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

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.

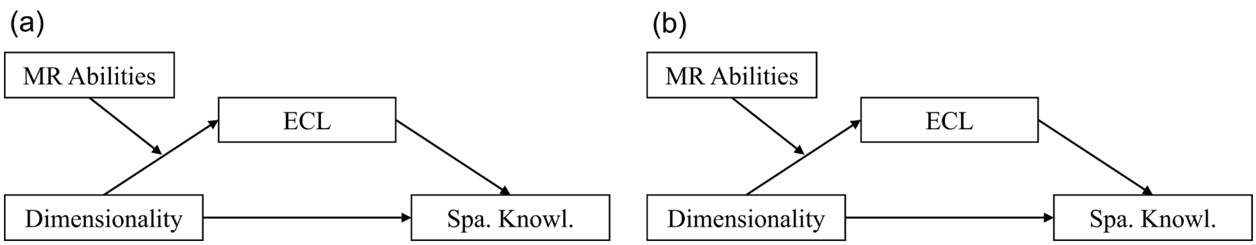

Fig. 1. Moderated mediation in H3a (a) and RQ3b (b). In H3a, the influence of dimensionality on ECL in the mediation model proposed in H2d is hypothesized to be moderated by mental rotation abilities. In RQ3b, the influence of dimensionality on GCL in the mediation model proposed in H2e is hypothesized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

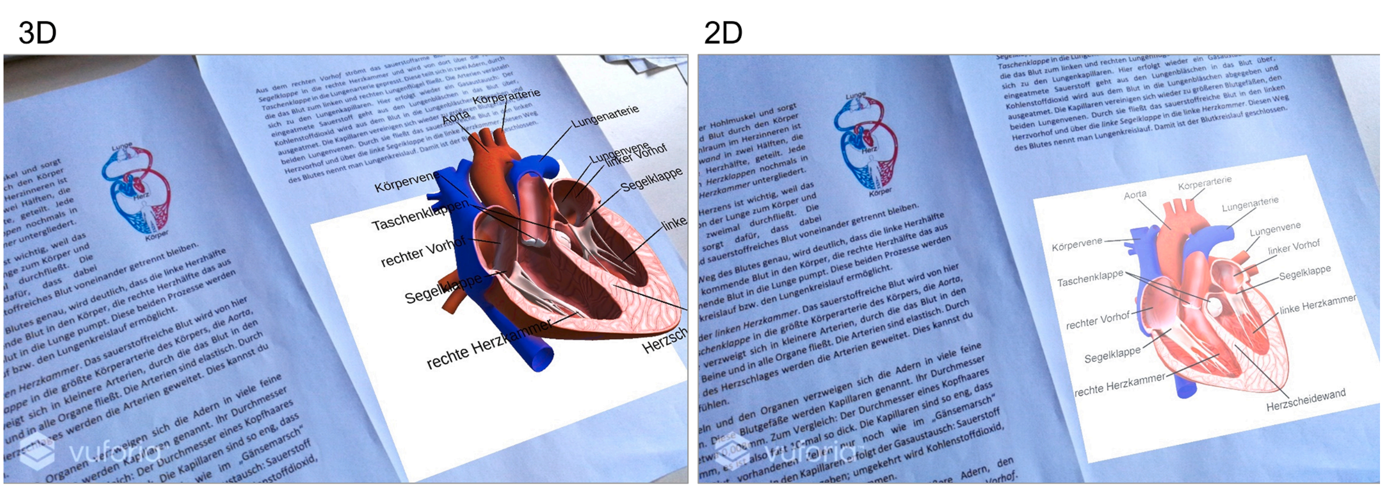

Fig. 2. Screenshot from the application in the 3D condition (left) and the 2D condition (right).

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

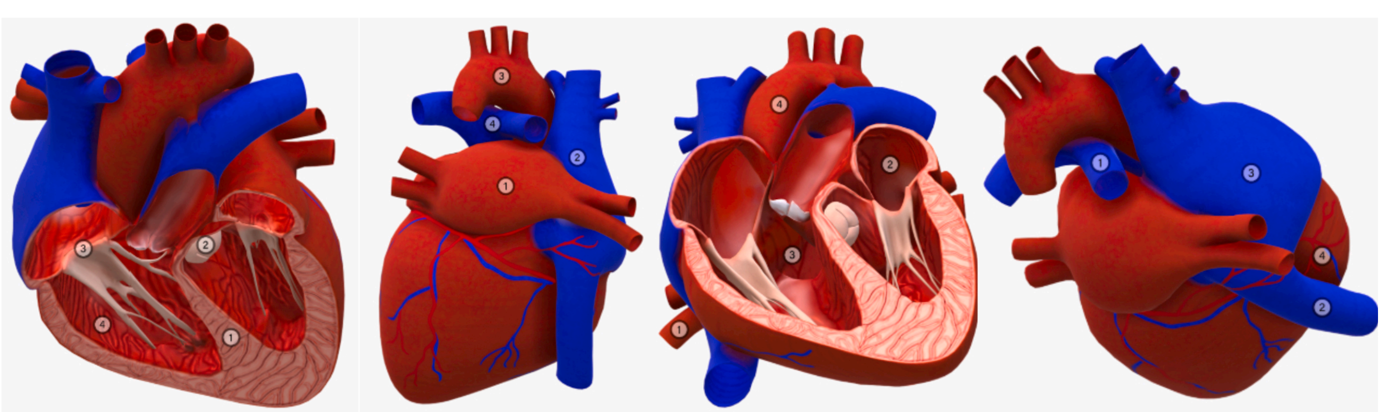

Fig. 3. Pictures used for the component part of the knowledge test.

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].

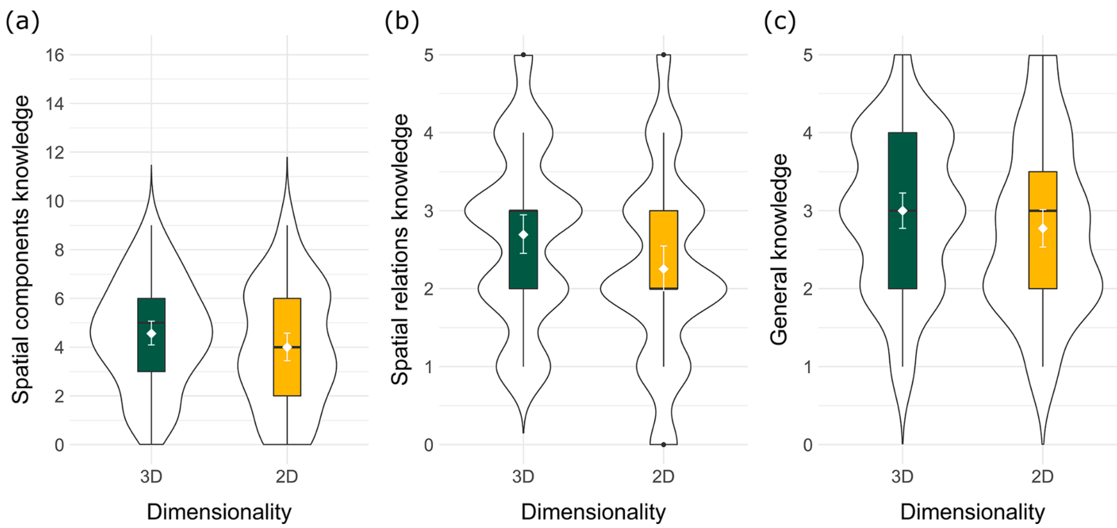

Fig. 4. Distribution of (a) spatial components knowledge, (b) spatial relations knowledge and (c) general knowledge test scores split by group [boxplot with IQR (filled), mean with bootstrapped 95% CI (white), violin plot for distribution (outline)].

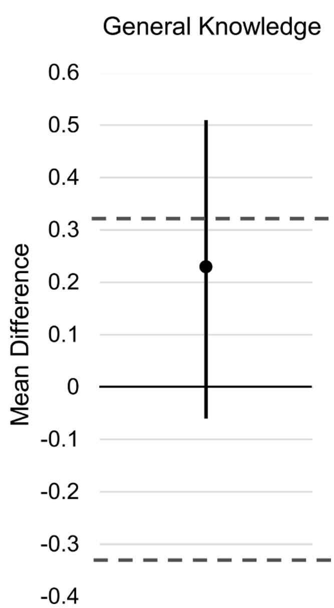

Fig. 5. Result TOST equivalence test for general knowledge showing the 90% CI for raw mean difference [-0.06, 0.51] and raw bounds at +/-0.33. The 90% CI ends above the upper bound, showing no equivalence inside these bounds.

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.

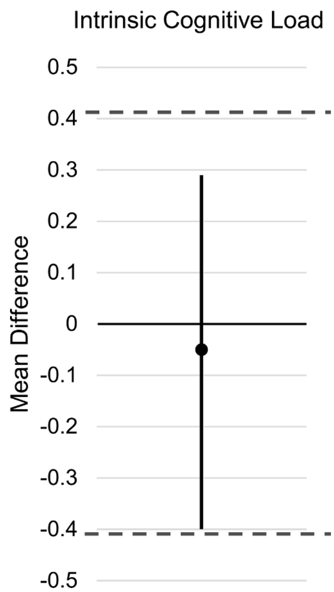

Fig. 6. Result TOST equivalence test for ICL showing the 90% CI for raw mean difference [-0.40, 0.29] and raw bounds at +/-0.41. The 90% CI ends below the upper bound and above the lower bound, showing equivalence inside these bounds.

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

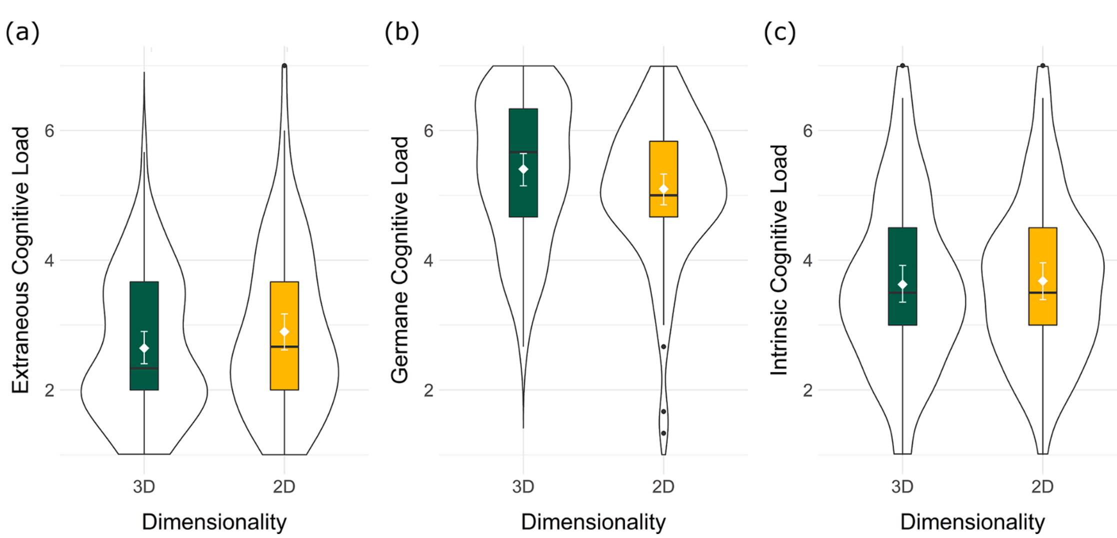

Fig. 7. Distribution of (a) ECL, (b) GCL and (c) ICL subscale scores split by group [boxplot with IQR (filled), mean with bootstrapped 95% CI (white), violin plot for distribution (outline)].

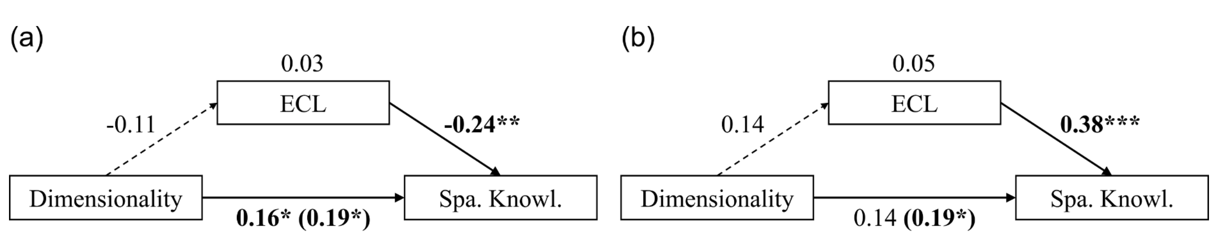

Fig. 8. Meditation model for (a) H2d and (b) H2e including completely standardized effect sizes (β) and significance levels for all effects, including the indirect effect. Significance levels: * $p < .05$, ** $p < .01$, *** $p < .001$.

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

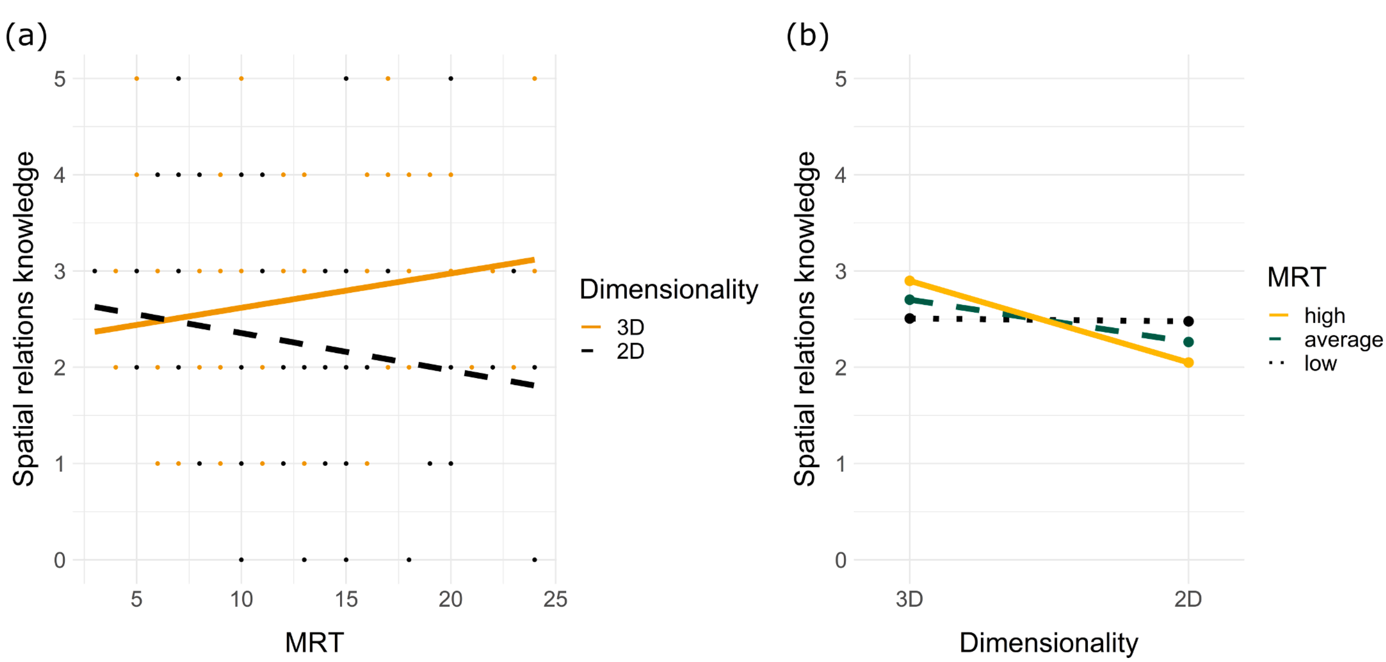

Fig. 9. (a) Interaction effect between dimensionality of visualization and MRT score on spatial relations knowledge, and (b) simple slopes of the MRT score levels concerning the effect of the visualization on spatial relations knowledge.

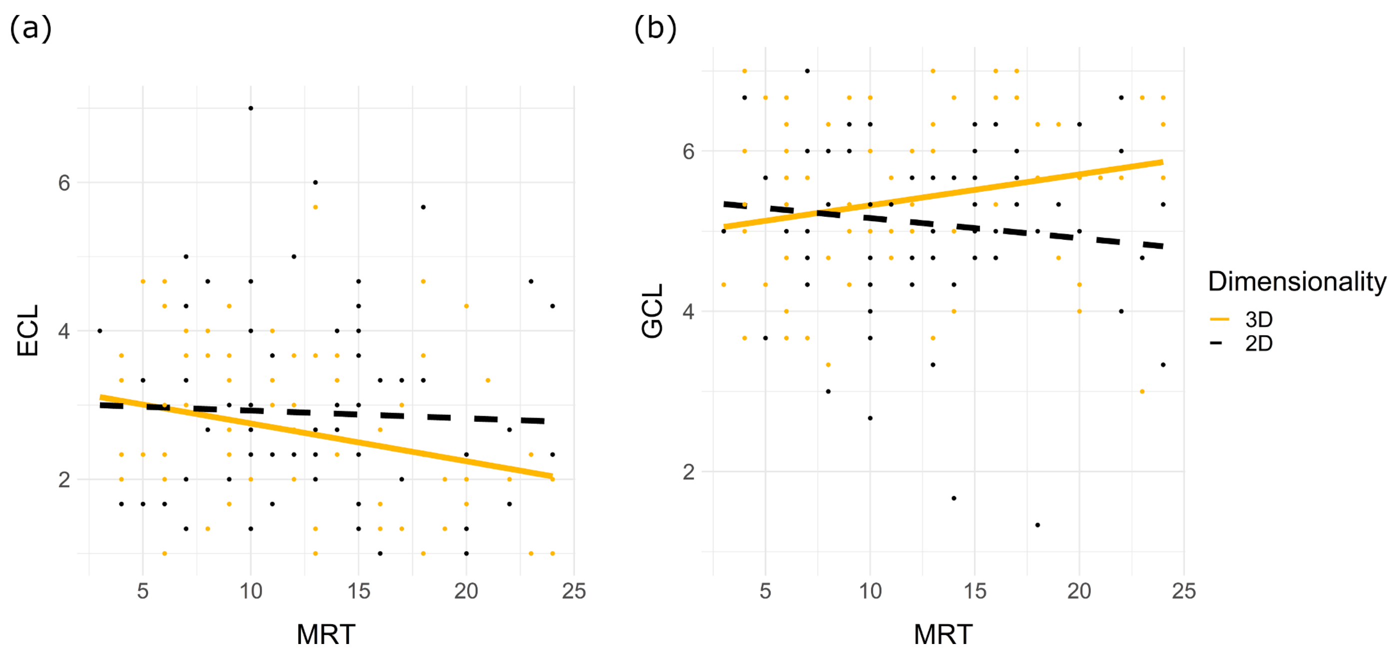

Fig. 10. Scatterplot including interaction of MRT and visualization on (a) ECL and (b) GCL.

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.