coherence principle. The *spatial contiguity principle* describes that corresponding pictures and words in multimedia presentations should be presented in a visuo-spatially integrated way instead of a separated presentation [12]. It is assumed that when material is presented in a separated way, more visual searching is necessary and cognitive resources need to be used to keep the individual elements in working memory before being able to integrate them mentally. This increases extraneous processing, using up resources that are then not available for essential and generative processing. The same idea is described in the split-attention effect in CLT [13]. Many empirical studies have reported a positive effect on learning outcomes when following the principle in instructional design [14]. Specifically in AR, this principle can be followed for combinations of virtual and physical pictorial as well as textual representations, which can be displayed in an integrated way, e.g., through video or optical see-through technology in AR systems. The *coherence principle* describes that extraneous elements that might disturb learning, such as interesting but irrelevant or unnecessarily detailed visual or auditory elements, should be excluded from multimedia presentations [15]. The addition of irrelevant but interesting elements, also called seductive details, can divert learners' attention away from, lead to difficulties in organization within and mislead the integration of the relevant learning content. Adding extraneous material can thus lead to an increase in extraneous processing, depleting cognitive working memory resources that cannot be used for essential and generative processing. Specifically in AR, this extraneous material can include real and virtual elements, which can contain both visual and auditory elements that can be more or less coherent considering the learning goal. In the two potential implementations of AR learning experiences described in the current paper, we examine the spatial contiguity principle with purely visual physical and virtual material in the first study and take a look at the coherence principle with audiovisual real and virtual material in the second study.

1.4. AR Characteristics

AR as a form of combining virtual and physical information has already been implemented in diverse topical areas, suggesting a general significance for formal and informal educationally relevant fields. For example, wearable AR has been used to provide additional textual information about artists and paintings in art galleries [16], mobile AR has been used to train working memory in elderly people through an interactive serious game [17] and AR as well as virtual reality (VR) materials combined in a mixed reality (MR) application have been used to teach mathematical foundations to architecture students [18]. In various reviews of research on AR in formal and informal educational settings, its positive effects on learning outcomes, motivation, engagement, attitudes and cognitive load in comparison to non-AR implementations has been established (e.g., [2,4,19–22]). Still, more research on the effectivity and efficiency of specific design decisions in educational AR is necessary.

AR-based learning experiences have different attributes or affordances that are enabled through the features of AR technologies. Bower and colleagues (2014), for example, identified the rescaling of virtual objects and overlaying contextually relevant information as key affordances of AR [23]. Wu and colleagues (2013) described 3D-based, situated, ubiquitous and collaborative learning, the senses of immersion and presence, visualization of the invisible and the bridging of informal and formal learning in that respect [20]. Contextuality, interactivity and spatiality are three characteristics of AR experiences identified by Krüger, Buchholz and Bodemer (2019) [3]. The identification of AR-specific characteristics provides researchers and designers with another structure to conceptualize relevant research on and implementations of AR. Two of the characteristics that we will focus on in the current paper are *spatiality*, which includes the potential to place virtual objects in spatial proximity to corresponding physical objects, and *contextuality*, which includes the potential of AR to enable learning supported by virtual elements inside a relevant real-world environment [3]. While this starting point focusing on the technology's and experience's capabilities is a more technology-centered approach to multimedia design, the multimedia principles described in Section 1.2 come from a learner-centered approach that focuses on how multimedia technology can be designed to facilitate cognitive processing [24].

With a technologically enabled experience such as AR, it is important to step out of a technological perspective and connect capabilities with how they can be adapted based on what we know about human cognition. Mystakidis and colleagues executed a systematic mapping review of AR applications in the specific context of STEM learning in higher education and identified five instructional strategies and five instructional techniques often used in AR [25]. They clustered these into a taxonomy including five categories ranging from passive, teacher-centered information presentation to autonomous, student-controlled project work. In other reviews, CTML has been described as a relevant approach in AR-based learning, which should be and already is used as a basis for AR design. Sommerauer and Müller, for example, explicitly suggest on the content layer of their conceptual design framework that CTML should be used in the instructional design of AR applications and that any combination of Mayer's multimedia principles should be implemented [26]. In a review by da Silva et al. CTML is described as one of the most used theories in studies evaluating AR-based educational technology [27], which is supported by the results of a review on pedagogical approaches in AR-based education in which Garzón and colleagues found that CTML is a very popular approach that has been used in various content areas and levels of education [28]. In another systematic review with a focus on cognitive load and performance in AR-based learning research, Buchner and colleagues also describe CTML as an important theory, specifically mentioning the necessity for research examining the effect of either following or violating multimedia principles in AR [29]. Concerning the above-mentioned multimedia principles, the spatially close placement of real and virtual information enabled through spatiality is important for following the spatial contiguity principle. Spatial and temporal contiguity have specifically been mentioned as relevant principles that can be applied through AR [29]. In a study on a tablet-based AR implementation in STEM laboratory courses, the integration of real and virtual information following the spatial contiguity principle led to better acquisition of conceptual knowledge, although cognitive load was not rated differently [30]. When placing information in a relevant context or enriching a situation with contextually relevant information, which is enabled through contextuality, the above-mentioned coherence principle needs to be considered. Mayer's immersion principle states that highly immersing features, which are also apparent in AR, can be seen as similar to seductive details [31], such that the coherence principle may be very significant for AR environments.

As described above in Section 1.3, implementing multimedia principles in instructional design has an influence on learners' cognitive processing of information. In research on AR in education, cognitive load has been included in various studies, most often using subjective measures such as the NASA TLX questionnaire for data collection but very rarely using measures of the three types of cognitive load described in CTML and CLT [32]. The specification of different types of cognitive load may be valuable in research on the implementation of multimedia principles in AR in particular. In general, the results concerning cognitive load in educational AR environments are inconclusive and both the potential decrease and potential increase in load are postulated by different researchers [29]. It is apparent that the elicitation of cognitive load through the specific design of applications combining virtual and physical elements for learning purposes needs to be further examined.

The following two studies will describe potential implementations of AR experiences using the different above-mentioned variables concerning modalities, multimedia principles and AR characteristics. Study one focusses on an application based on purely visual representations implementing the spatial contiguity principle concerning spatially integrated physical and virtual elements. Study two focusses on an application based on combined visual and auditory representations implementing the coherence principle concerning contextually coherent physical and virtual elements. Because active cognitive processing is at the center of CTML, both studies examine how cognitive processing resources are used, with a focus on cognitive load, task load and the resulting knowledge.

2. Study One: Spatial Contiguity Principle

In the first study, the goal is to examine the spatial integration of virtual and physical elements. The learning material is purely visual, and it is focused on the implementation of the spatial contiguity principle in AR. In AR-specific implementations, the principle can be applied to the spatially integrated visualization of virtual elements and physical elements in the real-world environment. In the study, virtual textual information is integrated into a real pictorial environment. Based on the spatial contiguity principle, we want to examine if this implementation of the principle in particular has a positive influence on cognitive factors, including cognitive load, task load and knowledge.

We hypothesize that complying with the spatial contiguity principle through AR leads to a decrease in extraneous processing and thus in extraneous cognitive load (H1.1a) by reducing visual search processes and decreasing the time that the individual elements need to be held in working memory for mental integration. In turn, the working memory capacities that are made available can be used for generative processing, thus increasing germane cognitive load when material is integrated instead of separated (H1.1b).

Furthermore, we hypothesize that the spatial integration of the learning material has an influence on task load. We expect that through decreasing the necessity of holding individual elements in working memory for a longer time when the visualization is integrated, mental demand is decreased (H1.2a). The decreased necessity for visual search processes leads to fewer necessary eye movements and thus a decrease in physical demand (H1.2b). We furthermore expect that temporal demand is decreased when the presentation is integrated (H1.2c), because fewer search and processing steps need to be taken within the same time. We propose that easier processing of the content with the integrated presentation leads to feelings of higher performance (H1.2d) in addition to lower effort (H1.2e) and frustration (H1.2f).

Through the decrease in extraneous cognitive load and the task-load-related factors as well as the resulting increase in germane cognitive load, we would also expect increased resulting knowledge when information is spatially integrated (H1.3). All hypotheses of study one are summarized in Table 1.

Table 1. Hypotheses in study one.

Hypotheses in Study One	
H1.1a:	learning with an integrated presentation of real and virtual information leads to lower *extraneous cognitive load* than learning with a separated presentation.
H1.1b:	learning with an integrated presentation of real and virtual information leads to higher *germane cognitive load* than learning with a separated presentation.
H1.2a:	learning with an integrated presentation of real and virtual information leads to lower *mental demand* than learning with a separated presentation.
H1.2b:	learning with an integrated presentation of real and virtual information leads to lower *physical demand* than learning with a separated presentation.
H1.2c:	learning with an integrated presentation of real and virtual information leads to lower *temporal demand* than learning with a separated presentation.
H1.2d:	learning with an integrated presentation of real and virtual information leads to higher perceived *performance* than learning with a separated presentation.
H1.2e:	learning with an integrated presentation of real and virtual information leads to lower *effort* than learning with a separated presentation.
H1.2f:	learning with an integrated presentation of real and virtual information leads to lower *frustration* than learning with a separated presentation.
H1.3:	learning with an integrated presentation of real and virtual information leads to higher resulting *knowledge* than learning with a separated presentation.

2.1. Methods

In a between-subjects design with two conditions, the integration of the visual information was manipulated in a video-based simulation of a location-based informational AR application. One group received an integrated design, which resembled a see-through AR application because the virtual, textual information was placed as an overlay of the relevant pictorial information in a recorded video. The other group received a separated design, in which the information was displayed on a tablet in the video, such that the information was separated from the respective pictorial real information. Dependent variables are cognitive load, task load and the resulting knowledge.

2.1.1. Participants

The participants were reached through online platforms for participant sampling of the department and convenience sampling. Students could receive participant hours for taking part. The final dataset included N = 80 people after one outlier was filtered out based on high age. Primarily (95%) undergraduate students took part, of which most were in the study programs of applied cognitive and media science (84%) and psychology (14%), in which there are no classes related to the learning topic of the study. They were aged 17 to 33 (M = 22.21, SD = 3.14) and 20 indicated being male, 60 being female. On average, the participants did not indicate high prior knowledge beliefs concerning the focal learning topic of plants in a subjective rating (M = 1.91, SD = 0.73; 5-point response format from 1, low to 5, high). The participants on average indicated having rarely used general AR applications (M = 2.23, SD = 0.98) and AR learning applications (M = 1.69, SD = 0.96; both measured in 5-point response format, with 1—"never", 2—"rarely", 3—"now and then", 4—"often" and 5—"regularly"). They were randomly distributed into the two groups. In Table 2, the number of participants, gender, age, prior knowledge beliefs and prior usage of AR applications per condition are shown. The distribution is quite balanced for all variables. All subjects gave their informed consent for inclusion before they participated in the study. The study with the ID psychmeth_2020_AR13_29 was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the department's Ethics Committee (vote ID: 2011PFBS7216).

Table 2. Distribution of number of participants, gender, age, prior knowledge beliefs and prior usage of AR applications split by condition in study one.

Condition	n	Gender		Age	Prior Knowledge	Usage AR Applications	
						General	Learning
		Male	Female	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Integrated	39	8	31	21.72 (2.76)	1.93 (0.60)	2.31 (0.92)	1.82 (1.00)
Separated	41	12	29	22.68 (3.43)	1.89 (0.85)	2.15 (1.04)	1.56 (0.92)

2.1.2. Materials

The independent variable was manipulated by showing simulated AR experiences to the participants through two different videos during the learning phase. In both videos, plants in a botanical garden were filmed. Additionally, in both videos, textual information about different plants were added: their common and scientific name, their height, the color of their blossoms and where they usually grow. The simulated AR experience in this study thus included the video of the botanical garden in the background as the real-world environment in which the participants should imagine themselves to be in. The additional textual information about the different plants were included as the virtual elements in the AR experience which the participants should imagine viewing while walking through the garden. In the video in the integrated visualization condition, information was placed as an overlay directly in front of the plants, resembling a visualization that may be possible with an AR glasses or other see-through version of an application (left picture in Figure 1). In the video in the separated visualization condition a tablet was held by the person filming the video and the textual information was displayed on that tablet, such that it was spatially separated from the real plants (right picture in Figure 1). In both conditions, the participants were asked to imagine that they were walking around in the botanical garden and using the application in the real world themselves. The videos were around three minutes long.

Figure 1. Screenshots from the videos used in study one: integrated (AR) visualization on the **left**; separated (non-AR) visualization on the **right**.

To measure subjective prior knowledge of the participants for the sample description, three questions from the ability belief subscale of the expectancy–value questionnaire by Wigfield and Eccles [33] were used in a reframed version, inquiring knowledge beliefs with translated items adapted to the content area. The 5-point response format from 1 (low) to 5 (high) had different anchor phrases. Cronbach's alpha was good for this scale ($\alpha = 0.82$).

To measure cognitive load the questionnaire by Klepsch et al. [34] was used, measuring both extraneous and germane cognitive load with three items each. The 7-point response format ranged from "not at all true" (1, low) to "completely true" (7, high), and means were calculated for each subscale. Cronbach's alpha was acceptable for the germane cognitive load subscale (α = 0.78) but questionable for the extraneous cognitive load (α = 0.61) subscale, which we nonetheless kept as it was because of the already low number of three items.

Task load was measured through a translated, German version of the NASA TLX by Hart and Staveland [35,36] including six one-item scales for the six variables mental demand, physical demand, temporal demand, performance, effort and frustration. Most items were answered in a 21-point response format from "low" (1) to "high" (21). The performance item was measured from "good" (1) to "bad" (21), but we inverted the scores for the analysis. This provides an easier interpretability of the scores so that in the reported results high scores mean high perceived performance and low scores mean low perceived performance. Only the scores on the individual subscales were used for the analyses, no summarized version for general task load.

Learning outcomes were measured through a knowledge test that included different kinds of questions. In total, ten questions were administered in the form of multiple-choice questions with five possible answers (one correct answer). In five of these questions the participants had to match textual information about a plant to the picture of the plant (e.g., "Which family does this plant belong to?") and in five other questions they had to match textual information to the name of the plant (e.g., "Where does the ox-eye grow?"). One point was given for a correct answer. In addition to the multiple-choice questions, at the end a picture of each of the five plants had to be matched to its respective name, which could earn one point per correct answer. In total, 15 points could be achieved.

2.1.3. Procedure

The study took place during social distancing measures due to the COVID-19 pandemic in December 2020; therefore, it was fully online with a researcher supervising each participant through synchronous (voice/video) chat. After the researcher welcomed the participants, they read the conditions and were asked for their consent. They answered the questions concerning their prior knowledge on plants. After that, the participants viewed the video showing either the integrated AR or the separated tablet-view of real plants in the botanical garden and virtual textual information. The participants were asked to imagine that they were in a real-world situation using the respective application that was displayed in the video. Afterwards, the cognitive load questionnaire and NASA TLX were administered, followed by the knowledge test. In the end, demographic data were requested, the participants were debriefed and the session was completed by the researcher. This procedure can also be seen in Figure 2.

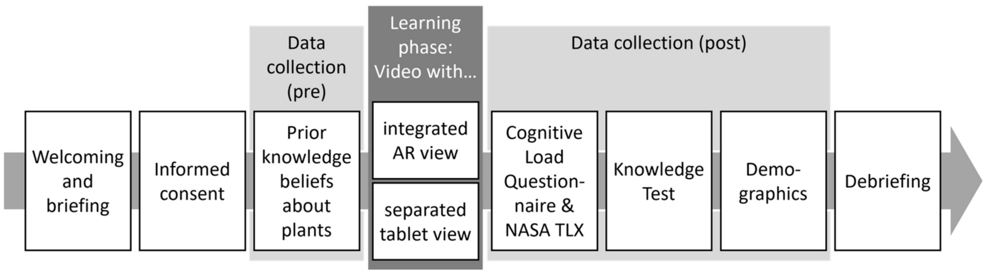


Figure 2. Procedure of study one.

2.2. Results

All hypotheses were statistically tested through independent one-sided *t*-tests, with the integration of material (integrated vs. separated) as the grouping variable and the appropriate score as the outcome variable. As suggested by Delacre and colleagues [37] we used Welch's *t*-test for all analyses, although Levene's test indicated the homogeneity of variances for all variables except physical demand (see Appendix A). The means and standard deviations of all variables can be seen in Table 3.

Table 3. Means and standard deviations of different variables in study one.

Mean and SD per Condition [a]	Possible Range	Integrated *M* (*SD*)	Separated *M* (*SD*)
H1.1a: extraneous CL	1–7	2.57 (1.19)	**2.90 (1.13)**
H1.1b: germane CL	1–7	**4.91 (1.29)**	4.46 (1.44)
H1.2a: mental demand	1–21	9.62 (4.42)	**10.51 (4.88)**
H1.2b: physical demand	1–21	1.82 (1.50)	**2.68 (3.38)**
H1.2c: temporal demand	1–21	6.90 (4.85)	**8.85 (5.44)**
H1.2d: performance	1–21	**13.31 (4.75)**	10.68 (4.99)
H1.2e: effort	1–21	8.77 (4.49)	**10.24 (5.05)**
H1.2f: frustration	1–21	5.97 (4.81)	**6.95 (5.79)**
H1.3: knowledge	0–15	**10.51 (2.81)**	10.49 (3.53)

[a.] Highest mean per subscale in bold.

2.2.1. H1.1: Cognitive Load

To test hypotheses H1.1a and H1.1b, concerning the influence of the integration of the material on extraneous cognitive load, the *t*-tests included the extraneous cognitive load and germane cognitive load subscale scores as outcome variables. H1.1a, concerning extraneous cognitive load, was tested with a one-sided *t*-test proposing lower scores for the integrated than the separated condition, while H1.1b, concerning germane cognitive load, was conversely tested proposing higher scores for the integrated than the separated condition. Boxplots showing the data for both types of cognitive load can be seen in Figure 3.

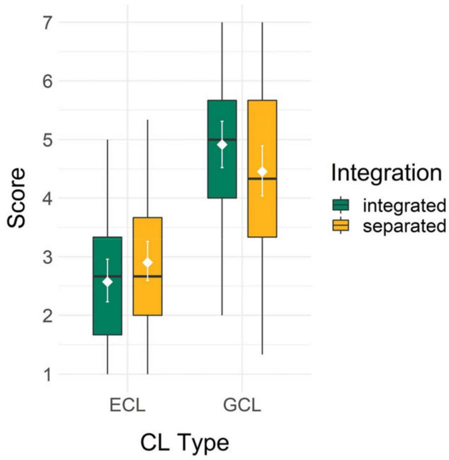


Figure 3. Distribution of extraneous cognitive load and germane cognitive load scores split by group in study one (boxplot with IQR (black), mean with bootstrapped 95% confidence interval (white)).

For H1.1a, concerning the lowering influence of the integration of the material on extraneous cognitive load, the score was indeed descriptively lower for the integrated ($M = 2.57$, $SD = 1.19$) than the separated ($M = 2.90$, $SD = 1.13$) condition. However, this effect with a small effect size was not significant, $t(77.14) = -1.27$, $p = 0.104$ and $d = -0.28$.

For H1.1b, concerning the positive influence of the integration of the material on germane cognitive load, a descriptively higher score was indeed found in the integrated ($M = 4.91$, $SD = 1.29$) than the separated ($M = 4.46$, $SD = 1.44$) condition. Again, this effect with a small effect size was not significant, $t(77.72) = 1.50$, $p = 0.069$ and $d = 0.34$.

Although descriptively the data match our expectations with small effects, H1.1a and H1.1b were not completely supported due to the non-significance of the effects: no significant advantages of the integration of the material concerning extraneous and germane cognitive load were found.

2.2.2. H1.2: Task Load

In order to test hypotheses H1.2a, H1.2b, H1.2c, H1.2d, H1.2e and H1.2f, concerning the influence of the integration of the material on task load, the *t*-tests included mental demand, physical demand, temporal demand, performance, effort and frustration subscale scores as outcome variables. Most hypotheses were tested with one-sided *t*-tests proposing lower scores for the integrated than the separated condition, except for H1.2d, concerning performance, which was tested proposing higher scores for the integrated than the separated condition. Boxplots showing the data for all six subconstructs of task load can be seen in Figure 4.

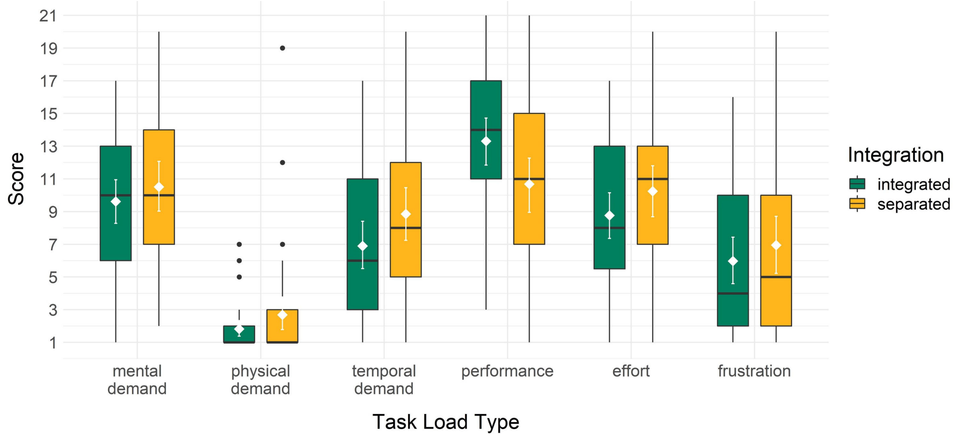


Figure 4. Distribution of task load subscale scores split by group in study one (boxplot with IQR (black), mean with bootstrapped 95% confidence interval (white)).

For H1.2a, concerning the lowering influence of the integration of the material on mental demand, the score was indeed descriptively lower for the integrated ($M = 9.62$, $SD = 4.42$) than the separated ($M = 10.51$, $SD = 4.88$) group, although this difference was not significant, $t(77.81) = -0.86$, $p = 0.195$ and $d = -0.19$. For H1.2b, concerning the lowering influence of the integration of the material on physical demand, the score was indeed descriptively lower for the integrated ($M = 1.82$, $SD = 1.50$) than the separated ($M = 2.68$, $SD = 3.38$) group. Although Cohen's d shows a small effect size, no significant difference was found between the two groups, $t(55.81) = -1.49$, $p = 0.071$ and $d = -0.33$. For H1.2c, concerning the lowering influence of the integration of the material on temporal demand, the score was indeed descriptively lower for the integrated ($M = 6.90$, $SD = 4.85$) than the separated ($M = 8.85$, $SD = 5.44$) group. This difference was significant with a small effect size, $t(77.68) = -1.70$, $p = 0.047$ and $d = -0.38$.

For H1.2d, concerning the positive influence of the integration of the material on perceived performance, the score was indeed descriptively higher for the integrated ($M = 13.31$, $SD = 4.75$) than the separated ($M = 10.68$, $SD = 4.99$) group. This difference was significant with a medium effect size, $t(78) = 2.41$, $p = 0.009$ and $d = 0.54$. For H1.2e, concerning the lowering influence of the integration of the material on effort, the score was indeed descriptively lower for the integrated ($M = 8.77$, $SD = 4.49$) than the separated ($M = 10.24$, $SD = 5.05$) group. No significant difference was found between the two groups, although a small effect size is apparent, $t(77.64) = -1.38$, $p = 0.086$ and $d = -0.31$. For H1.2f, concerning the lowering influence of the integration of the material on frustration, the score was indeed descriptively lower for the integrated ($M = 5.97$, $SD = 4.81$) than the separated ($M = 6.95$, $SD = 5.79$) group. No significant difference was found between the two groups, $t(76.66) = -0.82$, $p = 0.207$ and $d = -0.18$.

We thus only found a significant difference concerning temporal demand and performance in the expected directions, supporting H1.2c and H1.2d: temporal demand was perceived as lower while performance was perceived as higher when the material was integrated instead of separated. For the other variables no significant effects were found, although the directions of the mean differences were as expected, with at least small effects for physical demand (H1.2b) and effort (H1.2e). Not even small effects were found for mental demand (H1.2a) and frustration (H1.2f).

2.2.3. H1.3: Knowledge

To test H1.3 on the positive influence of the integration of the material on knowledge, a one-sided *t*-test proposing higher scores for the integrated condition included knowledge test score as the outcome variable. Descriptively, the knowledge test scores in the integrated (M = 10.51, SD = 2.81) and separated (M = 10.49, SD = 3.53) groups barely differed and no significant difference was found, $t(75.68) = 0.04$, $p = 0.486$ and $d = 0.01$. Hypothesis H1.3 was thus not supported: no significant advantages of the integration of the material concerning knowledge were found.

As different kinds of items were used in the knowledge test, we also took a closer exploratory look at these. Concerning the six items with ten possible points, in which the picture of a plant had to be matched with a textual characteristic or name, the pattern was the same as in the complete knowledge test results: higher score for the integrated (M = 6.90, SD = 1.90) compared to the separated (M = 6.80, SD = 2.51) visualization. This difference was descriptively only minimally bigger than the general knowledge difference and not significant, $t(74.34) = 0.19$, $p = 0.426$ and $d = 0.04$. Concerning the five items in which only text where included, thus matching the name of a plant with a characteristic, this pattern was the opposite. The separated condition had a higher score (M = 3.68, SD = 1.40) than the integrated (M = 3.62, SD = 1.23) condition. This difference was also very small and not significant, $t(77.46) = -0.23$, $p = 0.819$ and $d = -0.05$ (two-sided *t*-test due to the descriptive direction of difference being opposite than expected for knowledge in general). Although these differences between the groups are very small, the items in which pictures and text needed to be matched thus descriptively showed an advantage for the integrated presentation, while the items in which different textual elements needed to be matched descriptively showed an advantage for the separated conditions. Boxplots showing the data split by item type and for the complete knowledge test can also be see in Figure 5.

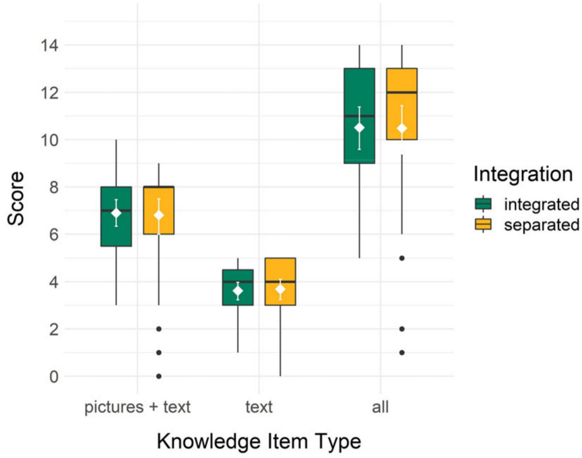


Figure 5. Distribution of knowledge test scores by item types split by group in study one (boxplot with IQR (black), mean with bootstrapped 95% confidence interval (white)).

2.3. Discussion

The goal of study one was to implement and examine the spatial contiguity principle with a combination of visual real and virtual elements, leveraging the AR characteristic spatiality. We examined the influence of the integration of the material on cognitive load, task load and knowledge. Only two of the hypotheses were supported by the data, namely H1.2c, concerning the decrease in temporal demand, and H1.2d, concerning the increase in perceived performance through the integrated presentation of the real and virtual elements.

All other hypotheses were not fully supported, although the tendencies in the data showed expected differences. Descriptively, the cognitive load scores were as expected, with small effects describing decreased extraneous cognitive load (H1.1a) and increased germane cognitive load (H1.1b) in the integrated presentation. Further non-significant effects with small effect sizes were found for the task load subconstructs physical demand (H1.2b) and effort (H1.2e), and even smaller, non-significant differences for mental demand (H1.2a) and frustration (H1.2f), all showing a decrease in the integrated presentation. Additionally, concerning H1.3, no effects on resulting knowledge were found, although interestingly different patterns were revealed for knowledge items in which pictures and text had to be combined and in which only text was included, with descriptively better results for the integrated presentation in the former and better results for the separated presentation in the latter. These results suggest that the spatial integration of pictorial real-world elements and textual virtual elements may have slightly strengthened the building of connections between the pictures and texts, while it may have even more slightly weakened the building of connections within the added texts. This differentiation between representational connections is in accordance with Seufert [38], who described connections within a representation as intra-representational and connections between representations as inter-representational coherence formation, and with Seufert and Brünken [39], who described these concepts as local and global coherence formation. They argue that for deeper understanding both forms of relating need to be achieved [38], and that learners can be supported in their coherence formation on a surface feature level or a deep structure level [39]. Following the spatial contiguity principle may be a form of support on the surface feature level, as the goal is to directly show learners which information belongs together. In future studies it may be interesting to also look at guidance that supports coherence formation in AR-based multimedia representations on a deep structure level, potentially extending the results of the current study.

Different factors may have led to these results showing mainly small, non-significant effects, although all with tendencies in the expected direction. The elements in the presentation could be understood independent from each other and did not have to be mentally integrated to learn the content, such that this pre-requisite for the spatial contiguity principle [12] and the split-attention effect [13] was not given. However, considering the learning goal of integrating the real-world elements and virtual elements, their combination was at least necessary for the knowledge test items in which pictures and textual information needed to be matched. Descriptively, the difference between the pattern in these items and the purely textual items supports this idea, but the simple connection between one picture and few textual characteristics for each plant may not be enough to lead to a strong effect.

Another factor leading to only small effects may have been that the learning situation was not demanding enough for the spatial contiguity principle to show effects on subjective cognitive load and task load. A boundary condition for the spatial contiguity principle is that the material needs to be complex for a strong effect to occur [12], and the split-attention effect is said to mainly appear in materials with higher element interactivity [13]. The learning materials concerning the different plants in a botanical garden were not very complex, including only real-world pictures and a few characteristics of said plants. More complex materials with higher element interactivity may lead to larger effects concerning the two types of cognitive load and the subconstructs of task load. This should be tested in future studies with a focus on material in which virtual textual elements need to be mentally combined into a coherent representation with the physical objects in order to understand their structure and function.

The simplicity of the materials due to the video-based implementation of the AR environment may also have been the reason for the missing effects. The video-based implementation provided different opportunities, including the decrease in individual differences that may usually be caused by non-standardized interaction with AR-based learning applications. This can, for example, include the duration that learners spend engaging with specific components or the exact view of the spatially integrated or separated materials, which could be standardized through the videos. This way, the focus was on the implementation of the spatial contiguity principle decreasing potentially confounding variables and increasing comparability between participants. A limitation is that the usage of the video-based implementation also removed some factors of complexity that real, location-based AR experiences have. In a real-world AR experience, more cognitive load and potentially overload might have been evoked through a combination of all impressions and necessary skills [40], requiring a decrease in cognitive load through the instructional design. The goal of the study was to examine the specific case of the spatial integration of real and virtual elements, which may have been negatively influenced by the video implementation. Participants may not have perceived the real-world plants in the video as real because they saw them on a screen compared to seeing them physically in front of them. Although it was a video recording of the real plants and the participants were asked to imagine seeing the plants in the real world, it is not sure if they achieved this or if the real elements were perceived as virtual. It is thus not clear how well the results from this study can be transferred onto the usage of real AR applications. A future step would be to implement the spatial contiguity principle in a field study in a location-based AR environment, including a more authentic, naturalistic setting. The advantages and limitations of the video-based implementations in both studies will be discussed more broadly in the general discussion.

Other limitations of the study were that, although a complete standardization of the experiences would have been possible through the video-based implementation, there were some differences between the two videos that were not controlled. Because the videos were filmed on different days, it was sunny in the AR-view video and cloudy in the tablet-view video. Furthermore, the individual plants and information were shown for around 25 s in the tablet-view video and around 30 s in the AR-view video. While there was only little textual information that all participants should have been able to read in 25 s, the time difference and the weather may have had a systematic influence on the results, which should be taken into account.

In conclusion, study one showed (descriptive) results suggesting a potential decrease in extraneous cognitive load as well as task load and a simultaneous increase in germane cognitive load and knowledge concerning the linking of real and virtual elements when learning with an integrated instead of a separated presentation. This suggests that the spatial contiguity principle could be transferable onto AR environments, although the results need to be supported in additional research with more complex material and real AR applications.

3. Study Two: Coherence Principle

In the second study the goal is to examine the contextual coherence of virtual information. The learning material is audiovisual, and it is focused on the implementation of the coherence principle in AR. In AR-specific implementations the principle can be applied to real and virtual, visual and auditory elements. In the study, virtual sounds that either match or do not match the topic of the learning material are added into an application that also includes virtual texts and pictures in addition to real environmental sounds. These sounds are no direct part of the learning task and are compared to the omission of virtual auditory elements. Based on the coherence principle, we want to examine if the implementation of the principle in particular has a positive influence on cognitive factors, including cognitive load, task load and knowledge.

We hypothesize that complying with the coherence principle in AR and thus not adding any sounds leads to a decrease in extraneous processing and thus in extraneous cognitive load (H2.1a) by reducing the number of elements that have to be processed. In turn, the working memory capacities that are made available can be used for generative processing, thus increasing germane cognitive load when no sounds are added, although we expect matching sounds to also increase germane cognitive load within the limits due to motivational effects, which non-matching sounds do not elicit (H2.1b).

Furthermore, we hypothesize that following the coherence principle has an influence on task load. We expect that through the necessity to attend to less sensory input and the explicit usage of fewer sensory organs, both mental (H2.2a) and physical demand (H2.2b) are decreased when no sounds are added. We also expect temporal demand to be decreased with the omission of additional sounds (H2.2c), because less sensory input can be processed within the same time. We propose that the decreased potential for distraction through additional sensory input when no additional sounds are presented leads to feelings of higher performance (H2.2d) and lower effort (H2.2e). Frustration is on one hand expected to be smaller when no distracting sounds are added at all, but on the other hand is expected to be smaller for the addition of matching sounds compared to non-matching sounds because the reason for adding these is not apparent and might lead to even higher frustration (H2.2f).

Through the decrease in extraneous cognitive load as well as the task-load related factors, and the resulting increase in germane cognitive load, we would also expect increased resulting knowledge when no sounds are added, although through motivating effects and decreased frustration we would further expect that matching sounds lead to higher resulting knowledge than non-matching sounds (H2.3). All hypotheses of study two are summarized in Table 4.

Table 4. Hypotheses in study two.

Hypotheses in Study Two	
H2.1a:	learning with material including real and virtual information without additional virtual sounds leads to lower *extraneous cognitive load* than when virtual sounds are added.
H2.1b:	learning with material including real and virtual information without additional virtual sounds leads to higher *germane cognitive load* than when virtual sounds are added, where adding matching sounds leads to higher *germane cognitive load* than adding non-matching sounds.
H2.2a:	learning with material including real and virtual information without additional virtual sounds leads to lower *mental demand* than when virtual sounds are added.
H2.2b:	learning with material including real and virtual information without additional virtual sounds leads to lower *physical demand* than when virtual sounds are added.
H2.2c:	learning with material including real and virtual information without additional virtual sounds leads to lower *temporal demand* than when virtual sounds are added.
H2.2d:	learning with material including real and virtual information without additional virtual sounds leads to higher perceived *performance* than when virtual sounds are added.
H2.2e:	learning with material including real and virtual information without additional virtual sounds leads to lower *effort* than when virtual sounds are added.
H2.2f:	learning with material including real and virtual information without additional virtual sounds leads to lower *frustration* than when virtual sounds are added, where adding matching sounds leads to lower *frustration* than adding non-matching sounds.
H2.3:	learning with material including real and virtual information without additional virtual sounds leads to higher *knowledge* than when virtual sounds are added, where adding matching sounds leads to higher *knowledge* than adding non-matching sounds.

3.1. Methods

In a between-subjects design with three conditions, the addition or omission of matching or non-matching virtual sounds was manipulated in a video-based simulation of a location-based informational AR application. One group heard no additional sounds, one heard sounds matching the learning topic and one heard sounds that did not match the topic. Dependent variables are cognitive load, task load and knowledge.

3.1.1. Participants

The participants were reached through the same online platforms for participant sampling of the department as in study one and convenience sampling. Students could receive participant hours for taking part. The final dataset included N = 130 people after two outliers were filtered out based on very long study duration. Primarily (86%) students took part, of which most were in the study programs of applied cognitive and media science (76%) and psychology (17%), in which there are no classes related to the learning topic of the study. They were aged 18 to 61 (M = 23.72, SD = 7.95) and 34 indicated being male, 96 being female. On average, the participants did not indicate high prior knowledge beliefs concerning the focal learning topic of regional birds in a subjective rating (M = 1.59, SD = 0.50; 5-point response format from 1—low to 5—high). The participants on average indicated having very rarely used general AR applications (M = 1.81, SD = 0.81) and AR learning applications (M = 1.31, SD = 0.68; both measured in a 5-point response format, with 1—"never", 2—"rarely", 3—"now and then", 4—"often" and 5—"regularly"). The participants were randomly distributed into the three groups. In Table 5 the number of participants, gender, age, prior knowledge beliefs and prior usage of AR applications per condition are shown. The distribution is quite balanced for all variables except for age, with a descriptively higher age in the matching group. All subjects gave their informed consent for inclusion before they participated in the study. The study with the ID psychmeth_2020_AR14_30 was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the department's Ethics Committee (vote ID: 2012PFKL8474).

Table 5. Distribution of the number of participants, gender, age, prior knowledge beliefs and prior usage of AR applications split by condition in study two.

Condition	n	Gender		Age	Prior Knowledge	Usage AR Applications	
						General	Learning
		Male	Female	M (SD)	M (SD)	M (SD)	M (SD)
No sounds	43	11	32	22.30 (6.13)	1.54 (0.42)	1.84 (0.72)	1.35 (0.81)
Matching	44	13	31	26.52 (11.10)	1.61 (0.45)	1.73 (0.79)	1.36 (0.72)
Non-matching	43	10	33	22.28 (4.23)	1.61 (0.62)	1.86 (0.91)	1.21 (0.47)

3.1.2. Materials

The independent variable was manipulated by showing simulated AR experiences to the participants through three different videos during the learning phase. The visual material and real environmental sounds were the same for all videos. The scene was filmed in a forest, in which six different locations were walked towards and focused on with the camera, where a picture of and additional textual information (common name, scientific name, size and food) about a different bird then appeared in each location. The simulated AR experience in this study thus included the video of the forest including sounds in the background as the real-world environment which the participants should imagine themselves to be in. The additional pictorial (i.e., the bird) and textual (i.e., bird characteristics) information was included as virtual elements in the AR experience which the participants should imagine as being viewed through AR glasses. The videos differed in the additional virtual sounds that were played when focusing on a bird: either no additional sound (no sounds), the chirping of that bird (matching sounds) or another unrelated sound, such as a bell (non-matching sounds). In Figure 6, screenshots of two frames including birds from the videos are shown. In all conditions, the participants were asked to imagine that they were walking through the forest and using the AR application in the real world themselves. The videos took 6 min.

Figure 6. Two screenshots from the video used in study two.

Subjective prior knowledge was measured with the same scale by Wigfield and Eccles [33] used in study one, although the content area was adapted. Cronbach's alpha was questionable (α = 0.58), but the scale was kept the same because it only has three items.

Extraneous and germane cognitive load were measured as in study one with the cognitive load scale by Klepsch and colleagues [34]. Cronbach's alpha was questionable for extraneous cognitive load at α = 0.59 and germane cognitive load at α = 0.62, but we kept the scales as they were because of the already low number of three items per scale.

Task load including mental demand, physical demand, temporal demand, performance, effort and frustration was also measured with the six one-item scales from the NASA TLX used in study one [35,36]. Again, the score for the performance item was inverted for easier interpretation of the scores, with high scores meaning high perceived performance and low scores meaning low perceived performance.

Learning outcomes were measured through a knowledge test (8 items) that included different kinds of questions, all multiple-choice questions with four possible answers (one correct answer). In four questions the participants had to match textual information about a bird to the picture of the bird (e.g., "What is the name of this bird?"), in four other questions they had to match textual information to the name of the bird (e.g., "What does the bullfinch eat?"). One point was given for a correct answer, so that in total 8 points could be reached.

For exploratory analyses, a sound–picture matching test was also administered in the two groups that received additional sounds in the learning test. In two items the participants were asked to match a sound to the bird's name and in two other items to the bird's picture.

3.1.3. Procedure

As in study one, the study took place during social distancing measures due to the COVID-19 pandemic in December 2020, such that it was fully online with a researcher supervising each participant through synchronous (voice/video) chat. The procedure is very similar to that of study one. After the researcher welcomed the participants, they read the conditions and were asked for their consent. They answered the questions concerning their prior knowledge on regional birds. After that, the participants viewed the video showing the virtual birds and textual information in the real forest and adding either no, matching or non-matching sounds. The participants were asked to imagine that they were in the real-world situation using the respective application that was displayed in the video. Afterwards, the cognitive load questionnaire and NASA TLX were administered, followed by the knowledge test. In the end, demographic data were requested, the participants were debriefed, and the session was completed by the researcher. This procedure can also be seen in Figure 7.

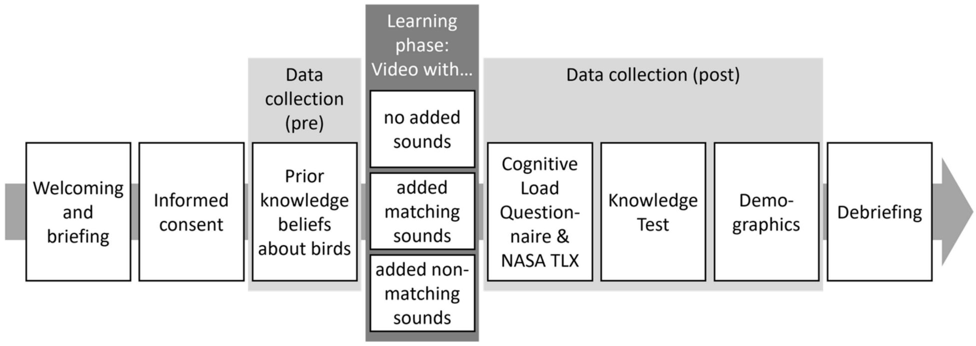


Figure 7. Procedure in study two.

3.2. Results

All hypotheses were statistically tested through one-way ANOVAs with type of sounds (no vs. matching vs. non-matching) as the predictor and the appropriate score as the outcome variable. Levene's test indicated the homogeneity of variances for all variables (see Appendix A). A-priori-determined contrasts based on the individual hypotheses are used for the analyses. The first comparison in the contrast analysis always focuses on no added sounds compared to added sounds, while the second comparison focuses on matching sounds compared to non-matching sounds. When the focus of the hypothesis is on the first comparison in the contrast, the second comparison is explored for a more complete picture. Means and standard deviations of all variables can be seen in Table 6.

Table 6. Means and standard deviations of different variables in study two.

Mean and SD per Condition [a]	Possible Range	No Sound *M* (*SD*)	Matching *M* (*SD*)	Non-Matching *M* (*SD*)
H2.1a: extraneous CL	1–7	*2.24 (1.08)*	**2.55 (1.00)**	2.53 (1.16)
H2.1b: germane CL	1–7	4.81 (1.41)	4.76 (1.26)	**5.12 (1.22)**
H2.2a: mental demand	1–21	*8.60 (4.47)*	**9.98 (4.71)**	9.74 (4.33)
H2.2b: physical demand	1–21	**4.12 (4.23)**	3.93 (3.39)	*3.40 (3.58)*
H2.2c: temporal demand	1–21	*6.84 (4.57)*	**7.64 (4.92)**	7.47 (4.67)
H2.2d: performance	1–21	11.95 (4.89)	*10.95 (5.25)*	**13.00 (4.72)**
H2.2e: effort	1–21	*7.95 (4.54)*	**9.70 (4.52)**	8.21 (4.38)
H2.2f: frustration	1–21	5.65 (4.83)	**6.11 (4.94)**	*5.00 (4.89)*
H2.3: knowledge	0–8	*3.70 (1.74)*	3.86 (1.77)	**4.02 (1.57)**

[a.] Highest mean per subscale in bold, lowest mean per subscale in italics.

3.2.1. H2.1: Cognitive Load

To test hypotheses H2.1a and H2.1b on the influence of added sounds on cognitive load, the one-way ANOVAs included extraneous cognitive load and germane cognitive load subscale scores as outcome variables. Boxplots showing the data for both types of cognitive load can be seen in Figure 8.

In H2.1a, extraneous cognitive load was hypothesized to be lower when no sounds were added than when either matching or non-matching sounds were added. This pattern was indicated descriptively in the group means, showing a lower extraneous cognitive load in the no sounds (M = 2.24, SD = 1.08) than the very similar matching sounds (M = 2.55, SD = 1.00) and non-matching sounds (M = 2.53, SD = 1.16) conditions. The overall model of the one-way ANOVA was not significant, $F(2, 127) = 1.14$, $p = 0.324$ and $\omega^2 < 0.01$. Additionally, no significant differences were found in the first comparison of the Helmert contrast analysis, comparing no sounds and added sounds, $t = -1.51$, $p = 0.135$, or in the second comparison, comparing matching and non-matching sounds, $t = 0.08$, $p = 0.938$.

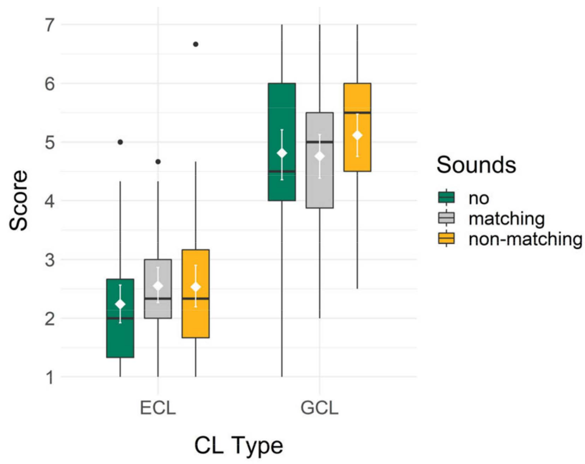


Figure 8. Distribution of extraneous cognitive load and germane cognitive load scores split by group in study two (boxplot with IQR (black), mean with bootstrapped 95% confidence interval (white)).

In H2.1b, germane cognitive load was hypothesized to be higher when no sounds were added than when either matching or non-matching sounds were added, and higher when matching than when non-matching sounds were added. Descriptively, a different pattern was shown with the highest germane cognitive load in the non-matching (M = 5.12, SD = 1.22) condition, then the no sounds (M = 4.81, SD = 1.41) and then the matching sounds (M = 4.76, SD = 1.26) conditions. The overall model of the one-way ANOVA was not significant, $F(2, 127) = 0.94$, $p = 0.393$ and $\omega^2 < 0.00$. Additionally, no significant differences were found in the first comparison of the Helmert contrast analysis, $t = -0.52$, $p = 0.607$, or in the second comparison, $t = -1.27$, $p = 0.205$.

We thus found no support for H2.1a and H2.1b: no significant advantages of leaving out additional sounds were found concerning extraneous or germane cognitive load.

3.2.2. H2.2: Task Load

For testing H2.2a, H2.2b, H2.2c, H2.2d, H2.2e and H2.2f on the influence of added sounds on cognitive load, the one-way ANOVAs included mental demand, physical demand, temporal demand, performance, effort and frustration subscale scores as outcome variables. Boxplots showing the data for all six subconstructs of task load can be seen in Figure 9.

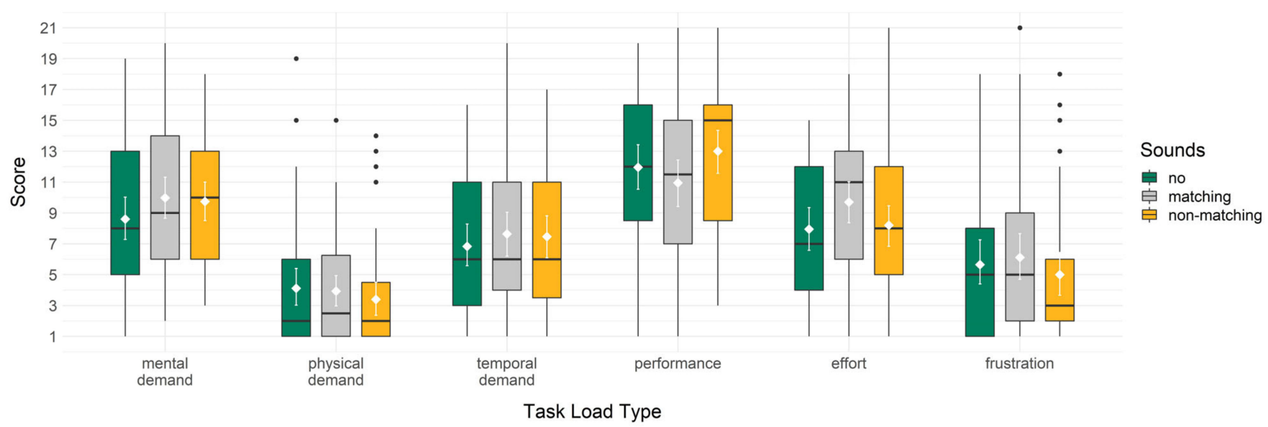


Figure 9. Distribution of task load subscale scores split by group in study two (boxplot with IQR (black), mean with bootstrapped 95% confidence interval (white)).

In H2.2a, mental demand was hypothesized to be lower when no sounds were added than when either matching or non-matching sounds were added, which was indicated descriptively in the group means, showing a lower mental demand in the no sounds (M = 8.60, SD = 4.47) than the matching sounds (M = 9.98, SD = 4.71) and the non-matching sounds (M = 9.74, SD = 4.33) conditions. The overall model of the one-way ANOVA was not significant, $F(2, 127) = 1.15$, $p = 0.320$ and $\omega^2 < 0.01$. Additionally, no significant differences were found in the first comparison of the Helmert contrast analysis, $t = -1.50$, $p = 0.137$, or in the second comparison, $t = 0.24$, $p = 0.810$.

In H2.2b, physical demand was hypothesized to be lower when no sounds were added than when either matching or non-matching sounds were added. An opposite pattern was indicated descriptively in the group means, showing the lowest physical demand in the non-matching sounds (M = 3.40, SD = 3.58), higher physical demand in the matching sounds (M = 3.93, SD = 3.39), and the highest physical demand in the no sounds (M = 4.12, SD = 4.23) condition. The overall model of the one-way ANOVA was not significant, $F(2, 127) = 0.45$, $p = 0.638$, $\omega^2 = -0.01$. Additionally, no significant differences were found in the first comparison of the Helmert contrast analysis, $t = 0.66$, $p = 0.512$, or in the second comparison, $t = 0.69$, $p = 0.494$.

In H2.2c, temporal demand was hypothesized to be lower when no sounds were added than when either matching or non-matching sounds were added, which was indicated descriptively in the group means, showing a lower temporal demand in the no sounds (M = 6.84, SD = 4.57) than the matching sounds (M = 7.64, SD = 4.92) and the non-matching sounds (M = 7.47, SD = 4.67) condition. The overall model of the one-way ANOVA was not significant, $F(2, 127) = 0.34$, $p = 0.710$, $\omega^2 = -0.01$. Additionally, no significant differences were found in the first comparison of the Helmert contrast analysis, $t = -0.81$, $p = 0.419$, or in the second comparison, $t = 0.17$, $p = 0.866$.

In H2.2d, performance was hypothesized to be perceived as higher when no sounds were added than when either matching or non-matching sounds were added. A different pattern was indicated descriptively in the group means, showing the highest perceived performance in the non-matching sounds (M = 13.00, SD = 4.72), lower perceived performance in the no sounds (M = 11.95, SD = 4.89) and lowest in the matching sounds (M = 10.95, SD = 5.25) conditions. The overall model of the one-way ANOVA was not significant, $F(2, 127) = 1.85$, $p = 0.162$ and $\omega^2 = 0.01$. Additionally, no significant differences were found in the first comparison of the Helmert contrast analysis, $t = -0.03$, $p = 0.980$, or in the second comparison, $t = -1.92$, $p = 0.057$.

In H2.2e, effort was hypothesized to be lower when no sounds were added than when either matching or non-matching sounds were added, which was indicated descriptively in the group means, showing a lower effort in the no sounds (M = 7.95, SD = 4.54) than the matching sounds (M = 9.70, SD = 4.52) and the non-matching sounds (M = 8.21, SD = 4.38) conditions. The overall model of the one-way ANOVA was not significant, $F(2, 127) = 1.95$, $p = 0.147$ and $\omega^2 = 0.01$. Additionally, no significant differences were found in the first comparison of the Helmert contrast analysis, $t = -1.20$, $p = 0.232$, or in the second comparison, $t = 1.56$, $p = 0.122$.

In H2.2f, frustration was hypothesized to be lower when no sounds were added than when either matching or non-matching sounds were added, and lower when matching than when non-matching sounds were added. A different pattern was indicated descriptively in the group means, showing the lowest frustration in the non-matching sounds (M = 5.00, SD = 4.89), higher frustration in the no sounds (M = 5.65, SD = 4.83) and the highest frustration in the matching sounds (M = 6.11, SD = 4.94) conditions. The overall model of the one-way ANOVA was not significant, $F(2, 127) = 0.57$, $p = 0.568$ and $\omega^2 = -0.01$. Additionally, no significant differences were found in the first comparison of the Helmert contrast analysis, $t = 0.10$, $p = 0.918$, or in the second comparison, $t = 1.06$, $p = 0.290$.

We thus did not find support for hypotheses H2.2a, H2.2b, H2.2c, H2.2d, H2.2e and H2.2f, showing no significant advantage for not adding sounds concerning mental demand, physical demand, temporal demand, perceived performance, effort and frustration.

3.2.3. H2.3: Knowledge

To test H2.3 on the influence of added sounds on knowledge, the one-way ANOVA included the knowledge test score as the outcome variable. Knowledge was hypothesized to be higher when no sounds were added than when either matching or non-matching sounds were added, and higher when matching than when non-matching sounds were added. Descriptively, an opposite pattern was shown with the highest knowledge in the non-matching sounds ($M = 4.02$, $SD = 1.57$) condition, then the matching sounds ($M = 3.86$, $SD = 1.77$) and the no sounds ($M = 3.70$, $SD = 1.74$) conditions. The overall model of the one-way ANOVA was not significant, $F(2, 127) = 0.40$, $p = 0.674$ and $\omega^2 = -0.01$. Additionally, no significant differences were found in the first comparison of the Helmert contrast analysis, $t = -0.78$, $p = 0.438$, or in the second comparison, $t = -0.44$, $p = 0.661$. Hypothesis H2.3 was thus not supported: no significant advantages of following the coherence principle by leaving out additional sounds were found concerning knowledge outcomes.

As different kinds of items were used in the knowledge test, we also took a closer exploratory look at these. Concerning the four items in which the picture of a bird had to be matched with a textual characteristic or name, the pattern was the same as in the complete knowledge test results: highest score for the non-matching ($M = 2.30$, $SD = 1.15$) compared to the matching ($M = 2.09$, $SD = 1.22$) and no sounds ($M = 2.09$, $SD = 1.19$) conditions. The full model was not significant, $F(2, 127) = 0.45$, $p = 0.636$ and $\omega^2 = -0.01$. Additionally, in the contrast analysis neither the first comparison, $t = -0.47$, $p = 0.640$, nor the second comparison, $t = -0.83$, $p = 0.407$, were significant. Concerning the four items in which only text was included, thus matching the name of a bird with a characteristic, this pattern was a little different. Here, the matching condition had the highest score ($M = 1.77$, $SD = 1.10$), then the non-matching ($M = 1.72$, $SD = 1.08$) and then the no sounds ($M = 1.60$, $SD = 1.03$) conditions. Again, the full model was not significant, $F(2, 127) = 0.28$, $p = 0.755$ and $\omega^2 = -0.01$, and in the contrast analysis neither the first comparison, $t = -0.71$, $p = 0.476$, nor the second comparison, $t = 0.23$, $p = 0.821$, were significant. Although these differences between the groups are very small, the items in which pictures and text needed to be matched thus descriptively showed an advantage for the non-matching sounds condition, while the items in which different textual elements needed to be matched descriptively showed an advantage for the matching sounds condition. Boxplots showing the data split by item type and for the complete knowledge test can also be seen in Figure 10.

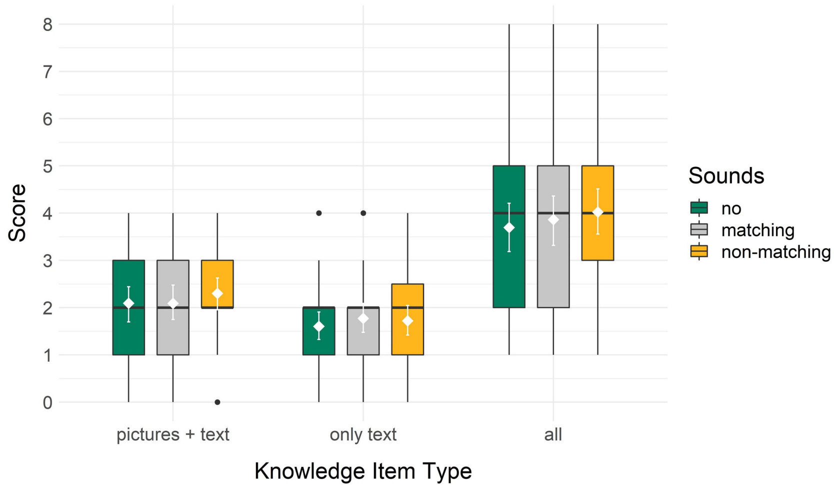


Figure 10. Distribution of knowledge test scores by item types split by group in study two (boxplot with IQR (black), mean with bootstrapped 95% confidence interval (white)).

In additional exploratory analyses, we compared how much the sounds were connected to the pictures and names of the birds in the participants' memory. In this analysis, we thus only compared the groups that received sounds, thus the matching and non-matching group, in a Welch's *t*-test with the score in the audio-matching knowledge test as the outcome variable. From eight possible points, the mean was higher in the non-matching (M = 1.79, SD = 1.10) than the matching (M = 1.39, SD = 0.84) condition. Descriptively, the data thus show that the condition with non-matching sounds remembered the connection between sound and bird more correctly than the matching group. The effect has a small effect size and is not significant, $t(78.53) = -1.92$, $p = 0.059$ and $d = -0.41$.

3.3. Discussion

The goal of study two was to implement and examine the coherence principle in a learning environment including a combination of (contextually matched) real and virtual, visual and auditory elements, leveraging the AR characteristic contextuality. We examined the influence of omitting or adding either matching or non-matching sounds on cognitive load, task load and the resulting knowledge. None of the hypotheses were supported, although the tendencies in the data partly showed expected effects. Descriptively, the extraneous cognitive load (H2.1a), mental demand (H2.2a), temporal demand (H2.2c) and effort (H2.2e) scores were as expected, with lower scores when no sounds were added in comparison to matching and non-matching added sounds. For all these scores, matching sounds led to the highest scores, which may suggest that learners tried to remember the sounds, which may have been more demanding and effortful. By adding sounds that were relevant for the learning material, learners may have been more distracted because they may have thought that the sounds were important to listen to and remember, which was probably not the case for the non-matching sounds. This is in accordance with the boundary conditions for the coherence principle specified by Mayer, which describe that the principle is more important when the added extraneous material is more interesting for learners [15]. While participants may also have tried to remember the sounds in the non-matching condition, the difference may have been that those were easier to distinguish and remember. We descriptively found that participants who received the non-matching sounds could better connect them to the birds that appeared simultaneously with the sound. This may have mainly been due to the recognizability of those sounds and that they were easily distinguishable in comparison to the different bird sounds. In the future, research could look at different kinds of sound, especially more and less distinguishable in addition to more and less familiar sounds.

Further non-significant effects that also had different descriptive patterns than those expected show that germane cognitive load (H2.1b) and perceived performance (H2.2d) were highest while physical demand (H2.2b) and frustration (H2.2f) were lowest when non-matching sounds were added. Maybe the non-matching sounds provided the learners with a way to stay alert when the visual information was shown. Additionally, they could have been able to tune these non-relevant sounds out after some time. Concerning the knowledge outcomes, hypothesis H2.3, describing a positive effect of sound omission in comparison to sound addition and a positive effect of matching in comparison to non-matching sounds, was also not supported. Still, interesting descriptive results were found when comparing the different forms of test items. While for the picture–text items (matching the pictures of birds to their characteristics) the highest mean score was in the non-matching sounds condition, for the completely textual items (matching the names of birds to their characteristics) the highest mean score was in the matching sounds condition. Although none of the differences between the groups are significant, this may indicate a difference for different kinds of test items. In general, the lack of significant differences concerning all variables may indicate that the difference between the conditions through the addition of only small sounds such as the ones used in the study is not big enough to completely disrupt learning and confirm an effect based on the coherence principle.

In the current study, we only examined cognitive factors, not affective or motivational aspects. In his description of the theoretical basis for the coherence principle, Mayer contrasts the idea that interesting elements can lead to higher motivation in learners with the active processing assumption of CTML, describing that humans need to actively select and process information for learning, such that no extraneous information should be added to disturb these processes [15]. Still, he further writes that based on the cognitive affective model of learning with media (see [41]) and the integrated cognitive affective model of learning with multimedia (see [42]), cognitive and affective processing can influence each other, describing that the first interest may be attracted through seductive details but that further interest must come from personal value that learners individually find in the material [15]. The cognitive in comparison to the affective factors of seductive details have been discussed in the literature, and it has been found that in high-load situations specifically information that cannot be completely ignored, such as auditory narrations, may interfere with learning [43]. While the information in the current study was not a narration, it was auditory. As expected, the bird sounds may have had a distracting effect, which added load onto the mental tasks of immersing oneself in the AR experience, watching the video and integrating pictorial and textual information, which is more mentally demanding than the combination of pictures and auditory narration (modality principle, [10]).

The descriptive results in the current study seem to at least partly and descriptively support the cognitive detriments of adding contextually relevant and interesting auditory elements to AR-based learning experiences. Still, research on AR often focuses on affective and motivational aspects, which have been identified as important factors in AR-based instructional environments (e.g., [4,19,21]). It may thus be interesting for future research to take a closer look at how the addition of atmospheric, relevant sounds in AR experiences have an influence on the enjoyment of the experience, and for example on learners' wishes to use the application more often, while still keeping the cognitive aspects in mind. The usage of environmental sound in the form of spatial AR soundscapes for realistic experiences through specific recording and reproduction techniques, taking into account movement and an interaction of virtual with real sounds [11], might be considered for this. A learning-specific focus might be on the relevance of atmospheric and specific sounds for motivational and cognitive learning goals within a specific context.

Again, the study was executed with a video-based implementation simulating an AR-based learning experience. In addition to the real and virtual visual elements presented in study one, in study two there were also additional real environmental sounds, and in two of the three conditions virtual sounds were added. This video-based implementation might thus have provided a more sensorily immersive experience than in study one. An advantage of using videos again was being able to keep the conditions the same except for the manipulated variable. The participants all received exactly the same visual material with the same real sounds in the background; only the added virtual sounds were different. In a study in a real forest, factors such as the weather, background sounds and present animals may differ between participants and lead to different coherence-related circumstances and distractions. Additionally, the interaction and duration might differ when people have control over the usage, where they may focus on different aspects of the material and environment. Through the video-based implementation, confounding factors are thus decreased and (descriptive) differences can be attributed to the omission or addition of sounds. The limitations due to this implementation are again mainly based on the question of the transferability of the results onto real AR-based learning experiences. Especially concerning coherence, the interaction with, movement in and presence within an immersive environment may play a big role and can be both motivating and distracting, as stated in Mayer's immersion principle [31]. These factors could not be considered in the same way in the present study, because the immersion in a video-based presentation is different than physically being in the environment. The transferability should be tested in future studies. The advantages and challenges of the video-based implementation will be further discussed in the general discussion.

Other limitations of the study were that the real-world background sound of the forest included bird sounds from the environment that may have been distracting, especially in the matching sounds condition. As this is realistic, it increases comparability to a setting in the real world but cannot be excluded as a confounding factor and should be taken into account when interpreting the results. Furthermore, some participants had technical difficulties with the video playback, although this was not systematic in one condition but randomly distributed.

In conclusion, it can be said that the descriptive results are not conclusive in supporting the coherence principle in the case of omitting or adding virtual matching or non-matching sounds in an environment with real and virtual elements. The addition of small sounds might not have that big of an influence, and future research should test for the transferability of the results onto more complex materials and real AR-based learning environments.

4. General Discussion

The goal of the paper is to apply multimedia design principles to the specific case of AR and evaluate them in experimental studies concerning their effects on cognitive load, task load and the resulting knowledge. Only a few of the hypotheses that we tested in the two studies were supported, although interesting and mostly expected descriptive tendencies emerged. In study one, the data did not fully support the expected positive effect of following the spatial contiguity principle through the integration of visual real and virtual elements on cognitive load, task load, and knowledge. Only the hypotheses describing decreased temporal demand and increased perceived performance through the integration of the material were supported. Descriptively, we also found the expected results of integration of the material concerning decreased extraneous cognitive load, mental demand, physical demand, effort, and frustration, in addition to increased germane cognitive load and positive results for the picture–text knowledge test items. In study two, the data did also not fully support the expected positive effects of following the coherence principle by omitting virtual sounds that are matching or non-matching to the learning material on cognitive load, task load, and knowledge. We did find descriptively lowest extraneous cognitive load, mental demand, temporal demand, and effort when no sounds were added, but opposed to what we expected found lowest physical demand and highest germane cognitive load and perceived performance when non-matching sounds were added. Again, different kinds of knowledge test items showed different patterns, with textual items revealing the highest score when matching sounds were added, and picture–text items revealing the highest score when non-matching sounds were added.

4.1. Methodological Approach

As already described in the study-specific discussions, there are several advantages but also some limitations due to the video-based, simulated implementation of the AR environment. While the implementation of the video format in a fully online study was a great opportunity for us to conduct AR-related research during social distancing measures in the COVID-19 pandemic, beyond this there are additional advantages for experimental research. Due to the possibility of standardizing the experience for all participants, individual differences in the usage of real, interactive AR applications (e.g., duration of engagement with specific components, focus on other aspects of the environment) and subsequently evoked random or systematic error can be decreased. Furthermore, the novelty effect that is often reported as a confounding variable in research on new technologies, such as AR [44], can probably be decreased when participants just imagine using real AR. Chang and colleagues, for example, found that an AR application increased learners' motivation in comparison to a learning video [45]. With a focus on cognitive aspects, removing potential motivational effects may increase the interpretability of the data. Additionally, disruptions or distractions that the operation of unknown technologies might bring are decreased. Through these advantages, effects can more securely be attributed to the studies' manipulations.

There are also apparent limitations of the approach. While the above-mentioned advantages all help with increased internal validity and the possibility of attributing effects to the experimental manipulations, those advantages also bring limitations when it comes to the question of the transferability of the results to the real usage of AR. The immersion principle by Mayer [31] states that immersive virtual environments may not always lead to better learning results because distractions of the environment may evoke an increase in necessary cognitive processing and eliminate motivational effects. This may also be the case for immersion in real-world environments that might provide motivational and atmospheric advantages but also increased cognitive demands due to many distractions. While removing these distracting and motivational aspects for research may help in decreasing confounding variables, those factors are not removed when using real AR applications. Furthermore, it is not clear if participants could really imagine being in the situations and if the perception of the real elements in the videos is comparable with the perception of those elements in physical reality. We did not ask the participants about their experience of imagining the situation taking place in the real world. Interaction and self-direction are key features in AR and every learner can have a different, individual experience when learning with AR, such that, for example, coherence-related factors may differ for each learner. In the video-based implementations the pace of the influx of information could not be regulated by the learners, although we did try to give them a lot of time to process information. This may have been more time than necessary for many participants, which may have led to boredom. While we do not know of cases in which this approach has already been used to evaluate AR, there has been research on general instructional videos in which virtual elements are added for extra information or guidance. In a study by Tsiatsos and colleagues, for example, virtual footprint symbols were added onto recorded videos of dancers for the teaching of Greek traditional dances [46]. These kinds of video overlays were classified as design patterns for video annotation and described as video augmentation through synchronized information overlays emphasizing the contents, linking to related content or containing additional relevant information themselves [47]. This may, in general, be an interesting alternative to fully location-based AR environments when their development or implementation is not possible, but the combination of real and virtual elements is an important focus of a learning area.

Both studies took place online with a supervising researcher over a (video/voice) call. The researcher was in direct contact at the beginning as well as the end and was available for questions during the study. Furthermore, participants were asked to minimize all distractions to account for the online situation. Still, especially concerning the learning phase, it is not completely clear how seriously the participants watched the videos, and external distractions could not be controlled for. In study two it might even be the case that participants did not listen to the sounds. They were asked to use headphones and adjust their volume based on a sound before the video, but they might still have turned sounds off or removed their headphones afterwards. In a laboratory those factors could have been better controlled.

Another limitation is based on the usage of questionnaires for measuring cognitive load and task load. While this method provided us with differentiated outcomes based on the different subconcepts, such as the different types of cognitive load, for future research the addition of objective, continuous measures of cognitive load may be considered. Retrospective measures cannot capture changes in load during the learning task, which may be particularly interesting for learning environments in which the focus on learning materials shifts over time due to navigational periods when looking for the next material (e.g., walking from one to the next plant or bird) or when environmental distracting variables change. Physical, continuous measures, such as heart rate or eye movement, may be able to catch those changes without disrupting learning processes and should be considered as complementary methods in the future. Objective measures can be valuable for verifying subjective results, which are sometimes biased because not all learners may rate cognitive load in an expected way [48].

4.2. Future Research

Implementing different multimedia design principles in AR adds new features to the mix, including the dimension of reality. As seen in the presented studies, following the principles of spatial contiguity and coherence was possible with combined presentations of virtual and real elements, and at least descriptively led to partly expected results. Based on the limitations described above, evaluating the implementations in more complex, naturalistic environments with real AR-based learning material and with more complex learning topics is necessary for their confirmation. In general, as already suggested in meta-reviews by [26–29], Mayer's CTML and following its multimedia design principles are important approaches used in multiple studies on AR in education. Although we focused on those in the current study, spatial contiguity and coherence are not the only principles that may play an important role in AR-based learning and instruction.

The *multimedia principle*, which states that text and pictures should be combined in instructional material for better learning [49], is in general important due to the multiple representations that can be used in AR. Elements of the real world are often pictorial, but both textual and pictorial virtual elements can be added, as seen in the presented studies. The combination of virtual texts and pictures with different real-world locations and elements might be interesting for research concerning specific multimedia mechanisms. In study two, the focus was on the *coherence principle* with additional more or less relevant sounds. While sounds are part of the third version of the coherence principle, the first version focuses more on visual aspects that distract learners' attention away from relevant visual elements [15]. In AR, additional information that may not be necessary for learning but can lead to a different context or atmosphere is often part of learning experiences due to the environment in which the learning takes place. Relevant visual backgrounds give a context and should thus increase immersion and motivation, but may distract from the learning material and use additional resources for unnecessary processing. A potentially detrimental seductive details effect concerning visual aspects in AR experiences should be examined in future studies.

When pictorial and textual information are combined, it is always good to take a look at the *modality principle* [50], and thus the combination of pictures and spoken words. AR can be used with a lot of different media, such that the addition of narration to pictures may bring a great potential to follow the modality principle without a lot of effort here. While AR is often described as a visual experience, the inclusion of auditory elements is possible and should be examined more closely. When narrations are added to AR-based experiences, following the *temporal contiguity principle* [51] also seems quite straightforward for the combination of virtual elements with real-world environments. The spatial integration of information could be easily achieved in the representations in study one. The temporal contiguity principle expands this idea towards narrations, suggesting that spoken text is provided exactly when corresponding pictorial virtual elements are shown and, in an AR-specific case, exactly when looking at corresponding physical elements in the real world. Here, it should be distinguished between information where one part of the elements is only understandable when the other part is also available (e.g., verbally describing processes in the real world that are not apparent or visible) and information where each part is understandable on its own, but where adding the other part may give more information (e.g., naming and characterizing objects in the real world through spoken text).

The focus of the current studies is on very specific features of instructional design of AR applications. In reality, AR applications are often more complex and incorporate more features than just a short presentation of spatially and contextually embedded information. In the taxonomy that Mystakidis and colleagues developed based on a systematic review, five clusters of instructional methods in AR-based learning interventions are described and the approaches in the current studies best fit into the most passive, teacher-centered category where presentation and observation are in focus [25]. The other four clusters describe methods that focus on activities that are more or completely learner-controlled, including project-based and experiential learning. In order to establish an empirical basis for the integration of multimedia principles into those more complex and interactive instructional methods, research in more authentic, naturalistic settings in which learners can interact with the representations is necessary. An important aspect that distinguishes AR settings from non-AR multimedia learning is the focus on being immersed and feeling present within a specific environment as an important affordance of the experience [20]. While Mayer's immersion principle states that immersion does not necessarily lead to improved learning, the motivational aspects of immersion are recognized [31]. Its interaction with the application of the other multimedia principles may provide an interesting research area. The feeling of presence is also relevant in VR settings, and maybe even more so in remote collaboration through VR, where it was suggested to be included into a taxonomy of computer-supported cooperative work (CSCW) [52]. In further steps, collaborative or cooperative learning settings, which often come naturally with more interactive and project-based settings, may be the focus of future research on the implementation of multimedia principles in AR-based learning.

Concerning the methods used, for future examinations of implementations of multimedia design principles in AR, it seems to be important to further take into account different types of cognitive load (e.g., extraneous and germane cognitive load), which compared to each other descriptively showed different patterns in both studies. Additionally, the different subconstructs of task load either supported each other, as in study one, or also showed different, more differentiated patterns, as in study two. Adding this measure of the NASA TLX, which is often used in research on AR [32], can provide researchers with more detailed information on potential effects. Furthermore, both studies showed that different kinds of knowledge and thus different test items (i.e., combining pictures and text; text only) should be considered for learning outcome examination. When learning with AR in the outside world, the real-world environment is often a pictorial and not a textual representation. Additional visual virtual materials can be both pictures and text, and multimedia effects should be further examined including the factor of virtual and physical realities in AR.

In general, we can conclude that AR brings many new opportunities concerning the application of multimedia design principles in the context of physical and virtual elements. The characteristics of AR, including spatiality, interactivity and contextuality, can help in considering AR-specific potentials when implementing those principles. Spatial contiguity and coherence, as two well-studied and -supported principles, were the focus of the current paper, although more should be implemented and investigated in future studies.

Author Contributions: Conceptualization, J.M.K. and D.B.; Methodology, J.M.K.; Formal Analysis, J.M.K.; Investigation, J.M.K.; Data Curation, J.M.K.; Writing—Original Draft Preparation, J.M.K. and D.B.; Writing—Review and Editing, J.M.K. and D.B.; Visualization, J.M.K.; Supervision, D.B. and J.M.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The studies were conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of the Department of Computer Science and Applied Cognitive Science of the University of Duisburg-Essen (study one: ID 2011PFBS7216, 26.11.2020; study two: ID 2012PFKL8474, 7 December 2020).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the studies.

Data Availability Statement: Publicly available datasets were analyzed in this study. These datasets can be found here: study one: osf.io/5gzyv/; study two: osf.io/qtnr7/.

Acknowledgments: We thank Laura Kusber and Stefan Baudeck for creating the video-based material and collecting the data in the studies.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Result of Levene's test for all variables.

Study One—Levene's Test	*F*-Value	df1	df2	*p*
H1.1a: extraneous cognitive load	0.09	1	78	0.764
H1.1b: germane cognitive load	1.00	1	78	0.320
H1.2a: mental demand	0.64	1	78	0.427
H1.2b: physical demand	4.07	1	78	0.047 *
H1.2c: temporal demand	0.53	1	78	0.469
H1.2d: performance	0.18	1	78	0.671
H1.2e: effort	0.15	1	78	0.696
H1.2f: frustration	1.25	1	78	0.266
H1.3: knowledge	0.78	1	78	0.379
Study two—Levene's Test	***F*-Value**	**df1**	**df2**	***p***
H2.1a: extraneous cognitive load	0.53	2	127	0.592
H2.1b: germane cognitive load	0.54	2	127	0.585
H2.2a: mental demand	0.20	2	127	0.819
H2.2b: physical demand	0.93	2	127	0.397
H2.2c: temporal demand	0.46	2	127	0.633
H2.2d: performance	0.06	2	127	0.944
H2.2e: effort	0.46	2	127	0.633
H2.2f: frustration	0.05	2	127	0.952
H2.3: knowledge	1.48	2	127	0.232

Note. * $p < 0.05$.

References

1. Mayer, R. 2 Science of Learning: Determining How Multimedia Learning Works. In *Multimedia Learning*; Cambridge University Press: Cambridge, UK, 2020; pp. 29–62, ISBN 978-1-316-94135-5.
2. Radu, I. Augmented Reality in Education: A Meta-Review and Cross-Media Analysis. *Pers. Ubiquitous Comput.* **2014**, *18*, 1533–1543. [CrossRef]
3. Krüger, J.M.; Buchholz, A.; Bodemer, D. Augmented Reality in Education: Three Unique Characteristics from a User's Perspective. In Proceedings of the 27th International Conference on Computers in Education, Kenting, Taiwan, 2–6 December 2019; Chang, M., So, H.-J., Wong, L.-H., Yu, F.-Y., Shih, J.L., Eds.; Asia-Pacific Society for Computers in Education: Kenting, Taiwan, 2019; pp. 412–422.
4. Garzón, J. An Overview of Twenty-Five Years of Augmented Reality in Education. *Multimodal Technol. Interact.* **2021**, *5*, 37. [CrossRef]
5. Mayer, R.E. *Multimedia Learning*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2009; ISBN 978-0-511-81167-8.
6. Baddeley, A.D. *Essentials of Human Memory*; Cognitive psychology; Psychology Press: Hove, UK, 1999; ISBN 978-0-86377-545-1.
7. Paivio, A. *Mental Representations: A Dual Coding Approach*; Oxford Psychology Series; Oxford University Press: Oxford, UK; Clarendon Press: New York, NY, USA, 1986; ISBN 978-0-19-503936-8.
8. Sweller, J.; van Merrienboer, J.J.G.; Paas, F.G.W.C. Cognitive Architecture and Instructional Design. *Educ. Psychol. Rev.* **1998**, *10*, 251–296. [CrossRef]
9. Sweller, J.; van Merriënboer, J.J.G.; Paas, F.G.W.C. Cognitive Architecture and Instructional Design: 20 Years Later. *Educ. Psychol. Rev.* **2019**, *31*, 261–292. [CrossRef]
10. Mayer, R. 13 Modality Principle. In *Multimedia Learning*; Cambridge University Press: Cambridge, UK, 2020; pp. 281–300. ISBN 978-1-316-94135-5.
11. Hong, J.Y.; He, J.; Lam, B.; Gupta, R.; Gan, W.-S. Spatial Audio for Soundscape Design: Recording and Reproduction. *Appl. Sci.* **2017**, *7*, 627. [CrossRef]
12. Mayer, R. 9 Spatial Contiguity Principle. In *Multimedia Learning*; Cambridge University Press: Cambridge, UK, 2020; pp. 207–226. ISBN 978-1-316-94135-5.
13. Ayres, P.; Sweller, J. The Split-Attention Principle in Multimedia Learning. In *The Cambridge Handbook of Multimedia Learning*; Mayer, R.E., Ed.; Cambridge University Press: Cambridge, UK, 2014; pp. 206–226. ISBN 978-1-139-54736-9.
14. Schroeder, N.L.; Cenkci, A.T. Spatial Contiguity and Spatial Split-Attention Effects in Multimedia Learning Environments: A Meta-Analysis. *Educ. Psychol. Rev.* **2018**, *30*, 679–701. [CrossRef]
15. Mayer, R. 6 Coherence Principle. In *Multimedia Learning*; Cambridge University Press: Cambridge, UK, 2020; pp. 143–165, ISBN 978-1-316-94135-5.

16. Jung, T.; tom Dieck, M.C.; Lee, H.; Chung, N. Relationships among Beliefs, Attitudes, Time Resources, Subjective Norms, and Intentions to Use Wearable Augmented Reality in Art Galleries. *Sustainability* **2020**, *12*, 8628. [CrossRef]
17. Han, K.; Park, K.; Choi, K.-H.; Lee, J. Mobile Augmented Reality Serious Game for Improving Old Adults' Working Memory. *Appl. Sci.* **2021**, *11*, 7843. [CrossRef]
18. Cabero-Almenara, J.; Barroso-Osuna, J.; Martinez-Roig, R. Mixed, Augmented and Virtual, Reality Applied to the Teaching of Mathematics for Architects. *Appl. Sci.* **2021**, *11*, 7125. [CrossRef]
19. Akçayır, M.; Akçayır, G. Advantages and Challenges Associated with Augmented Reality for Education: A Systematic Review of the Literature. *Educ. Res. Rev.* **2017**, *20*, 1–11. [CrossRef]
20. Wu, H.-K.; Lee, S.W.-Y.; Chang, H.-Y.; Liang, J.-C. Current Status, Opportunities and Challenges of Augmented Reality in Education. *Comput. Educ.* **2013**, *62*, 41–49. [CrossRef]
21. Bacca, J.; Baldiris, S.; Fabregat, R.; Graf, S. Kinshuk Augmented Reality Trends in Education: A Systematic Review of Research and Applications. *Educ. Technol. Soc.* **2014**, *17*, 133–149.
22. Goff, E.E.; Mulvey, K.L.; Irvin, M.J.; Hartstone-Rose, A. Applications of Augmented Reality in Informal Science Learning Sites: A Review. *J. Sci. Educ. Technol.* **2018**, *27*, 433–447. [CrossRef]
23. Bower, M.; Howe, C.; McCredie, N.; Robinson, A.; Grover, D. Augmented Reality in Education–Cases, Places and Potentials. *Educ. Media Int.* **2014**, *51*, 1–15. [CrossRef]
24. Mayer, R. 1 The Promise of Multimedia Learning. In *Multimedia Learning*; Cambridge University Press: Cambridge, UK, 2020; pp. 3–28, ISBN 978-1-316-94135-5.
25. Mystakidis, S.; Christopoulos, A.; Pellas, N. A Systematic Mapping Review of Augmented Reality Applications to Support STEM Learning in Higher Education. *Educ. Inf. Technol.* **2021**. [CrossRef]
26. Sommerauer, P.; Müller, O. Augmented Reality for Teaching and Learning-a Literature Review on Theoretical and Empirical Foundations. In Proceedings of the Twenty-Sixth European Conference on Information Systems (ECIS2018), Portsmouth, UK, 23–28 June 2018.
27. Da Silva, M.M.O.; Teixeira, J.M.X.N.; Cavalcante, P.S.; Teichrieb, V. Perspectives on How to Evaluate Augmented Reality Technology Tools for Education: A Systematic Review. *J. Braz. Comput. Soc.* **2019**, *25*, 3. [CrossRef]
28. Garzón, J.; Baldiris, S.; Gutiérrez, J.; Pavón, J. How Do Pedagogical Approaches Affect the Impact of Augmented Reality on Education? A Meta-Analysis and Research Synthesis. *Educ. Res. Rev.* **2020**, *31*, 100334. [CrossRef]
29. Buchner, J.; Buntins, K.; Kerres, M. The Impact of Augmented Reality on Cognitive Load and Performance: A Systematic Review. *J. Comput. Assist. Learn.* **2021**. [CrossRef]
30. Altmeyer, K.; Kapp, S.; Thees, M.; Malone, S.; Kuhn, J.; Brünken, R. The Use of Augmented Reality to Foster Conceptual Knowledge Acquisition in STEM Laboratory Courses—Theoretical Background and Empirical Results. *Br. J. Educ. Technol.* **2020**. [CrossRef]
31. Mayer, R. 18 Immersion Principle. In *Multimedia Learning*; Cambridge University Press: Cambridge, UK, 2020; ISBN 978-1-316-94135-5.
32. Buchner, J.; Buntins, K.; Kerres, M. A Systematic Map of Research Characteristics in Studies on Augmented Reality and Cognitive Load. *Comput. Educ. Open* **2021**, *2*, 100036. [CrossRef]
33. Wigfield, A.; Eccles, J.S. Expectancy–Value Theory of Achievement Motivation. *Contemp. Educ. Psychol.* **2000**, *25*, 68–81. [CrossRef] [PubMed]
34. Klepsch, M.; Schmitz, F.; Seufert, T. Development and Validation of Two Instruments Measuring Intrinsic, Extraneous, and Germane Cognitive Load. *Front. Psychol.* **2017**, *8*, 1997. [CrossRef] [PubMed]
35. Hart, S.G.; Staveland, L.E. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology*; Elsevier: Amsterdam, The Netherlands, 1988; Volume 52, pp. 139–183, ISBN 978-0-444-70388-0.
36. Hart, S.G. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2006**, *50*, 904–908. [CrossRef]
37. Delacre, M.; Lakens, D.; Leys, C. Why Psychologists Should by Default Use Welch's *t*-Test Instead of Student's *t*-Test. *Rips* **2017**, *30*, 92–101. [CrossRef]
38. Seufert, T. Supporting Coherence Formation in Learning from Multiple Representations. *Learn. Instr.* **2003**, *13*, 227–237. [CrossRef]
39. Seufert, T.; Brünken, R. Supporting Coherence Formation in Multimedia Learning. In Proceedings of the Instructional design for effective and enjoyablecomputer-supported learning. In Proceedings of the First Joint Meeting of the EARLI SIGs Instructional Design and Learning and Instruction with Computers, Tübingen, Germany, 7–9 July 2004; Gerjets, P., Kirschner, P., Elen, J., Joiner, R., Eds.; Knowledge Media Research Center: Tübingen, Germany, 2004; pp. 138–147.
40. Dunleavy, M.; Dede, C.; Mitchell, R. Affordances and Limitations of Immersive Participatory Augmented Reality Simulations for Teaching and Learning. *J. Sci. Educ. Technol.* **2009**, *18*, 7–22. [CrossRef]
41. Moreno, R.; Mayer, R.E. Interactive Multimodal Learning Environments. *Educ. Psychol. Rev.* **2007**, *19*, 309–326. [CrossRef]
42. Plass, J.L.; Kaplan, U. Chapter 7-Emotional Design in Digital Media for Learning. In *Emotions, Technology, Design, and Learning*; Tettegah, S.Y., Gartmeier, M., Eds.; Emotions and Technology; Academic Press: San Diego, CA, USA, 2016; pp. 131–161, ISBN 978-0-12-801856-9.
43. Park, B.; Flowerday, T.; Brünken, R. Cognitive and Affective Effects of Seductive Details in Multimedia Learning. *Comput. Hum. Behav.* **2015**, *44*, 267–278. [CrossRef]

44. Derby, J.L.; Chaparro, B.S. The Challenges of Evaluating the Usability of Augmented Reality (AR). *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2021**, *65*, 994–998. [CrossRef]
45. Chang, R.-C.; Chung, L.-Y.; Huang, Y.-M. Developing an Interactive Augmented Reality System as a Complement to Plant Education and Comparing Its Effectiveness with Video Learning. *Interact. Learn. Environ.* **2016**, *24*, 1245–1264. [CrossRef]
46. Tsiatsos, T.; Stavridou, E.; Grammatikopoulou, A.; Douka, S.; Sofianidis, G. Exploiting Annotated Video to Support Dance Education. In Proceedings of the 2010 Sixth Advanced International Conference on Telecommunications, Barcelona, Spain, 9–15 May 2010; pp. 100–105.
47. Seidel, N. Interaction Design Patterns for Spatio-Temporal Annotations in Video Learning Environments. In *Proceedings of the Proceedings of the 20th European Conference on Pattern Languages of Programs*; Association for Computing Machinery: New York, NY, USA, 2015; pp. 1–21.
48. Zu, T.; Munsell, J.; Rebello, N.S. Subjective Measure of Cognitive Load Depends on Participants' Content Knowledge Level. *Front. Educ.* **2021**, *6*, 647097. [CrossRef]
49. Mayer, R. 5 Multimedia Principle. In *Multimedia Learning*; Cambridge University Press: Cambridge, UK, 2020; pp. 117–138, ISBN 978-1-316-94135-5.
50. Mayer, R.E.; Pilegard, C. Principles for Managing Essential Processing in Multimedia Learning: Segmenting, Pre-Training, and Modality Principles. In *The Cambridge Handbook of Multimedia Learning*; Mayer, R., Ed.; Cambridge University Press: Cambridge, UK, 2014; pp. 316–344, ISBN 978-1-139-54736-9.
51. Mayer, R.E. Temporal Contiguity Principle. In *Multimedia Learning*; Mayer, R.E., Ed.; Cambridge University Press: Cambridge, UK, 2009; pp. 153–169, ISBN 978-0-511-81167-8.
52. Cruz, A.; Paredes, H.; Morgado, L.; Martins, P. Non-Verbal Aspects of Collaboration in Virtual Worlds: A CSCW Taxonomy-Development Proposal Integrating the Presence Dimension. *JUCS-J. Univers. Comput. Sci.* **2021**, *27*, 913–954. [CrossRef]

DOI: 10.17185/duepublico/78994
URN: urn:nbn:de:hbz:465-20231018-120400-8