Journal of Computer Assisted Learning

WILEY

Journal of Computer Assisted Learning

REVIEW ARTICLE OPEN ACCESS

Unlocking Augmented Reality Learning Design Based on Evidence From Empirical Cognitive Load Studies—A Systematic Literature Review

Jana Gonnermann-Müller[1,2] | Jule M. Krüger[3]

[1]Chair of Business Informatics, Processes and Systems, University of Potsdam, Potsdam, Germany | [2]Weizenbaum Institute for the Networked Society, Berlin, Germany | [3]Chair for Digital Education, University of Potsdam, Potsdam, Germany

Correspondence: Jana Gonnermann-Müller (jana.gonnermann@wi.uni-potsdam.de)

Received: 22 February 2024 | **Revised:** 9 October 2024 | **Accepted:** 1 November 2024

Funding: This research was funded by the Bundesministerium für Bildung und Forschung and the BMBF Project 'German Internet Institute' (16DII137).

Keywords: augmented reality | cognitive load | design guidelines | design principles | instructional design | workload

ABSTRACT

Background: Despite the numerous positive effects of augmented reality (AR) on learning, previous research has shown ambiguous results regarding the cognitive demand on the learner arising from, for example, the overlay of virtual elements or novel interaction techniques. At the same time, the number of evidence-based guidelines on designing AR is limited or focuses on global effects, primarily relying on media comparison studies, whose validity is criticised.

Objective: To guide the meaningful design of learning and training settings, this paper systematically reviews empirical research on AR design and synthesises the findings to develop evidence-based recommendations for designing AR systems considering cognitive load.

Methods: We conducted a systematic literature review, initially screening 810 distinct papers and ultimately analysing findings from 27 publications, which report on 29 distinct experimental studies. This selection was based on rigorously defined inclusion and exclusion criteria, adhering to the PRISMA guidelines.

Results and Conclusion: The central value of this paper is the aggregation of existing evidence from empirical studies, resulting in 15 recommendations for AR design based on six design dimensions: Spatiality-related, Interaction-related, Contextuality-related, Content-related, Guidance-related and Display Selection. Additionally, with three points for future research, this systematic literature review, first, stresses the need for more empirical evidence and value-added studies. Second, learner characteristics that might influence cognitive load in AR-based learning should be examined. Third, it advocates for the inclusion of measurements beyond the NASA-TLX, and including more physiological measurements (e.g., eye-tracking, EEG) to enhance the applicability of the results for learning and training situations.

1 | Introduction

Augmented reality (AR) offers a wide range of possibilities for learning and training since it presents additional virtual information, cues and 3D objects directly into the physical learning environment. Devices that enable these functions include AR glasses, tablets, or projections. The learner can interact with the virtual representations in real-time (Azuma 1997; Milgram and Kishino 1994) and directly apply the learning content in the physical space, supporting knowledge and skill development in realistic environments (Chu et al. 2019). Several positive effects of AR have been identified, including improved attitudes towards learning, learning achievements, satisfaction and flow (e.g., Buchner, Buntins, and Kerres 2022; Hsu 2019; Lin

Summary

- What is already known about this topic
  - Augmented reality (AR) positively affects learning but can also add cognitive load.
  - AR-specific design guidelines to foster a meaningful AR design for learning are rare.
  - Current research on AR requires a systematisation of empirical evidence to identify which specific aspects of AR contribute to cognitive load and learning.
- What this paper adds
  - This review systematically evaluates existing literature to assess the impact of AR attributes on cognitive load in educational and training contexts and postulates AR-specific recommendations.
  - The article identifies three points of future research targeting the need for more empirical research and value-added studies.
- Implications for practice
  - The central value of this paper is the aggregation of existing evidence from empirical cognitive load studies, resulting in 15 recommendations for AR instructional design on six design dimensions: Spatiality-related, Interaction-related, Contextuality-related, Content-related, Guidance-related and Display Selection.
  - The findings offer practical guidance for instructional designers and practitioners in AR-based learning and higher education training.

and Yu 2023; Yu 2023). While the potential to decrease mental demand through integrated representations in AR is high (Altmeyer et al. 2020), the empirical results are not yet conclusive, showing positive (Lin and Yu 2023) and negative (Bautista, Maradei, and Pedraza 2023) impacts.

Meaningful design and instructions are necessary to leverage the benefits of AR concerning cognitive demand (Buchner, Buntins, and Kerres 2021, 2022). The design of AR-based environments can be based on general multimedia principles (Krüger and Bodemer 2022; Mayer 2021), but more AR-specific recommendations for instructional design are still missing. Abstract guidelines for using AR based on media comparison studies (e.g., 'Recommending using AR for learning of applied knowledge') are common, but concrete guidelines (e.g., 'AR should integrate attribute X in manner Y to foster learning') are sparse. To define the first design recommendations based on empirical evidence, this paper systematically reviews the literature on how instructional AR design attributes affect cognitive load and learning or training performance. Based on this review, the goal is to cluster identified effects and formulate recommendations for AR design. In the following, important constructs and related research will be described, leading to the research questions for this systematic review.

1.1 | Augmented Reality

AR is a technological paradigm that overlays computer-generated information onto the physical environment, supporting the learner with additional information and instructions (Daling and Schlittmeier 2022; Hou et al. 2013). It enables dynamic interactions by juxtaposing digital and physical realms, contributing to an enriched and contextually relevant user experience. Rauschnabel et al. (2022) distinguish AR use cases through AR devices (Stationary, Mobile, Wearable, On-body and In-body), AR enablers (App, Web AR, Platform AR and Stationary setup) and AR displays (AR mirror, video see-through, optical see-through and projection-based AR). The most common AR devices and displays include head-mounted displays (HMD), handheld displays (HHD) and spatial displays (Carmigniani et al. 2011). HMDs, as 'wearable' AR devices, sit on the user's head to deliver virtual content into the field of view, freeing their hands for concurrent tasks. They can utilise optical-see-through (OST) or video-see-through (VST) techniques as classified by Rauschnabel et al. (2022). OST employs transparent displays to overlay virtual objects in the user's field of view, while VST captures the physical environment through cameras to digitally integrate it with virtual elements on screens. In contrast, HHDs are 'mobile' AR devices and usually VST displays that users hold in their hands like smartphones and tablets. Spatial augmented reality (SAR) is usually enabled through 'stationary' AR devices, external to the user and often seamlessly embedded into the natural environment through projection devices (Carmigniani et al. 2011).

AR devices and display technologies augment reality through virtual content. With AR, information can be integrated into the work process, delivering contextual insights and real-time assistance for applying learning content by overlaying or connecting the real-world setting with virtual objects (Azuma 1997; Milgram and Kishino 1994). Research on AR has increased since it emerged as a promising solution for delivering supplementary information and crucial cues essential for process-integrated learning. The effectiveness of learning with AR is studied in application contexts and occupational settings like industry, assembly (Howard and Davis 2023; Wang et al. 2022) and medical education and training (Fischer et al. 2016; Lu et al. 2020; Uruthiralingam and Rea 2020). Systematic reviews show that a general positive effect of AR on learning achievements can be found when comparing it to more traditional learning media, showing effect sizes (Hedge's *g*) between 0.49 and 0.92 depending on the type of measurements and achievements (Chang et al. 2022; Garzón et al. 2020; Garzón and Acevedo 2019; Xu et al. 2022; Zhang et al. 2022).

However, concerning information processing and related factors like mental demand, the outcomes are inconclusive. On the one hand, additional information presented via in-place and image recognition may contribute to less demand than traditional applications (Lin and Yu 2023). For example, AR can simplify information processing within a physical learning environment by integrating elements, compared to instruction on a separate screen, where the learner must extract the required information while integrating it with the application context (Altmeyer et al. 2020). On the other hand, AR attributes may introduce distractions and overwhelm learners. Integrating and simultaneously processing virtual and natural information was perceived as challenging (Bautista, Maradei, and Pedraza 2023; Uruthiralingam and Rea 2020). Another pitfall is the effort required to process information while performing tasks, for example, when integrating virtual elements and unfamiliar interaction techniques, such as gestures or voice commands. This could lead to additional demand since the learner must simultaneously process various real and virtual stimuli. These findings contrast with studies indicating that AR can provide users with cognitive support, guidance and assistance (Daling and Schlittmeier 2022; Hou et al. 2013). To understand general learning outcomes, it is thus important to investigate the cognitive processes and demands that environments may evoke.

1.2 | Cognitive Load and Workload

Different terms among different disciplines describe human cognitive demand, like cognitive workload, cognitive load, or mental workload (Kosch et al. 2023). In education and learning research, cognitive demand is often described using cognitive load theory (CLT). CLT is based on assumptions about human cognitive architecture and how information is processed, stored and transferred between working and long-term memory (Bannert 2002; Gerjets, Scheiter, and Cierniak 2009; Kalyuga 2023; Sweller, van Merriënboer, and Paas 2019). It states that humans' abilities to process novel information are restricted due to the limited capacity of the working memory. The theory distinguishes three dimensions of cognitive load: intrinsic cognitive load (ICL), germane cognitive load (GCL) and extraneous cognitive load (ECL) (Bannert 2002; Gerjets, Scheiter, and Cierniak 2009; Kalyuga 2023; Sweller, van Merriënboer, and Paas 2019). ICL refers to the complexity of the learning material in interaction with learners' skills to recognise schemata and structures. It is associated with prior expertise and knowledge influencing how new information is processed. GCL describes the load necessary for a successful learning process, integrating newly perceived information within working memory and with prior knowledge from long-term memory. ECL is influenced by external stimuli like instructional and didactical design and attributes of the learning environment (Mayer 2021; Sweller, van Merriënboer, and Paas 2019) and can be decreased by designing effective instructions and reducing unnecessary or redundant information. Therefore, ECL is defined as negatively associated with learning performance and should be considered when evaluating instruction and interface design (Bannert 2002; Kalyuga 2023; Mayer 2021; Sweller, van Merriënboer, and Paas 2019; van Merriënboer, Jelsma, and Paas 1992). During learning activities, most cognitive capacity should be focused on the learning content (ICL) and not on unnecessary information or instructions (ECL). Design principles based on CLT, for example, in multimedia learning, aim to reduce ECL and keep ICL within an appropriate level to support efficient learning of novel information without overstraining the learner (Gerjets, Scheiter, and Cierniak 2009; Mayer 2021).

Besides CLT, other conceptualisations of load have been used in research on AR. Buchner, Buntins, and Kerres (2021) found in a mapping review that most studies on AR used the NASA Task Load Index (NASA-TLX) conceptualisation developed by Hart and Staveland (1988). It refers to six global constructs of workload, including mental, physical and temporal demand as task-based constructs, perceived performance and effort as behaviour-based constructs and frustration as a person-based construct (Hart and Staveland 1988). The proposed workload components go beyond mental demand and have been applied in learning but in various contexts (Hart 2006). Another conceptualisation in multiple AR studies distinguishes between mental effort and load (Buchner, Buntins, and Kerres 2021). This distinction describes mental effort as the load that learners actively invest related to GCL, and mental load as passively elicited by the task affordances related to ICL (Klepsch and Seufert 2021). These different conceptualisations show the variety of research on cognitive load in AR-based learning. In the current systematic review, all conceptualisations will be included and the distinction will be used to structure the collected papers.

1.3 | Multimedia Design Guidelines

Mayer's (2021) Cognitive Theory of Multimedia Learning (CTML), like CLT, assumes that humans' information processing capacity is limited. The theory provides 15 guidelines for designing effective multimedia materials, reducing ECL, thereby preserving learning capacity, managing ICL or enhancing GCL (Mayer 2024). Multimedia learning describes learning from text-picture combinations, sometimes including other modes, such as haptics, smells, or tastes (Niegemann and Heidig 2012). CTML assumes that learners process information through two separate channels (visual and auditory information). Combining spoken words and displayed images allows simultaneous information processing, fostering learning without overloading the learner (Mayer 2021; Wickens 1981).

AR offers value through the combination of information sources, images, sound and physical integration. However, employing AR does not inherently yield enduring benefits, which necessitates meaningful design and instructions (Buchner, Buntins, and Kerres 2021, 2022). While many AR applications have been released, the ability to derive specific guidelines for AR design is limited. Most AR applications are still in the prototype stages or focus on technological implementation, requiring a more human-centred perspective (Bottani et al. 2021). CTML has been identified as a relevant theory for the instructional design of AR-based learning environments (Buchner, Buntins, and Kerres 2021; da Silva et al. 2019; Garzón et al. 2020; Sommerauer and Müller 2014). The multimedia, spatial and temporal contiguity, signalling, segmenting and modality principles are examples of principles that have been applied in AR-based learning setups (Lin and Tsai 2021; Sommerauer and Müller 2014). However, in a systematic literature review of articles from 2020 and before, Çeken and Taşkın (2022) found no papers that specifically tested the effectiveness of multimedia principles in AR. A few studies apply relevant principles when comparing AR and non-AR or might be transferable to AR. The following section presents studies in the context of virtual reality (VR) or AR that implement CTML principles and report their effects.

Dealing with integrating different information sources, the spatial and temporal contiguity effect advocates for the spatial and temporal alignment of related information to optimise cognitive resources and foster a more seamless learning experience (Mayer 2021). In studies comparing AR and non-AR, Thees, Altmeyer and colleagues tested the influence of the spatial contiguity principle (i.e., corresponding information placed far apart causing ECL) by comparing integrated

AR-based and separate screens to deliver physics knowledge in a hands-on experiment in multiple studies (Altmeyer et al. 2020; Thees et al. 2020, 2022). However, the results in the individual studies differed, showing both positive and negative effects of AR on ECL and knowledge outcomes. The authors argue that the choice of AR device (i.e., tablet or HMD) may have made an impact, with a mismatch between technological affordances and targeted cognitive processes leading to issues in coherence formation processes (Thees et al. 2022). This underscores the cognitive challenge posed when information is presented disparately across modalities, emphasising the importance of cohesive presentation for enhanced comprehension based on task goals.

Describing the potential usage of multimedia design in AR-based environments, Krüger and Bodemer (2022) evaluated the instructional design of combined physical and virtual information in environments simulating AR learning situations. The studies focusing on spatial contiguity and coherence principles found no confirmation of those principles in these designs (Krüger and Bodemer 2022). While the testing was not done with real AR, the results may partially apply similarly to real AR environments. Other results that may be partially transferable are established in research on VR. Based on the modality effect, it is assumed that an audio-visual presentation, compared to a visual-only presentation, can lead to higher learning outcomes (Mayer 2021), which was tested in VR by Albus and Seufert (2023). The results show a reverse modality effect, with higher learning outcomes and GCL when only visual material instead of combined audio-visual material was used, and no difference in ECL. The signalling effect, which describes the usage of cues, highlights, or annotations to guide learning processes through more organised material (Mayer 2021), was tested for HMD-based VR, where learners' recall and GCL were improved through annotations, but ECL was not different (Albus, Vogt, and Seufert 2021). In a 360° desktop VR implementation, signalling increased recall and comprehension and decreased ECL but did not influence transfer performance and GCL (Albus and Seufert 2022).

The results from the described studies, which only partially support multimedia principles in AR and similar settings, suggest starting from AR-specific affordances when effectively applying already existing design guidelines for task execution and cognitive processing. Adaptation of these or the creation of new guidelines that are specifically applicable in AR-based learning environments may be necessary. In the current paper, the ultimate goal is to develop specific recommendations for designing AR learning and training based on empirical evidence. Therefore, the authors perform a systematic literature review to evaluate and aggregate relevant study results. To achieve this, studies that specifically test design decisions within AR will be considered. For this purpose, we will focus on value-added and learner-treatment interaction study designs, excluding media comparison studies from the review.

1.4 | Media Comparison Versus Value-Added and Learner-Treatment Interaction Studies

Research criticises using media comparison studies for design statements (Buchner, Buntins, and Kerres 2022; Hartmann and Bannert 2022; Howard and Davis 2023; Lin and Yu 2023). Media comparison studies investigate the global effects of AR compared to other learning applications, for example, textbooks, virtual learning and web-based learning, comparing their effectiveness but allowing no specific conclusions for AR design. The methodology has persisted for several reasons, such as a straightforward application of research design, the high likelihood of significant effects, and explicit user feedback for novel technologies. One reason for positive effects is that individuals initially exhibit heightened interest or engagement simply due to the novelty of the technology itself, potentially leading to motivational effects when using AR (Hartmann and Bannert 2022). However, this can confound the assessment, requiring researchers to carefully distinguish between genuine sustained impact and short-term novelty-driven responses. Moreover, researchers argue it is difficult to establish comparable conditions for experimental and control groups in media comparisons due to the non-comparability of media, for example, textbooks and lectures varying both in presentation style and human contact (Hartmann and Bannert 2022; Surry and Ensminger 2001). Media comparison studies focus on global effects while excluding specific feature-based or human-system interaction effects.

Starting a general debate concerning the effect of technology on learning, Clark (1983) argued that a medium merely delivers instructional methods and does not impact the learning process itself. Kozma (1994), on the other hand, argued that media attributes like symbol systems and processing capabilities can be consequential for cognitive processes. He states that 'a particular medium can be described in terms of its capability to present certain representations and perform certain operations' (p. 11). In AR, for example, a multitude of virtual representations can be implemented, including 2D and 3D, text and pictures, static and dynamic and visual and auditory representations. Its unique nature involves the possibility to combine these virtual representations with representations in the physical world. Furthermore, on an operational level, real-time interaction with physical and virtual elements is possible and almost unlimited in the case of virtual representations. Distinguishing specific media attributes leads to an alternative to media comparison studies, namely value-added studies, where controlled variations of one learning medium are compared (Mayer 2019). Sometimes called intra-medium studies, an attribute or characteristic of the medium is manipulated as an independent variable (Surry and Ensminger 2001). Howard and Davis (2023) advocate this kind of research design to investigate the impact of specific AR attributes and be able to formulate practical guidelines for reasonable use and design. Another alternative to media comparisons is learner-treatment interaction studies, which examine how different learner characteristics interact with specific design decisions (Surry and Ensminger 2001). AR-based research has been mostly technology-centred and requires a more human-centred perspective (Bottani et al. 2021). The focus on cognitive load further includes human cognitive capacities. In the current systematic review, value-added studies and learner-treatment interaction studies will be included so that causal statements about the effectiveness of specific AR attributes and instructional features concerning cognitive load can be made.

So far, effects and conclusions about the design or use of AR have been based on systematic literature reviews and meta-analyses that mostly build on media comparison studies (Buchner and Kerres 2023). In contrast, empirical studies investigating the effects of specific design attributes on cognitive load and learning have not been systematically summarised (see Appendix 1). A systematisation that can guide instructional designers in the development of AR-based learning and instruction building on human-centred factors, like cognitive processing and demand, is still missing (Bottani et al. 2021, 2021; Buchner, Buntins, and Kerres 2021, 2022; Hartmann and Bannert 2022).

1.5 | Related Literature Reviews

Systematic literature reviews on AR attributes and instructions affecting cognitive load and learning are sparse, as shown in a systematic mapping review (Buchner, Buntins, and Kerres 2021). The authors questioned the approach of media comparison studies, which were predominantly used in their analysed studies. Previous systematic literature reviews and meta-analyses that are concerned with learning through AR either did not focus on cognitive load or focus on methodological concerns. A meta-analysis by Lin and Yu (2023) examined the effects of AR in interactive learning environments compared to traditional learning tools. The authors included 70 studies between 2012 and 2022 and analysed the impact of learning with AR on different learning outcomes compared to conventional learning. Bautista, Maradei, and Pedraza (2023) systematically discerned a design grounded in CLT through a comprehensive systematic literature review encompassing VR, MR and AR. Looking into the part concerned with AR, the authors delineated attributes, categorising them into three topics: (1) spatial proximity, (2) visual attributes and (3) content segmentation. However, in the literature overview (Bautista, Maradei, and Pedraza 2023), effects of media comparison studies were mixed with value-added studies and covered the continuum of AR, MR and VR, which makes statements on the usefulness of AR more challenging. Buchner, Buntins, and Kerres (2022) conducted a precise, systematic examination of the nexus between AR learning and cognitive load. The authors undertook a comprehensive literature analysis up to 2019, discerning trends and patterns in the relationship between AR implementation and cognitive load dynamics. Their findings suggested that AR evokes a lower cognitive burden than other applications, fostering learning outcomes. Nevertheless, the observed trends were contingent upon outcomes from media comparison studies. Appendix 1 summarises the literature on AR effects, cognitive load and learning. Since research is dominated by work providing global effects of AR on learning created by media comparison studies, it needs to be discussed how AR-specific designs affect cognitive load and learning.

1.6 | The Current Systematic Review: Goal and Research Questions

Based on the review of the theoretical background and with a different focus than previous reviews, this work systematically reviews research articles that address AR-specific instructional design and cognitive load for learning and training purposes. The central research question is:

> Which evidence-based instructional design recommendations to optimise cognitive load in augmented reality-based learning and training can be derived from the literature?

To address the central research question, this article analyses research articles related to three sub-research questions:

RQ1. How do different AR displays affect cognitive load during learning and training?

RQ2. How do different instructional designs in AR affect cognitive load during learning and training?

RQ3. What effects do the learner characteristics have in AR on learning and cognitive load?

The focus on learning and training will be applied to all three sub-research questions, based on the central research question. As described in Section 1.1, the focus on only AR will allow for technology-specific design recommendations for similar symbol systems and information processing operations (Section 1.4). The review will consider different AR devices and displays as defined by Rauschnabel et al. (2022). RQ1 focuses on study designs that compare different AR display types. As described in Section 1.4, comparisons of AR with other types of media are excluded from this review, which focuses on providing recommendations for designing AR-based learning and training experiences. Comparisons with other media do not provide enough AR-specific insights for the pursued design recommendations. However, AR representations can be implemented in different AR displays. These displays may present information in slightly different ways (e.g., OST vs. VST), leading to potential differences in cognitive load and performance, while keeping the essential AR experience which uniquely combines physical and virtual elements. RQ1 will consider this. The focus on cognitive load outcomes, as described in Section 1.2, allows a human-centred perspective, which is currently still missing in many research approaches on AR (Bottani et al. 2021; Buchner, Buntins, and Kerres 2021, 2022; Hartmann and Bannert 2022; Section 1.4). It will incorporate different types of workload and cognitive load conceptualisations, focusing on a distinction of desirable and non-desirable cognitive load as in CLT (Section 1.2). An examination of value-added and learner-treatment interaction studies will allow for creating specific design recommendations and a more human-centred approach, as described in Section 1.4. RQ 2 will be answered through studies with value-added designs, and RQ 3 will be answered through studies with learner-treatment interaction designs.

2 | Methodology

This article presents a systematic literature analysis on AR learning and cognitive load. The authors followed the PRISMA

TABLE 1 | Search string applied.

Search string				
Title/topic	**And**	**Title-abs-key**	**And**	**Title-abs-key**
('augmented reality' OR 'extended reality' OR 'mixed reality')		('cognitive load' OR 'mental load' OR 'workload' OR 'task load' OR 'mental demand' OR 'dual task' OR 'overload')		('learning' OR 'education' OR 'training' OR 'instruction')

guidelines to pursue a methodologically rigorous and comprehensive approach (Page et al. 2021).

2.1 | Search Terms

The review commenced by delineating precise search terms. The literature search was conducted using four databases, Scopus, Web of Science, PubMed and ERIC, to ensure profound access to the topic. Three terms were searched in combination, where '*augmented reality*' was supplemented with '*extended reality*' and '*mixed reality*' as the terminological distinction is not used clearly (Rauschnabel et al. 2022). The second term consists of '*cognitive load*' and synonyms like '*mental load*' *OR* '*task load*', '*workload*', '*mental demand*', '*dual task*', *OR* '*overload*'. Since this systematic literature review focuses on the application context of learning and training, '*learning*', '*education*', '*training*', OR '*instruction*' were added as the application context (see Table 1).

2.2 | Inclusion and Exclusion Criteria

A set of inclusion and exclusion criteria was established to delimit the scope of the review. Since the authors sought primary and empirical peer-reviewed research, only journal articles and conference proceedings were included, excluding reviews, meta-analyses, theses and book chapters. Additionally, only primary studies with a comparative study design contrasting at least two different AR settings (value-added studies) were considered. Exclusive consideration was accorded to peer-reviewed articles to maintain a high-quality standard. Only texts written in English were considered. To address the scope of the research questions, studies with the following criteria were excluded:

Studies that

- Do not evaluate cognitive load or similar concepts like workload, mental load, mental effort, task load, or mental demand.
- Do not include AR as defined in Section 1.1, specifically excluding VR and 360° video.
- Execute media comparisons, comparing AR to another learning application (e.g., paper-pencil, VR, 'traditional learning'; distinction based on Mayer 2019).
- Do not aim to foster learning or training performance, excluding other domains, such as automated driving.

Furthermore, since this article examines the use of AR in the tertiary education sector, studies on education for children (primary or secondary sector) were excluded.

2.3 | Search Strategy

The literature search was conducted between May and June 2024. Applying the search terms to the given databases with no publication year restrictions revealed 1188 references. Three hundred and seventy-eight duplicates were eliminated. The remaining 810 sources were screened based on the inclusion and exclusion criteria independently by both authors, who reviewed all titles and abstracts. In instances of disagreement, the abstract was reviewed collaboratively until a consensus was reached. This process resulted in 74 articles proceeding to full-text screening. These articles were divided between the authors for a final decision. If an author needed clarification about including an article, it was reviewed jointly to reach a collective decision. Ultimately, 27 articles were included in the review containing 29 distinct studies since some publications include more than one study, leading to a higher number of studies than articles. Figure 1 demonstrates the literature search and filtration processes for each step.

2.4 | Coding Scheme

Both authors collaborated on the coding process for the 29 studies. Each study included in the review was jointly assessed and coded. For the coding of the AR technologies, the authors use the classification by Rauschnabel et al. (2022), distinguishing AR devices into stationary, wearable, mobile, on-body and in-body and AR display types into optical see-through (OST), video see-through (VST), projection-based AR and AR mirrors. For the research questions, the studies were split into: (1) studies comparing different AR display types for RQ1, and (2) studies comparing different attributes within a single display type for RQ2, coding the instructional design comparison. The inclusion of the assessment of learner characteristics was coded for RQ3. Basic study descriptions were gathered from all papers, including the application context, sample size, study population and study design. Furthermore, the methods used to measure cognitive load and learning performance were collected from each study. Finally, we summarised the findings of each study regarding cognitive load/workload and learning performance. The results of the full coding can be found in the table provided under this Link. https://osf.io/u9rt7/?view_only=d6c4d503d593433386a84fe20c8db12c.

Based on this full coding, particularly relevant data for answering the research questions is summarised in the findings in Section 3. The type of AR display and the type of cognitive load measurement are provided as context variables. To address RQ1, the effects of different AR display types on cognitive load and performance reported in studies from the first cluster are

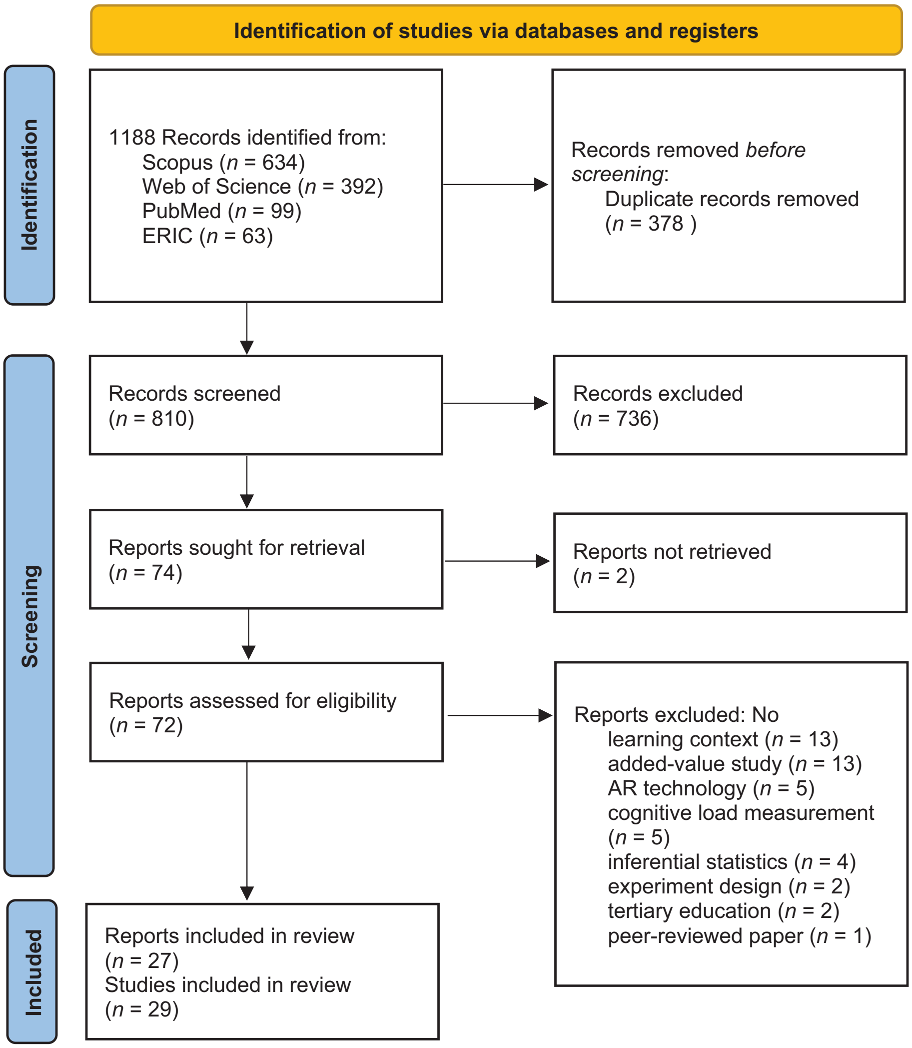

FIGURE 1 | Overview of the conducted review process following the Prisma guidelines.

summarised. For RQ2, the effects of the instructional design comparisons assessed in the second cluster of studies are categorised and summarised. To address RQ3 the studies were screened for effects of learner characteristics on cognitive load and learning. In order to develop design recommendations as postulated by the overarching goal of the systematic review, the observed effects are summarised into recommendations, which are clustered into categories within different design dimensions.

3 | Findings

In this section, we will describe the results of the systematic literature review, addressing the sub-research questions RQ1, RQ2 and RQ3. The translation of the findings into specific design recommendations, responding to the central research question, happens in Section 4. As described in Section 2.4, the 29 studies are distinguished into two clusters:

1. Five studies that compare different AR display types with each other (RQ1).
2. Twenty-four studies that compare specific design decisions within one AR display type (RQ2 and RQ3).

To address RQ1, we analyse and summarise the five studies comparing the effects of different AR display types on cognitive load and learning performance. For RQ2, the second cluster with 24 studies comparing specific design attributes within one display type is analysed and summarised concerning the effect of specific instructional design. Concerning RQ3, we found four studies that examine the influence of learner characteristics, which are all in the second cluster.

When an article presents two independent study designs, we classify and report them as two separate studies, which is why there are 29 studies from 27 articles. However, when studies employ a two-factor design, we treat them as a single study. When there is no observed interaction effect between the two factors, we separately report the two instructional designs and detail their respective effects as distinct entries in the tables.

3.1 | General Study Information

To describe the included studies and provide some context for the research questions, we will first give an overview of the AR devices, display types and cognitive load measurement methods utilised in the sample.

3.1.1 | AR Device and Display

Different types of AR devices and displays as categorised by Rauschnabel et al. (2022) were used in the studies (see Appendix 2). 21 of the studies used wearable AR devices, with 18 studies using OST HMD and three studies using VST HMDs. Eight studies used mobile AR on tablets (six studies) or smartphones (two studies). Four studies employed stationary AR, with three utilising projection-based applications and one using an AR mirror. Notably, all instances of stationary AR were compared against other AR displays, such as OST and VST. No study in the sample examined the effects of instructional design decisions within projection-based AR. Furthermore, we found no studies that used on-body or in-body AR.

3.1.2 | Measuring Cognitive Load and Workload

The studies incorporate various cognitive load measurements, which can be distinguished into mono-method (using one assessment method) and multi-method assessments (combining at least two assessment methods). Appendix 3 presents an overview of assessment methods and references, respectively. The majority of 23 studies used one method, namely questionnaires. Out of these, 18 studies used the NASA-TLX. This questionnaire assesses workload in total and on the subdimensions mental demand, physical demand, temporal demand, performance, effort and frustration (Hart and Staveland 1988). Another six studies used questionnaires that distinguish between mental load and mental effort by either Hwang, Yang, and Wang (2013), Krell (2015) or Paas, van Merriënboer, and Adam (1994). Only one study relied on the cognitive load theory conceptualisation in the subdimensions ICL, GCL and ECL using the questionnaire by Klepsch, Schmitz, and Seufert (2017).

In contrast to mono-method assessment, studies using a multi-method approach supported the cognitive load assessment with a second questionnaire or physiological measurements. Four studies used a second questionnaire, combining the ICL, ECL and GCL assessment with the workload assessment (NASA-TLX). Another three studies used physiological measurements such as heart rate variability, skin conductance (EDA) and eye-tracking in addition to a questionnaire.

The studies identified in our systematic literature review exhibit considerable diversity in the conceptualisation of cognitive load or workload, employing nine different measurement methods. In the subsequent presentation of results, we adhere to the conceptualisations as defined by the respective authors, distinguishing workload, which refers to the measurement NASA-TLX, and cognitive load, which covers the assessment of mental load and mental effort, as well as the conceptualisation of ICL, ECL and GCL.

3.2 | Effects of the Augmented Reality Display

To address RQ1, 'How do different augmented reality displays affect cognitive load during learning and training?', we summarise the findings of the five studies comparing two display types. Table 2 presents an overview of the studies and their findings. Projection-based AR is predominantly evaluated against see-through AR. This is done by juxtaposing VST, projection-based AR and OST methods (Baumeister et al. 2017) or by comparing projection-based AR against OST approaches (Boyce et al. 2022; Nowak et al. 2020). The fifth study compares OST with AR mirror (Karg et al. 2023).

In the two studies by Baumeister et al. (2017), reaction tasks were carried out with different versions of a control panel displaying high or low amounts of information. The findings show that projection-based AR in comparison to VST and OST resulted in lower cognitive load, as determined by the single-item assessment of Paas, van Merriënboer, and Adam (1994) and quicker secondary task response time, and also in better performance measured as faster primary task response times. The authors reason that the OST and VST HMD devices' limited field of view played a substantial role in the results. With a high amount of simultaneously presented information, OST outperformed VST when field of view was accounted for (Baumeister et al. 2017). The study by Boyce et al. (2022) reports similar findings regarding decision-making in military tactics. Projection-based AR in comparison to OST led to lower workload, as measured by the NASA-TLX, and shorter decision response times, although decision accuracy did not differ between the two AR displays.

Karg et al. (2023) and Nowak et al. (2020) investigated the usage of OST and, respectively, projection-based AR and AR mirrors in procedural tasks that demand motoric task execution, such as an assembly and wiring task. Nowak et al. (2020) did not find significant differences between the AR displays regarding workload or performance. The findings of Karg et al. (2023) indicated a significant benefit of using OST, which led to shorter assembly times and fewer errors. Additionally, participants using the OST reported less workload in terms of mental demand, effort, frustration and higher performance on the NASA-TLX subscales. In contrast, the AR mirror setup led to participants deviating more from the ideal position.

In total, the limited findings concerning RQ1 preliminarily suggest advantages in performance for projection-based AR and OST over VST but also increased performance when using OST compared to AR mirrors. The type of task and especially the consideration of the field of view in the task design seem to play an important role in these relationships, so that no simple response to RQ1 is possible based on this data. The focus on specific instructional design decisions and learner characteristics in the following sections can provide more specific insights.

3.3 | Effects of the Instructional Design Applied in Augmented Reality

To address RQ2, 'How do different instructional designs in augmented reality affect cognitive load during learning and training?', we summarise the effects of the 24 studies in the second cluster, which assess the impact of a specific design attribute within one display type. Table 3 provides an overview of the studies and effects of instructional design attributes on cognitive load/workload and learning performance.

TABLE 2 | Summary of cognitive load/workload, learning, and performance effects across studies ($N = 5$) comparing different AR display types.

AR display	AR display technology	Effects on cognitive load/workload	Effects on learning/performance	Reference
Video-see-through vs. Projection-based vs. Optical-see-through	VST: Gear VR SAR: Projector OST: HoloLens	Cognitive load: Low amount of information: Sign. lower mental effort in SAR than other displays ($p < 0.01$). High amount of information: Sign. lower mental effort in SAR than other displays ($p < 0.01$). OST was lower than VST ($p < 0.01$). Secondary Task: Low amount of information: Sign. faster response time in SAR than other displays ($p < 0.05$). High amount of information: Sign. faster response time in SAR than other displays ($p < 0.01$).	Task response time: Low amount of information: Sign. faster response times in SAR than other displays ($p < 0.01$). High amount of information: Sign. faster response times in SAR than other displays ($p < 0.01$), followed by OST ($p < 0.01$), followed by VST ($p < 0.01$).	Baumeister et al. (2017)
Video-see-through vs. Projection-based vs. Optical-see-through	VST: Gear VR SAR: Projector OST: HoloLens	Cognitive load: Low amount of information: Sign. lower mental effort in SAR than other displays ($p < 0.01$). High amount of information: Sign. lower mental effort in SAR than other displays ($p < 0.05$). OST sign. lower than VST ($p < 0.01$). Secondary Task: Low amount of information: Sign. faster response times in SAR than other displays ($p < 0.05$). VST sign. faster than OST ($p < 0.05$). High amount of information: Sign. faster response times in SAR than other displays ($p < 0.05$).	Task response time: Low amount of information: Sign. faster times in SAR than other displays ($p < 0.01$). OST faster than VST ($p < 0.01$). High amount of information: Sign. faster in SAR than other displays ($p < 0.01$). OST faster than VST ($p < 0.01$).	Baumeister et al. (2017)
Optical-see-through vs. Projection-based	OST: HoloLens 1 SAR: Sand table (a projection-based display onto a military sand table)	Workload: Sign. lower in SAR than OST ($p < 0.01$).	Decision Accuracy: No sign. differences. Response Time: Sign. lower in SAR than OST ($p < 0.01$).	Boyce et al. (2022)
Optical-see-through vs. Projection-based	OST: HoloLens SAR: In-situ projections	No sign. differences in total workload or sub-components.	TCT: No sign. differences. Errors: No sign. differences.	Nowak et al. (2020)
Optical-see-through vs. AR Mirror	OST: HoloLens AR Mirror	Workload: Sign. lower mental demand ($p < 0.01$), perceived effort ($p < 0.01$), and frustration ($p = 0.01$) but higher performance ($p \leq 0.045$) in OST compared to AR mirror.	TCT: Sign. longer assembly time using AR mirror ($p < 0.01$). Errors: No bricks were misplaced with OST, while a scattering around the ideal position occurred with the AR mirror setup.	Karg et al. (2023)

Abbreviations: errors, indicate the number of errors; OST, optical-see-through; SAR, projection-based; TCT, refers to task completion time; VST, video-see-through.

Eleven studies assess instructional design attributes for OST HMDs, while 13 studies compare instructional designs within VST displays, such as VST HMDs, tablets, or smartphones. The design attributes in the studies were clustered into different design categories: spatial integration, spatial visualisation, physical-virtual integration, content modality, content complexity, generative activity, interaction modalities, physical interaction, embodied assistance and adaptive guidance.

Concerning RQ2 it can be concluded that many different design attributes in AR-based learning environments can have an influence on cognitive load, workload and learning and training performance. The specific results of the different design attributes can be seen in Table 3. A more specific and multi-faceted response to RQ2 will be provided in Section 4.1, assessing and discussing the large variety of design attributes examined in the studies, clustering the design categories and turning the findings into recommendations.

3.4 | Effects of Learner Characteristics

To address RQ3, 'What effects do the learner characteristics have in augmented reality on learning and cognitive load?' and enrich the design recommendations based on the findings from Section 3.3, we examined the interaction effects of learner characteristics and design decisions in four of the 29 studies (see Appendix 4). Two of those studies included spatial abilities, measured through a mental rotation test, in interaction with spatial visualisation as a design dimension. Bogomolova et al. (2023) used an OST display and found no interaction effect of spatial abilities and spatial visualisation (stereoscopic 3D vs. monoscopic 3D) on cognitive load or learning-related variables. Krüger, Palzer, and Bodemer (2022), on the other hand, used a VST display and found a significant moderation effect concerning learning outcomes that describes that for learners with higher spatial abilities, 3D compared to 2D visualisations improve their knowledge outcome. They also found no interaction effect concerning cognitive load. Another study that examined a learner-treatment interaction was executed by Kim, Laine, and Åhlund (2021) with an OST display. The authors found that confidence played a role when working with a virtual instructor, finding that only underconfident but not overconfident students learned better with a virtual instructor than without one. Furthermore, only for overconfident students who learned with a virtual instructor a negative correlation between workload and learning performance could be found. A fourth study separately examined the effects for learners who were experienced in the targeted task execution and inexperienced learners (Lange-Nawka, Wünsche, and Thompson 2023). Different results were found in the groups concerning cognitive load, amount of eye movement, focus switches, focus time, focus depth and different types of errors. However, no clear pattern of differences can be identified as the results differed between different tasks, which the authors partly attributed to unknown task characteristics (Lange-Nawka, Wünsche, and Thompson 2023). In total, the limited number of findings concerning RQ3 preliminarily suggest that learner characteristics such as spatial abilities, confidence and prior experience can impact cognitive load and learning performance in interaction with design decisions in AR contexts.

Appendix 4 provides a more detailed picture of the results per study, and in Section 4.1, the findings are discussed and integrated into the design recommendations.

4 | Discussion

This review addresses the central research question: 'Which evidence-based instructional design recommendations to optimise cognitive load in augmented reality-based learning and training can be derived from the literature?'. To approach this research question, we systematically evaluate the existing literature concerning the effects of augmented reality (AR) attributes on cognitive load in educational and training contexts. The next step clusters these effects and develops specific, evidence-based recommendations for designing AR applications in learning (Section 4.1). The central value of this paper is the aggregation of existing evidence from empirical cognitive load studies, resulting in recommendations for AR instructional design. The findings provide practical guidance for instructional designers and practitioners involved in AR-based learning and training in tertiary education. The review summarises findings from added-value and learner-treatment interaction studies, focusing on the specific impact of AR design features and learner characteristics rather than media comparisons, which are criticised for their inability to examine the effect of specific AR features. The systematic screening approach to identify and synthesise empirical studies exploring the influence of AR design attributes on cognitive load and learning outcomes applied here leads to systematic insights into specific AR design decisions. Despite the promising potential of AR-based learning to enhance learning processes and achievements, the specific design needs to be considered. While aggregating evidence from empirical studies, this review also highlights significant gaps in the current literature, including the limited number of studies empirically examining the effects of various AR attributes on cognitive load using added-value methodologies, thus identifying critical areas for further research and development (Section 4.2).

4.1 | Integration of Results

Based on the findings from the systematic literature review described in Section 3, the results were aggregated, and design recommendations were established. To structure the results, the design attributes that were discovered in the studies were clustered into different categories. Based on the data from RQ1, the category of display selection was developed. Based on the data from RQ2 and RQ3, two categories of design decisions were identified: AR-specific design decisions and decisions that can be independent of AR. For AR-specific design decisions, the results were clustered into spatiality-related, interaction-related and contextuality-related results based on the ARcis framework (Krüger, Buchholz, and Bodemer 2019; Krüger 2023). Two further categories, that are important for instructional design in general but not necessarily connected to AR-specific characteristics of the learning situation, were clustered into content-related and guidance-related results. The aggregated results and recommendations can be found in Table 4.

TABLE 3 | Overview of cognitive load/workload, learning, and performance effects in studies ($N = 24$) comparing instructional designs in AR, clustered by design dimension and attribute.

AR display	Design categories	Design attributes	Effects on cognitive load	Effects on learning/performance	Reference
OST	Spatial Integration	Animation-based visualisation Video-based visualisation	Workload, EDA and HR: No sign. differences.	TCT: Sign. shorter for 3D animation than video ($p < 0.01$). Errors: Sign. less for 3D animation than video ($p < 0.01$).	Ariansyah et al. (2022)
	Spatial Integration	Superimposed guidepost Guideposts and 3D model	Workload: No sign. differences during exploration task; greater effort for guideposts only compared to combined layout during wayfinding task ($p = 0.02$).	TCT (exploration and wayfinding task), sketching map score, self-evaluation of learning: No sign. differences.	Zhang et al. (2021)
	Spatial Integration	Virtual information fixed to window Virtual information fixed to target	Workload: No sign. differences.	TCT, recognition reaction time and errors: No sign. differences.	Yu et al. (2022)
	Spatial Visualisation	Stereoscopic 3D Monoscopic 3D	Workload: No sign. differences.	Anatomical knowledge: No sig. differences (written or plastinated specimen test).	Bogomolova et al. (2023)
	Spatial Visualisation	2D illustrations 3D models	Workload: No sign. differences.	TCT, errors: No sign. differences.	Simmen et al. (2023)
	Content Modality	Visual-led learning material Auditory-led learning material	Workload: No sign. differences.	TCT: Sign. shorter for auditory than visual for driver-controller ($p = 0.01$) and swapping-train-end ($p = 0.03$) tasks. Errors: Sign. less for visual than auditory in icon-recognition ($p < 0.01$); sign. less for auditory than visual in driver-controller ($p < 0.01$) and swapping-train-end ($p = 0.02$) tasks. Recognition reaction: Sign. quicker reaction ($p < 0.01$) and less errors ($p < 0.01$) for auditory than visual for driver-controller task.	Yu et al. (2022)

(Continues)

TABLE 3 | (Continued)

AR display	Design categories	Design attributes	Effects on cognitive load	Effects on learning/ performance	Reference
	Content Modality	Text only Audio only Text and audio at the same time Audio with subtitles—synchronously Audio with subtitle—text 2 s. faster Audio with subtitle—text 2 s. slower	Workload and HRV: No sign. differences.	Knowledge: No sign. differences.	Zhao et al. (2023)
	Generative Activity	With assessment Without assessment	Workload: No sign. differences.	TCT: No sig. differences. Errors: Sign. less sequence mistakes when assessment is included than when not ($p = 0.04$, $r = 0.26$). Immediate recall completion time: Sign. faster when assessment than when no assessment ($p = 0.02$, $r = 0.29$).	Werrlich, Nguyen, and Notni (2018)
	Interaction Modalities	Gesture command checklist Voice command checklist	Workload: Sign. higher for gesture than voice ($p < 0.01$). Sub-components: Sign. higher mental ($p < 0.01$), physical ($p < 0.01$), temporal demand ($p < 0.01$), effort ($p < 0.01$), frustration ($p < 0.01$) and lower perceived performance ($p < 0.01$) with gesture.	Situational Awareness: Sign. lower with gesture than voice ($p < 0.01$).	Li et al. (2022)
	Interaction Modalities	Voice-command Hand-gesture	Workload, EDA and HR: No sign. differences.	TCT, errors: No sign. differences.	Ariansyah et al. (2022)
	Physical Interaction	Tablet-based control Control device fixed to AR-HMD Control device fixed to arm	Workload: Sign. higher with device at HMD than at arm ($p < 0.04$) or tablet-based control ($p < 0.04$). Sub-components: Sign. higher mental demand with device at HMD than tablet-based control ($p < 0.05$).	TCT: No sign. differences.	Van den Bergh and Heller (2020)

(Continues)

TABLE 3 | (Continued)

AR display	Design categories	Design attributes	Effects on cognitive load	Effects on learning/performance	Reference
	Embodied Assistance	Voice-only assistant Embodied assistant No assistant	Workload: Sign. higher with voice than embodied ($p = 0.03$) or no ($p = 0.02$) assistant. Sub-components: Sign. higher frustration with embodied ($p < 0.01$) or voice ($p = 0.02$) than no assistant.	Task performance: Sign. better with embodied ($p < 0.01$) or voice ($p < 0.01$) than no assistant.	de Melo et al. (2020)
	Embodied Assistance	With virtual assistance Without virtual assistance	Workload: Sign. lower with a virtual instructor than without ($p < 0.01$).	Learning performance: No sign. differences.	Kim et al. (2023)
VST	Spatial Visualisation	3D virtual model 2D virtual model	CL: Sign. higher GCL for 3D than 2D ($p = 0.045$, $d = 0.28$), no sign. difference for ECL.	Knowledge: Sign. higher for 3D than 2D ($p = 0.01$, $d = 0.38$) (spatial relations); no sign. difference for other types.	Krüger, Palzer, and Bodemer (2022)
	Physical–Virtual Integration	Physical objects to activate AR content Marker cards to activate AR content	ME/ML: Sign. lower mental load for objects than markers ($p = 0.04$); no sign. difference in mental effort.	Knowledge: Sign. higher for objects than markers ($p = 0.01$).	Huang, Huang, and Cheng (2022)
	Content Complexity	One task Two parallel tasks Four parallel tasks	Workload sub-components: Sign. higher temporal demand ($p = 0.02$, $r = 0.70$) and effort ($p = 0.01$, $r = 0.74$) with four compared to one task.	TCT: Sign. shorter when four than when one task ($p < 0.01$). Errors: Sign. more when four than when one ($p < 0.01$, $r = 0.84$) or two ($p < 0.01$, $r = 1.03$) tasks.	Illing et al. (2021)
	Content Complexity	Adaptive AR Non-adaptive AR	Workload: No sign. differences. EEG: Sign. lower mental workload when adaptive than when non-adaptive ($p < 0.01$).	Performance: Sign. increase from pre to post when nonadaptive ($p < 0.01$) but no sign. change when adaptive.	Maitz et al. (2023)
	Content Complexity	Instruction synchronous with the previous task Instructions asynchronous between tasks	Workload: Sign. higher when synchronous than when asynchronous ($p < 0.01$). Sub-component: Sign. lower perceived performance ($p = 0.04$), and higher effort ($p = 0.03$) and frustration ($p = 0.01$) when synchronous than when asynchronous.	TCT: No sign. differences in general; but sign. longer for writing task (not math or search task) when synchronous than when asynchronous ($p < 0.01$, $d = 0.81$).	Lenz et al. (2024)

(Continues)

TABLE 3 | (Continued)

AR display	Design categories	Design attributes	Effects on cognitive load	Effects on learning/performance	Reference
	Generative Activity	High mental interaction Low mental interaction	CL: No sign. effects on ICL, GCL, ECL Workload sub-components: No sign. effects.	Knowledge: No sign. main effect, but sign. interaction with physical interaction ($p < 0.01$, $\omega_p^2 = 0.06$): higher for high mental interaction when low physical interaction, but higher for low mental interaction when high physical interaction.	Krüger and Bodemer (2020)
	Generative Activity	With formative assessment Without formative assessment	ML/ME: Sign. greater mental effort with than without formative assessment ($p = 0.01$, $d = 0.83$), no sign. difference in mental load.	Knowledge: Sign. higher with than without formative assessment ($p = 0.03$, $d = 0.73$).	Chu et al. (2019)
	Physical Interaction	High physical interaction Low physical interaction	CL: No sign. effect on ICL, GCL, ECL Workload sub-components: No sign. effects.	Knowledge: No sign. main effect, but sign. interaction with mental interaction ($p < 0.01$, $\omega_p^2 = 0.06$): higher for high physical interaction when low mental interaction, but higher for low physical interaction when high mental interaction.	Krüger and Bodemer (2020)
	Embodied Assistance	Instructional annotation Instructional hand-gestures Instructional avatar Instructional volumetric playback	Workload sub-components: Sign. more mental load ($p < 0.01$), effort ($p < 0.01$), and frustration ($p < 0.01$) with avatar than volumetric and sign. more frustration with avatar than hand instruction ($p < 0.01$).	TCT: Sign. longer with avatar than volumetric ($p < 0.01$), hand ($p < 0.01$) or annotation instruction ($p < 0.01$).	Sasikumar et al. (2021)
	Embodied Assistance	Gaze direction cues Hand-gesture cues Combined gaze and hand-gesture cues	Workload sub-components: Sign. higher mental demand ($p = 0.02$) and frustration ($p = 0.03$) with gaze than with combined cues.	TCT: Sign. shorter with gestures than with gaze instructions ($p = 0.01$).	Sasikumar et al. (2021)
	Adaptive Guidance	Fading instructions Non-fading, constant instructions	Workload: No sign. differences.	Errors: Sign. smaller error reduction when non-faded than when faded ($p = 0.04$); more errors in non-faded than faded group. TCT or knowledge: No sign. differences.	Herbert et al. (2022)

(Continues)

TABLE 3 | (Continued)

AR display	Design categories	Design attributes	Effects on cognitive load	Effects on learning/performance	Reference
	Others	Piano instructions without a metronome Piano instructions with a metronome	CL, eye movement: No general sign. differences.	Errors: Sign. less start time errors ($p=0.04$) but more missed end counts ($p=0.02$ and $p=0.046$) in song #2 with metronome than without.	Lange-Nawka, Wünsche, and Thompson (2023)
	Others	With gamification Without gamification	ML/ME: No sign. differences.	Knowledge Test: No sign. differences.	Czok et al. (2023)
	Others	Team competitive strategy Individual competitive strategy	ML/ME: Sign. less mental effort in collaborative than individual setting ($p=0.03$), no sign. mental load difference.	Cultural literacy: Sign. higher in collaborative than individual settings ($p=0.02$), sub-component cultural identity sign. higher ($p=0.03$). Computational thinking: No sign. differences.	Zhan et al. (2024)

Note: p values reported for sign. differences and effect sizes added if reported in the paper.
Abbreviations: CL, ICL, GCL, ECL; EDA, skin conductance; errors, number of errors; HRV, heart rate variability; ML/ME, mental load/mental effort; OST, optical-see-through; TCT, task completion time; VST, video-see-through; Workload, measured with NASA-TLX.

The spatiality-related results include studies on spatial integration and visualisation. In general, the studies on *spatial integration* show that in-place visualisations can reduce the completion time in haptic tasks without overloading the learner and that mental load can be decreased when spatially integrated AR information is embedded into a holistic visualisation (Ariansyah et al. 2022; Yu et al. 2022; Zhang et al. 2021). As already mentioned in the theoretical background, AR has the unique potential of spatially integrating virtual and physical content (Altmeyer et al. 2020). The focus on this characteristic in multiple studies and the outcomes support this idea. Concerning *spatial visualisations*, the results are mixed. Showing a positive effect of using 3D objects for learning spatial relations only for students with high and average spatial abilities, Krüger, Palzer, and Bodemer (2022) describe the necessity of including learner characteristics in research on AR. Not all results support the advantage of a 3D representation, especially when the information itself can be easily represented in 2D, as in the study by Simmen et al. (2023). However, if a task requires spatial understanding, including most STEM-based content, it is thus useful to visualise the content in 3D, as it increases GCL, which is associated with schema building and long-term memory integration (Sweller, van Merriënboer, and Paas 2019), and improves spatial relation knowledge results. This is in accordance with a systematic mapping review on STEM learning in higher education that showed the prevalence of the usage of 3D models in AR-based education (Mystakidis, Christopoulos, and Pellas 2021). The type of 3D visualisation, monoscopic or stereoscopic, did not seem to play a role (Bogomolova et al. 2023).

The interaction-related results include studies on interaction modalities and physical interaction. Concerning *interaction modalities*, it was found that workload can be reduced when using voice commands during an otherwise physical task (Li et al. 2022), but not necessarily when the task does not need usage of both hands (Ariansyah et al. 2022). Different AR devices include different displays but also different possibilities for interaction. Gesture-based and voice-based interactions are possible in HMDs, but not always possible in handheld devices, which usually need at least one hand to hold the device. When designing a task in AR, the number of hands needed for interaction and the possibility of using voice command as an alternative should be considered. *Physical interaction* describes findings on how physical interaction can support learning if the mental task is not too demanding (Krüger and Bodemer 2020). This shows that it is important to provide meaningful physical interaction that supports and does not clash with mental demands. This is in accordance with Clark and Mayer's (2016) four quadrants of interaction, showing that mental interaction but not physical interaction is necessary for learning. Furthermore, when designing an interaction, the usability of the control mechanisms should be considered in a way that it does not increase ECL, for example through the placement of control devices (Van den Bergh and Heller 2020).

The contextuality-related results include only one study on physical-virtual integration (Huang, Huang, and Cheng 2022). This study shows that using physical objects as anchors of virtual elements in AR can lead to decreased mental load and increased knowledge. This is in accordance with the definition of contextuality in the ARcis framework, which describes the

TABLE 4 | Aggregation of empirical evidence-based study results.

Design dimension	Design categories	Evidence-based recommendations based on empirical study results	Studies
AR-specific: Spatiality-related	SpatialIntegration	1. Incorporating in-place visualisation can reduce the completion time in haptic tasks without overloading the learner. 2. To prevent mental overload, incorporate spatially integrated AR information into a comprehensive, holistic model.	Ariansyah et al. (2022) Zhang et al. (2021) Yu et al. (2022)
	Spatial Visualisation	3. Visualise content in 3D if a task requires spatial understanding (e.g., STEM), as it increases GCL and improves spatial relation knowledge results. 3* However, consider the spatial abilities of the person since they impact effectiveness.	Krüger, Palzer, and Bodemer (2022) Bogomolova et al. (2023) Simmen et al. (2023)
AR-specific: Interaction-related	Interaction Modality	4. Use voice control for procedural or motor tasks where hands are required to perform a task, as it reduces workload.	Ariansyah et al. (2022) Li et al. (2022)
	PhysicalInteraction	5. Provide meaningful physical interaction in the sense of good usability, as it activates the learner. 5* However, no or poorly implemented physical interaction increases the cognitive load in a negative sense.	Krüger & Bodemer (2020) Van den Bergh and Heller (2020)
AR-specific: Contextuality-related	Physical–Virtual Integration	6. Incorporate tangible content since it increases knowledge while reducing mental load.	Huang, Huang, and Cheng (2022)
General: Content-related	ContentModality	7. Provide auditory information for procedural or motor tasks since it reduces errors and time.	Yu et al. (2022) Zhao et al. (2023)
	Content Complexity	8. Balance the number of simultaneous tasks since too many tasks overload the learner and cause more errors. 9. Adapting the content presentation based on mental workload can decrease physiological measured mental workload.	Maitz et al. (2023) Lenz et al. (2024) Illing et al. (2021)
	GenerativeActivity	10. Incorporate generative activities since they lead to better knowledge test results and fewer errors while the learner is activated in terms of mental effort.	Chu et al. (2019) Werrlich, Nguyen, and Notni (2018) Krüger and Bodemer (2020)
General: Guidance-related	EmbodiedAssistance	11. Integrate virtual assistance that provides suggestions for improvement or guidance since it reduces workload, resulting in shorter task duration and better learning. 11* Effects are increased for underconfident learners. 12. Incorporate a human-like/embodied tutor in a volumetric rather than an avatar-like visualisation.	Kim et al. (2023) de Melo et al. (2020) Sasikumar et al. (2021)
	AdaptiveGuidance	13. Adaptively adjust learning content based on the user's learning progress, as this can lead to fewer application errors or increased knowledge. 13* Adaptive design is not necessarily linked to reducing cognitive load since learning occurs at an optimal workload level at which under- and overstimulation should be prevented.	Herbert et al. (2022)
Display-selection	OST or projection-based	14. If feasible and practical, consider incorporating projection-based AR or OST for tasks that require significant haptic interaction and tactile precision. 15. Avoid utilising AR mirrors for information displays and consider alternative AR methods for optimal results and reduced cognitive burden.	Baumeister et al. (2017) Boyce et al. (2022) Nowak et al. (2020)

Note: * markers the influence of learner attributes that interact with the reported effects.

connection of corresponding virtual and physical elements as one of the unique advantages of AR that should be leveraged for learning situations in which virtual information and physical elements are closely connected (Krüger 2023; Krüger, Buchholz, and Bodemer 2019).

Content-related results, which are not necessarily specific to AR, include studies on content modality, content complexity and generative activity. *Content modality* results describe the usefulness of providing auditory information, as errors and task completion time can be reduced, especially in procedural or motor tasks (Yu et al. 2022), but not in all types of tasks (Zhao et al. 2023). This is in accordance with the results on the interaction modality described above, showing the other side in which visual channels may already be focused on other tasks and thus not available to process visual information. Furthermore, looking at the research on VR described in Section 1.3, a reversed modality effect showed that unique characteristics of technology and tasks may come together to break established design principles (Albus and Seufert 2023). The topic of *content complexity* is another prevalent area in research on AR-based learning and cognitive load, showing that too many parallel tasks may overload learners, producing more errors (Illing et al. 2021) and that two independent task instructions lead to a higher workload when presented synchronously instead of after each other (Lenz et al. 2024). As AR allows for the presentation of much content in various formats, it is important to find a 'sweet spot' of how much information should be shown at what time to not overload but still challenge learners. The potential of adaptive environments automatically adapting task complexity to learners' workload is shown by Maitz et al. (2023), decreasing physiologically measured workload. However, learning improvements were higher with non-adaptive material, showing that the adaptivity needs to be well thought through. When looking at the results concerning *generative activity*, it is clear that introducing assessments as a form of deeper learning or elaboration strategy can increase knowledge and active load (s. Klepsch and Seufert 2021) components like mental effort while decreasing errors (Chu et al. 2019; Werrlich, Nguyen, and Notni 2018). This is in accordance with the generative learning principle (Mayer 2021). Another study that specifically examined the introduction of a mentally demanding task found that the mental interaction only led to increased knowledge outcomes when no physically demanding task was apparent (Krüger and Bodemer 2020). This shows that excessive demand could hinder learning, and the aim is to strive for an optimal demand.

Guidance-related results include studies on embodied assistance and adaptive guidance. The results on *embodied assistance* show that using embodied compared to voice-based or no assistance can decrease workload, especially for underconfident learners (Kim et al. 2023), and improve learning outcomes (de Melo et al. 2020). This effect also appeared for a volumetric rather than an avatar-based visualisation (Sasikumar et al. 2021). Furthermore, gesture-based visualisations of assistance have a positive effect on workload and task completion time (Sasikumar et al. 2021). *Adaptive guidance* includes the results of only one study by Herbert et al. (2022). Fading of instruction based on learning progression can lead to less application error or increased knowledge. However, adaptive design is not necessarily linked to reducing cognitive load, since learning occurs at an optimal workload level at which under- and overstimulation should be prevented. Further studies focusing not only on detrimental workload, but also on advantageous cognitive load (e.g., GCL) should be considered for a more complete picture on the influence of adaptivity in AR-based learning.

Concerning the display selection, the five reviewed studies indicate a preference for using OST AR over AR mirror since OST demonstrates shorter task completion times and fewer errors in procedural tasks (Karg et al. 2023). Furthermore, the reviewed studies show a preference for projection-based AR over OST and VST, with decreased cognitive load and response times (Baumeister et al. 2017; Boyce et al. 2022). The limited field of view in HMD devices should be considered when using AR for learning and training, either by considering the usage of projection-based AR or by designing the tasks accordingly. Projection-based AR excels because it seamlessly integrates virtual elements into the user's environment without occlusion, reducing cognitive load and facilitating quicker interactions. OST is furthermore advantageous over VST because it can merge digital and real-world visuals more effectively, avoiding latency and resolution issues associated with video feeds (Baumeister et al. 2017). Due to the small sample size of only five studies, the recommendations for display selection are very tentative. The specific task needs to be taken into account when deciding for an AR display.

Based on the findings from the studies, recommendations for the instructional design of AR-based learning and training can be found in Table 4. Three studies that could not be included in the established clusters looked at very specific phenomena. Czok et al. (2023), for example, looked at the integration of gamification with AR. No effects were found, and as many elements are included in the gamification, including a narrative, contextualisation and other visuals, and the exact goal of the gamification was not clear, no specific design recommendations could be established. Another study focused on a support mechanism for a very specific task, namely a metronome function for AR-based piano training (Lange-Nawka, Wünsche, and Thompson 2023). No general design recommendations could be established due to the mixture of modality and assistance, very individual results for different songs and mixed results due to learners' prior experience (see Section 3.4). In general, the study shows that it is important to consider the complexity of tasks, as different tasks may profit differently from assistance. The third study for which no general design recommendations could be established, describes the comparison of collaborative and individual settings with AR (Zhan et al. 2024). While both mental effort and learning outcomes profited from the collaborative settings, it is not clear how exactly this difference emerged. However, it is important for instructional designers to not only think about the design of the application but also potential context variables when implementing a design in the field.

4.2 | Implications for Future Research

As with any systematic literature review, the interpretation of results in this review is constrained by the studies selected according to predefined inclusion and exclusion criteria. Nonetheless, by adhering to a rigorous PRISMA guideline-based procedure, the authors have aimed to mitigate potential methodological biases. The current research landscape on cognitive load and AR learning reveals several significant challenges. This review identifies three key areas for further investigation that we consider essential for advancing a human-centred approach to AR design for learning.

1. Call for empirical evidence generated by value-added studies:

The review identified only 29 studies that examine the effects of different AR design attributes or display on cognitive load in tertiary learning and training contexts. More research on these effects is needed to provide valuable evidence for guiding technology developers, instructional designers and researchers in creating learner-centred AR solutions, particularly by focusing on cognitive concepts like cognitive load. Many studies initially included in our review were media comparison studies consistent with findings from Buchner, Buntins, and Kerres (2021, 2022). However, media comparison studies investigate the effects of a medium compared to other media rather than the underlying attributes impacting learning performance. Furthermore, none of the included studies tested established design guidelines from theoretical frameworks like CTML (Mayer 2021). While it is important to consider the specific affordances of AR, it may be helpful to base research on already established guidelines, as also suggested by Krüger and Bodemer (2022). Therefore, we suggest performing value-added studies that deliver empirical evidence by incorporating an investigation of the impact of different AR displays or instructional designs, which should be based on relevant theories or established guidelines applied in AR.

2. Call for incorporating cognitive load measurement:

The studies in our review employ a range of cognitive load and workload conceptualisations. The majority use the NASA-TLX questionnaire, which conceptualises six global constructs of workload, including mental, physical, and temporal demand, perceived performance and effort, and frustration (Hart and Staveland 1988). Initially developed to measure the task load of pilots, the questionnaire is applied in various contexts (Hart 2006). Although it continues to be widely used, criticism of the questionnaire's suitability for usability testing of technologies (Kosch et al. 2023) or in the context of learning emerges.

Other conceptualisations are based in the learning context. The differentiation into ICL, ECL and GCL allows for conclusions on learning. The three subdimensions can be assessed using questionnaires from Klepsch, Schmitz, and Seufert (2017) or Leppink et al. (2013). Moreover, the review revealed that only three studies included physiological measurements to assess cognitive load. We advocate for increased utilisation of physiological measurement methods, as they offer objective insights into learning processes and development over time (Gonnermann-Müller et al. 2024)—insights that cannot be fully captured through self-report questionnaires (Suzuki, Wild, and Scanlon 2024).

3. Call for empirical evidence including learner characteristics:

Only four of the studies addressed the effects of learner characteristics on AR learning and cognitive load. Preliminary findings within this review suggest that learner characteristics such as spatial abilities (e.g., mental rotation skills), confidence and prior experience significantly influence cognitive load and learning performance in AR contexts. Especially learner-treatment interaction study designs can offer insights into the question of which design decisions work for which learners, as also suggested by Buchner and Kerres (2023). More studies on various learner characteristics are necessary.

5 | Conclusions

In the current paper, a systematic literature review was conducted with the goal of defining evidence-based instructional design recommendations for augmented reality (AR)-based learning and training. The review highlights the potential of AR to optimise cognitive load and learning performance if it is designed well based on theory- and evidence-based design decisions. By targeting value-added and learner-treatment interaction studies, specific effects of design attributes and learner characteristics could be defined. These result in 15 design recommendations based on 29 empirical studies, which built the basis for 11 design categories Spatial Integration, Spatial Visualisation, Interaction Modality, Physical Interaction, Physical—Virtual Integration, Content Modality, Content Complexity, Generative Activity, Embodied Assistance, Adaptive Guidance, OST or projection-based in the six dimensions *Spatiality-related*, *Interaction-related*, *Contextuality-related*, *Content-related*, *Guidance-related* and *Display Selection*. The findings aim at offering practical insights for instructional designers and educators involved in implementing AR in learning and training settings.

However, more systematic research is necessary to establish a solid basis for well-grounded design principles for AR instruction. In the current review, many of the recommendations are based on the results from a few studies. To reach a more stable basis for future recommendations, three steps have been identified, with a focus on value-added and learner-treatment interaction study designs incorporating a variety of cognitive load and learning performance measurements. The categories and recommendations resulting from this review can serve as a starting point for further investigation and research in the area of AR-based learning and training.

Author Contributions

Jana Gonnermann-Müller: conceptualization, methodology, writing – original draft, writing – review and editing, visualization, investigation. **Jule M. Krüger:** conceptualization, writing – original draft, methodology, visualization, writing – review and editing, investigation.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are openly available in OSF at https://osf.io/u9rt7/?view_only=d6c4d503d593433386a84fe20c8db12c.

References

Albus, P., and T. Seufert. 2022. "Signaling in 360° Desktop Virtual Reality Influences Learning Outcome and Cognitive Load." *Frontiers in Education* 7: 916105. https://doi.org/10.3389/feduc.2022.916105.

Albus, P., and T. Seufert. 2023. "The Modality Effect Reverses in a Virtual Reality Learning Environment and Influences Cognitive Load." *Instructional Science* 51, no. 4: 545–570. https://doi.org/10.1007/s11251-022-09611-7.

Albus, P., A. Vogt, and T. Seufert. 2021. "Signaling in Virtual Reality Influences Learning Outcome and Cognitive Load." *Computers & Education* 166: 104154. https://doi.org/10.1016/j.compedu.2021.104154.

Altmeyer, K., S. Kapp, M. Thees, S. Malone, J. Kuhn, and R. Brünken. 2020. "The Use of Augmented Reality to Foster Conceptual Knowledge Acquisition in STEM Laboratory Courses—Theoretical Background and Empirical Results." *British Journal of Educational Technology* 51, no. 3: 611–628. https://doi.org/10.1111/bjet.12900.

Ariansyah, D., J. A. Erkoyuncu, I. Eimontaite, et al. 2022. "A Head Mounted Augmented Reality Design Practice for Maintenance Assembly: Toward Meeting Perceptual and Cognitive Needs of AR Users." *Applied Ergonomics* 98: 103597. https://doi.org/10.1016/j.apergo.2021.103597.

Azuma, R. T. 1997. "A Survey of Augmented Reality." *Presence Teleoperators and Virtual Environments* 6, no. 4: 355–385. https://doi.org/10.1162/pres.1997.6.4.355.

Bannert, M. 2002. "Managing Cognitive Load—Recent Trends in Cognitive Load Theory." *Learning and Instruction* 12, no. 1: 139–146. https://doi.org/10.1016/S0959-4752(01)00021-4.

Baumeister, J., S. Y. Ssin, N. A. M. ElSayed, et al. 2017. "Cognitive Cost of Using Augmented Reality Displays." *IEEE Transactions on Visualization and Computer Graphics* 23: 2378–2388. https://doi.org/10.1109/TVCG.2017.2735098.

Bautista, L., F. Maradei, and G. Pedraza. 2023. "Strategies to Reduce Visual Attention Changes While Learning and Training in Extended Reality Environments." *International Journal on Interactive Design and Manufacturing (IJIDeM)* 17, no. 1: 17–43. https://doi.org/10.1007/s12008-022-01092-9.

Bogomolova, K., M. A. T. M. Vorstenbosch, I. El Messaoudi, et al. 2023. "Effect of Binocular Disparity on Learning Anatomy With Stereoscopic Augmented Reality Visualization: A Double Center Randomized Controlled Trial." *Anatomical Sciences Education* 16, no. 1: 87–98. https://doi.org/10.1002/ase.2164.

Bottani, E., F. Longo, L. Nicoletti, et al. 2021. "Wearable and Interactive Mixed Reality Solutions for Fault Diagnosis and Assistance in Manufacturing Systems: Implementation and Testing in an Aseptic Bottling Line." *Computers in Industry* 128: 103429. https://doi.org/10.1016/j.compind.2021.103429.

Boyce, M. W., R. H. Thomson, J. K. Cartwright, et al. 2022. "Enhancing Military Training Using Extended Reality: A Study of Military Tactics Comprehension." *Frontiers in Virtual Reality* 3: 754627. https://doi.org/10.3389/frvir.2022.754627.

Buchner, J., K. Buntins, and M. Kerres. 2021. "A Systematic Map of Research Characteristics in Studies on Augmented Reality and Cognitive Load." *Computers and Education Open* 2: 100036. https://doi.org/10.1016/j.caeo.2021.100036.

Buchner, J., K. Buntins, and M. Kerres. 2022. "The Impact of Augmented Reality on Cognitive Load and Performance: A Systematic Review." *Journal of Computer Assisted Learning* 38, no. 1: 285–303. https://doi.org/10.1111/jcal.12617.

Buchner, J., and M. Kerres. 2023. "Media Comparison Studies Dominate Comparative Research on Augmented Reality in Education." *Computers & Education* 195: 104711. https://doi.org/10.1016/j.compedu.2022.104711.

Carmigniani, J., B. Furht, M. Anisetti, P. Ceravolo, E. Damiani, and M. Ivkovic. 2011. "Augmented Reality Technologies, Systems and Applications." *Multimedia Tools and Applications* 51, no. 1: 341–377. https://doi.org/10.1007/s11042-010-0660-6.

Çeken, B., and N. Taşkın. 2022. "Multimedia Learning Principles in Different Learning Environments: A Systematic Review." *Smart Learning Environments* 9, no. 1: 19. https://doi.org/10.1186/s40561-022-00200-2.

Chang, H.-Y., T. Binali, J.-C. Liang, et al. 2022. "Ten Years of Augmented Reality in Education: A meta-Analysis of (Quasi-) Experimental Studies to Investigate the Impact." *Computers & Education* 191: 104641. https://doi.org/10.1016/j.compedu.2022.104641.

Chu, H.-C., J.-M. Chen, G.-J. Hwang, and T.-W. Chen. 2019. "Effects of Formative Assessment in an Augmented Reality Approach to Conducting Ubiquitous Learning Activities for Architecture Courses." *Universal Access in the Information Society* 18, no. 2: 221–230. https://doi.org/10.1007/s10209-017-0588-y.

Clark, R. C., and R. E. Mayer. 2016. "Engagement in E-Learning." In *E-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning*, edited by R. C. Clark and R. E. Mayer, 4th ed., 219–238. Hoboken, NJ: John Wiley & Sons Inc. https://doi.org/10.1002/9781119239086.ch11.

Clark, R. E. 1983. "Reconsidering Research on Learning From Media." *Review of Educational Research* 53, no. 4: 445–459. https://doi.org/10.3102/00346543053004445.

Czok, V., M. Krug, S. Müller, J. Huwer, and H. Weitzel. 2023. "Learning Effects of Augmented Reality and Game-Based Learning for Science Teaching in Higher Education in the Context of Education for Sustainable Development." *Sustainability* 15, no. 21: 21. https://doi.org/10.3390/su152115313.

da Silva, M. M. O., J. M. X. N. Teixeira, P. S. Cavalcante, and V. Teichrieb. 2019. "Perspectives on How to Evaluate Augmented Reality Technology Tools for Education: A Systematic Review." *Journal of the Brazilian Computer Society* 25, no. 1: 3. https://doi.org/10.1186/s13173-019-0084-8.

Daling, L. M., and S. J. Schlittmeier. 2022. "Effects of Augmented Reality-, Virtual Reality-, and Mixed Reality-Based Training on Objective Performance Measures and Subjective Evaluations in Manual Assembly Tasks: A Scoping Review." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 66: 589–626. https://doi.org/10.1177/00187208221105135.

de Melo, C. M., K. Kim, N. Norouzi, G. Bruder, and G. Welch. 2020. "Reducing Cognitive Load and Improving Warfighter Problem Solving With Intelligent Virtual Assistants." *Frontiers in Psychology* 11: 554706. https://doi.org/10.3389/fpsyg.2020.554706.

Fischer, M., B. Fuerst, S. C. Lee, et al. 2016. "Preclinical Usability Study of Multiple Augmented Reality Concepts for K-Wire Placement." *International Journal of Computer Assisted Radiology and Surgery* 11, no. 6: 1007–1014. https://doi.org/10.1007/s11548-016-1363-x.

Garzón, J., and J. Acevedo. 2019. "Meta-Analysis of the Impact of Augmented Reality on students' Learning Gains." *Educational Research Review* 27: 244–260. https://doi.org/10.1016/j.edurev.2019.04.001.

Garzón, J., Kinshuk, S.Baldiris, J. Gutiérrez, and J. Pavón. 2020. "How Do Pedagogical Approaches Affect the Impact of Augmented Reality on Education? A meta-Analysis and Research Synthesis." *Educational Research Review* 31: 100334. https://doi.org/10.1016/j.edurev.2020.100334.

Gerjets, P., K. Scheiter, and G. Cierniak. 2009. "The Scientific Value of Cognitive Load Theory: A Research Agenda Based on the Structuralist View of Theories." *Educational Psychology Review* 21, no. 1: 43–54. https://doi.org/10.1007/s10648-008-9096-1.

Gonnermann-Müller, J., N. Leins, N. Gronau, and T. Kosch. 2024. "Value by Design: Reducing Cognitive Load by Using Visual Guidance in

Augmented Reality – An Eye-Tracking Study." In *ICIS 2024 Proceedings, (18)*. Bangkok Thailand: Association for Information Systems.

Hart, S. G. 2006. "Nasa-Task Load Index (NASA-TLX); 20 years later." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, no. 9: 904–908. https://doi.org/10.1177/154193120605000909.

Hart, S. G., and L. E. Staveland. 1988. "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research." In *Advances in Psychology*, vol. 52, 139–183. Amsterdam, Netherlands: Elsevier, North-Holland. https://doi.org/10.1016/S0166-4115(08)62386-9.

Hartmann, C., and M. Bannert. 2022. "Lernen in virtuellen Räumen: Konzeptuelle Grundlagen und Implikationen für künftige Forschung." *MedienPädagogik: Zeitschrift Für Theorie Und Praxis Der Medienbildung* 47: 373–391. https://doi.org/10.21240/mpaed/47/2022.04.18.X.

Herbert, B., G. Wigley, B. Ens, and M. Billinghurst. 2022. "Cognitive Load Considerations for Augmented Reality in Network Security Training." *Computers & Graphics* 102: 566–591. https://doi.org/10.1016/j.cag.2021.09.001.

Hou, L., X. Wang, L. Bernold, and P. E. D. Love. 2013. "Using Animated Augmented Reality to Cognitively Guide Assembly." *Journal of Computing in Civil Engineering* 27, no. 5: 439–451. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000184.

Howard, M. C., and M. M. Davis. 2023. "A Meta-Analysis of Augmented Reality Programs for Education and Training." *Virtual Reality* 27: 2871–2894. https://doi.org/10.1007/s10055-023-00844-6.

Hsu, T.-C. 2019. "Effects of Gender and Different Augmented Reality Learning Systems on English Vocabulary Learning of Elementary School Students." *Universal Access in the Information Society* 18, no. 2: 315–325. https://doi.org/10.1007/s10209-017-0593-1.

Huang, H.-M., T.-C. Huang, and C.-Y. Cheng. 2022. "Reality Matters? Exploring a Tangible User Interface for Augmented-Reality-Based Fire Education." *Universal Access in the Information Society* 21, no. 4: 927–939. https://doi.org/10.1007/s10209-021-00808-0.

Hwang, G.-J., L.-H. Yang, and S.-Y. Wang. 2013. "*Cognitive Load Questionnaire*." [Database record]. APA PsycTests. https://doi.org/10.1037/t34423-000.

Illing, J., P. Klinke, M. Pfingsthorn, and W. Heuten. 2021. "Less Is More! Support of Parallel and Time-Critical Assembly Tasks With Augmented Reality." *Proceedings of Mensch Und Computer* 2021: 215–226. https://doi.org/10.1145/3473856.3473861.

Kalyuga, S. 2023. "Evolutionary Perspective on Human Cognitive Architecture in Cognitive Load Theory: A Dynamic, Emerging Principle Approach." *Educational Psychology Review* 35, no. 3: 91. https://doi.org/10.1007/s10648-023-09812-7.

Karg, P., R. Stöhr, L. Jonas, J. Kreimeier, and T. Götzelmann. 2023. "Reflect-AR: Insights Into Mirror-Based Augmented Reality Instructions to Support Manual Assembly Tasks." In *Proceedings of the 16th International Conference on Pervasive Technologies Related to Assistive Environments*, 62–68. New York: Association for Computing Machinery. https://doi.org/10.1145/3594806.3594866.

Kim, J. C., T. H. Laine, and C. Åhlund. 2021. "Multimodal Interaction Systems Based on Internet of Things and Augmented Reality: A Systematic Literature Review." *Applied Sciences* 11, no. 4: 1738. https://doi.org/10.3390/app11041738.

Kim, J. H., C.-Y. Yu, K. Seo, F. Wang, and D. Oprean. 2023. "The Effect of Virtual Instructor and Metacognition on Workload in a Location-Based Augmented Reality Learning Environment." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 67, no. 1: 1550–1555. https://doi.org/10.1177/21695067231192938.

Klepsch, M., F. Schmitz, and T. Seufert. 2017. "Development and Validation of Two Instruments Measuring Intrinsic, Extraneous, and Germane Cognitive Load." *Frontiers in Psychology* 8: 1997. https://doi.org/10.3389/fpsyg.2017.01997.

Klepsch, M., and T. Seufert. 2021. "Making an Effort Versus Experiencing Load." *Frontiers in Education* 6: 645284. https://doi.org/10.3389/feduc.2021.645284.

Kosch, T., J. Karolus, J. Zagermann, H. Reiterer, A. Schmidt, and P. W. Woźniak. 2023. "A Survey on Measuring Cognitive Workload in Human–Computer Interaction." *ACM Computing Surveys* 55: 3582272. https://doi.org/10.1145/3582272.

Kozma, R. B. 1994. "Will Media Influence Learning? Reframing the Debate." *Educational Technology Research and Development* 42, no. 2: 7–19. https://doi.org/10.1007/BF02299087.

Krell, M. 2015. "Evaluating an Instrument to Measure Mental Load and Mental Effort Using Item Response Theory." *Science Education Review Letters* 2015: 1–6.

Krüger, J. M. 2023. *Augmented Reality for Learning—The Role of Contextuality, Interactivity, and Spatiality for AR-Based Learning Experiences*. Duisburg, Germany: Universität Duisburg-Essen.

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

Krüger, J. M., and D. Bodemer. 2022. "Application and Investigation of Multimedia Design Principles in Augmented Reality Learning Environments." *Information* 13, no. 2: 74. https://doi.org/10.3390/info13020074.

Krüger, J. M., A. Buchholz, and D. Bodemer. 2019. "Augmented Reality in Education: Three Unique Characteristics From a User's Perspective." In *Proceedings of the 27th International Conference on Computers in Education*, edited by M. Chang, H.-J. So, L.-H. Wong, F.-Y. Yu, and J. L. Shih, 412–422. Taiwan: Asia-Pacific Society for Computers in Education.

Krüger, J. M., K. Palzer, and D. Bodemer. 2022. "Learning With Augmented Reality: Impact of Dimensionality and Spatial Abilities." *Computers and Education Open* 3: 100065. https://doi.org/10.1016/j.caeo.2021.100065.

Lange-Nawka, D., B. C. Wünsche, and S. E. R. Thompson. 2023. "Auditory Temporal Hints in AR Piano Tutoring." In *38th International Conference on Image and Vision Computing New Zealand (IVCNZ)*, 1–6. Palmerston North, New Zealand: IEEE. https://doi.org/10.1109/IVCNZ61134.2023.10343724.

Lenz, L. S., A. R. Fender, J. Chatain, and C. Holz. 2024. "Comparing Synchronous and Asynchronous Task Delivery in Mixed Reality Environments." *IEEE Transactions on Visualization and Computer Graphics* 30, no. 5: 2776–2784. https://doi.org/10.1109/TVCG.2024.3372034.

Leppink, J., F. Paas, C. P. M. Van der Vleuten, T. Van Gog, and J. J. G. van Merriënboer. 2013. "Development of an Instrument for Measuring Different Types of Cognitive Load." *Behavior Research Methods* 45, no. 4: 1058–1072. https://doi.org/10.3758/s13428-013-0334-1.

Li, W.-C., J. Zhang, S. Court, P. Kearney, and G. Braithwaite. 2022. "The Influence of Augmented Reality Interaction Design on Pilot's Perceived Workload and Situation Awareness." *International Journal of Industrial Ergonomics* 92: 103382. https://doi.org/10.1016/j.ergon.2022.103382.

Lin, H.-Y., and S.-C. Tsai. 2021. "Student Perceptions Towards the Usage of AR-Supported STEMUP Application in Mobile Courses Development and Its Implementation Into English Learning." *Australasian Journal of Educational Technology* 37, no. 3: 88–103. https://doi.org/10.14742/ajet.6125.

Lin, Y., and Z. Yu. 2023. "A meta-Analysis of the Effects of Augmented Reality Technologies in Interactive Learning Environments

(2012–2022).” *Computer Applications in Engineering Education* 41: cae.22628. https://doi.org/10.1002/cae.22628.

Lu, S., Y. P. Sanchez Perdomo, X. Jiang, and B. Zheng. 2020. “Integrating Eye-Tracking to Augmented Reality System for Surgical Training.” *Journal of Medical Systems* 44, no. 11: 192. https://doi.org/10.1007/s10916-020-01656-w.

Maitz, F., L. Ribeiro Skreinig, D. Kalkofen, and S. C. Wriessnegger. 2023. “Towards Neuroadaptive Augmented Reality Piano Tutorials.” In *IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE)*, 450–455. Milano, Italy: IEEE. https://doi.org/10.1109/MetroXRAINE58569.2023.10405829.

Mayer, R. E. 2019. “Thirty Years of Research on Online Learning.” *Applied Cognitive Psychology* 33, no. 2: 152–159. https://doi.org/10.1002/acp.3482.

Mayer, R. E., and Fiorella, L. (Eds.). 2021. “Multimedia Learning with Media.” In *The Cambridge Handbook of Multimedia Learning*, 437–565. Cambridge: Cambridge University Press.

Mayer, R. E. 2024. “The Past, Present, and Future of the Cognitive Theory of Multimedia Learning.” *Educational Psychology Review* 36, no. 1: 8. https://doi.org/10.1007/s10648-023-09842-1.

Milgram, P., and F. Kishino. 1994. “A Taxonomy of Mixed Reality Visual Displays.” *IEICE Transactions on Information and Systems* E77-D, no. 12: 1321–1329.

Mystakidis, S., A. Christopoulos, and N. Pellas. 2021. “A Systematic Mapping Review of Augmented Reality Applications to Support STEM Learning in Higher Education.” *Education and Information Technologies* 27: 1883–1927. https://doi.org/10.1007/s10639-021-10682-1.

Niegemann, H. M., and S. Heidig. 2012. “Multimedia Learning.” In *Encyclopedia of the Sciences of Learning*, edited by N. M. Seel, 2372–2375. New York: Springer. https://doi.org/10.1007/978-1-4419-1428-6_285.

Nowak, A., P. Knierim, A. Romanowski, A. Schmidt, and T. Kosch. 2020. “What Does the Oscilloscope Say?: Comparing the Efficiency of In-Situ Visualisations During Circuit Analysis.” In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–7. New York: Association for Computing Machinery. https://doi.org/10.1145/3334480.3382890.

Paas, F. G. W. C., J. J. G. van Merriënboer, and J. J. Adam. 1994. “Measurement of Cognitive Load in Instructional Research.” *Perceptual and Motor Skills* 79, no. 1: 419–430. https://doi.org/10.2466/pms.1994.79.1.419.

Page, M. J., J. E. McKenzie, P. M. Bossuyt, et al. 2021. “The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews.” *Systematic Reviews* 10, no. 1: 89. https://doi.org/10.1186/s13643-021-01626-4.

Rauschnabel, P. A., R. Felix, C. Hinsch, H. Shahab, and F. Alt. 2022. “What Is XR? Towards a Framework for Augmented and Virtual Reality.” *Computers in Human Behavior* 133: 107289. https://doi.org/10.1016/j.chb.2022.107289.

Sasikumar, P., S. Chittajallu, N. Raj, H. Bai, and M. Billinghurst. 2021. “Spatial Perception Enhancement in Assembly Training Using Augmented Volumetric Playback.” *Frontiers in Virtual Reality* 2: 698523. https://doi.org/10.3389/frvir.2021.698523.

Simmen, Y., T. Eggler, A. Legath, D. Agotai, and H. Cords. 2023. “Non-overlayed Guidance in Augmented Reality: User Study in Radio-Pharmacy.” In *Intelligent Human Computer Interaction*, edited by H. Zaynidinov, M. Singh, U. S. Tiwary, and D. Singh, 516–526. Switzerland: Springer Nature. https://doi.org/10.1007/978-3-031-27199-1_52.

Sommerauer, P., and O. Müller. 2014. “Augmented Reality in Informal Learning Environments: A Field Experiment in a Mathematics Exhibition.” *Computers & Education* 79, no. 2014: 59–68. https://doi.org/10.1016/j.compedu.2014.07.013.

Surry, D. W., and D. Ensminger. 2001. “What's Wrong With Media Comparison Studies?” *Educational Technology* 41, no. 4: 32–35.

Suzuki, Y., F. Wild, and E. Scanlon. 2024. “Measuring Cognitive Load in Augmented Reality With Physiological Methods: A Systematic Review.” *Journal of Computer Assisted Learning* 40: 375–393. https://doi.org/10.1111/jcal.12882.

Sweller, J., J. J. G. van Merriënboer, and F. Paas. 2019. “Cognitive Architecture and Instructional Design: 20 Years Later.” *Educational Psychology Review* 31, no. 2: 261–292. https://doi.org/10.1007/s10648-019-09465-5.

Thees, M., K. Altmeyer, S. Kapp, et al. 2022. “Augmented Reality for Presenting Real-Time Data During students' Laboratory Work: Comparing a Head-Mounted Display With a Separate Display.” *Frontiers in Psychology* 13: 804742. https://doi.org/10.3389/fpsyg.2022.804742.

Thees, M., S. Kapp, M. P. Strzys, F. Beil, P. Lukowicz, and J. Kuhn. 2020. “Effects of Augmented Reality on Learning and Cognitive Load in University Physics Laboratory Courses.” *Computers in Human Behavior* 108: 106316. https://doi.org/10.1016/j.chb.2020.106316.

Uruthiralingam, U., and P. M. Rea. 2020. “Augmented and Virtual Reality in Anatomical Education—A Systematic Review.” In *Biomedical Visualisation*, edited by P. M. Rea, vol. 1235, 89–101. Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-030-37639-0_5.

Van den Bergh, J., and F. Heller. 2020. “Wearable Touchscreens to Integrate Augmented Reality and Tablets for Work Instructions?” In *Human-Centered Software Engineering*, edited by R. Bernhaupt, C. Ardito, and S. Sauer, 199–206. Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-030-64266-2_13.

van Merriënboer, J. J. G., O. Jelsma, and F. G. W. C. Paas. 1992. “Training for Reflective Expertise: A Four-Component Instructional Design Model for Complex Cognitive Skills.” *Educational Technology Research and Development* 40, no. 2: 23–43. https://doi.org/10.1007/BF02297047.

Wang, Z., X. Bai, S. Zhang, et al. 2022. “A Comprehensive Review of Augmented Reality-Based Instruction in Manual Assembly, Training and Repair.” *Robotics and Computer-Integrated Manufacturing* 78: 102407. https://doi.org/10.1016/j.rcim.2022.102407.

Werrlich, S., P.-A. Nguyen, and G. Notni. 2018. “Evaluating the Training Transfer of Head-Mounted Display Based Training for Assembly Tasks.” In *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference*. 297–302. New York: Association for Computing Maschinery. https://doi.org/10.1145/3197768.3201564.

Wickens, C. D. 1981. Processing Resources in Attention, Dual Task Performance, and Workload Assessment. In *Multiple task performance*, edited by D. Damos, 1st ed., 3–16. London: CRC Press.

Xu, W.-W., C.-Y. Su, Y. Hu, and C.-H. Chen. 2022. “Exploring the Effectiveness and Moderators of Augmented Reality on Science Learning: A meta-Analysis.” *Journal of Science Education and Technology* 31, no. 5: 621–637. https://doi.org/10.1007/s10956-022-09982-z.

Yu, D., X. Yao, K. Yu, D. Du, J. Zhi, and C. Jing. 2022. “The Effect of Presentation Position and Multichannel Approach on Learning Performance in the Use of an Augmented Reality–Head Worn Display Interface for Train-Driving Training.” *Interactive Learning Environments* 32, no. 5: 1–16. https://doi.org/10.1080/10494820.2022.2124426.

Yu, Z. 2023. “Meta-Analyses of Effects of Augmented Reality on Educational Outcomes Over a Decade.” *Interactive Learning Environments* 32, no. 8: 4739–4753. https://doi.org/10.1080/10494820.2023.2205899.

Zhan, Z., X. Zhou, S. Cai, and X. Lan. 2024. “Exploring the Effect of Competing Mechanism in an Immersive Learning Game Based on

Augmented Reality." *Journal of Computers in Education*, 1–27. https://doi.org/10.1007/s40692-024-00317-y.

Zhang, J., G. Li, Q. Huang, Q. Feng, and H. Luo. 2022. "Augmented Reality in K–12 Education: A Systematic Review and Meta-Analysis of the Literature From 2000 to 2020." *Sustainability* 14, no. 15: 9725. https://doi.org/10.3390/su14159725.

Zhang, J., X. Xia, R. Liu, and N. Li. 2021. "Enhancing Human Indoor Cognitive Map Development and Wayfinding Performance With Immersive Augmented Reality-Based Navigation Systems." *Advanced Engineering Informatics* 50: 101432. https://doi.org/10.1016/j.aei.2021.101432.

Zhao, M., S. Ma, Y. Liu, and W. Song. 2023. "Multimedia-Based Informal Learning in Museum Using Augmented Reality." In *Image and Graphics Technologies and Applications*, edited by W. Yongtian and W. Lifang, 53–64. Singapore: Springer Nature. https://doi.org/10.1007/978-981-99-7549-5_5.

Appendix 1

An Overview of the Scope and Limitations of Systematic Reviews and Meta-Analyses Opens the Research Gap This Systematic Literature Review Aims to Solve

Scope of recent systematic reviews and meta-analysis	Limitations	Reference
A systematic analysis of 53 articles between 1999 and 2022 on design principles grounded in cognitive load theory.	No further distinction of findings between different technologies (VR, MR and AR). No further distinction of results between media comparison studies and value-added studies.	Bautista, Maradei, and Pedraza (2023)
A systematic map of 60 articles (till 2019) on the effects of AR learning on cognitive load.	Only bibliometric, geographical, and methodological overview of the literature. No systematisation of design effects.	Buchner, Buntins, and Kerres (2021)
A systematic analysis of 58 studies (till 2019) on the effects of AR learning compared to other learning methods on cognitive load.	No further distinction of results between media comparison studies and value-added studies.	Buchner, Buntins, and Kerres (2022)
Meta-analysis of 70 articles (between 2012 and 2022) on the comparison of AR with traditional tools concerning the effects of AR on learning.	No further distinction of results between media comparison studies and value-added studies.	Lin and Yu (2023)

Appendix 2

Overview of AR Devices and Displays Used in the Set of Studies Identified ($N = 29$) in This Review

AR device	*N*	AR display	*N*
Wearable AR	21	OST	18
		VST (HMD)	3
Mobile AR	8	VST tablet	6
		VST smartphone	2
Stationary AR	4	Projection-based AR	3
		AR Mirror	1
In-body	0		
On-body	0		

Note: Classification following Rauschnabel et al. (2022). Due to some studies comparing multiple AR displays, the total number of AR devices and displays exceeds 29.
Abbreviations: OST, optical-see-through; VST, video-see-through.

Appendix 3

Cognitive Load Measurements Across the Full Study Sample ($N = 27$)

Assessment	Assessment method(s)	N	Specific measurement	Reference
Mono-method	Questionnaire assessing workload	18	NASA-TLX (Hart and Staveland 1988; Hart 2006)	Bogomolova et al. (2023), Boyce et al. (2022)*, de Melo et al. (2020), Herbert et al. (2022), Illing et al. (2021), Karg et al. (2023)*, Kim et al. 2023; Lenz et al. (2024), Li et al. (2022), Maitz et al. (2023), Nowak et al. (2020)*, Sasikumar et al. (2021), Simmen et al. (2023), Van den Bergh and Heller (2020), Werrlich, Nguyen, and Notni (2018), Yu et al. (2022), Zhang et al. (2021)
	Questionnaires assessing mental load and mental effort	4	Mental load and mental effort (Hwang et al. 2013)	Chu et al. (2019), Huang, Huang, and Cheng (2022), Zhan et al. (2024)
			Mental load and mental effort (Krell 2015)	Czok et al. (2023)
	Cognitive Load (ICL, ECL, GCL)	1	Cognitive load distinguished into ICL, ECL, GCL (Klepsch, Schmitz, and Seufert 2017)	Krüger, Palzer, and Bodemer (2022)
Multi-method	Questionnaire + questionnaire	1	Cognitive load (Klepsch, Schmitz, and Seufert 2017), NASA-TLX (Hart und Staveland, 1988)	Krüger and Bodemer (2020)
	Questionnaire + secondary task	2	Total cognitive load (Paas, van Merriënboer, and Adam 1994), response time in a dual task	Baumeister et al. (2017)*
	Questionnaire + physiological	3	NASA-TLX (Hart und Staveland, 1988), skin conductance (EDA), heart Rate	Ariansyah et al. (2022)
			Questionnaire (self-developed), eye-tracking	Lange-Nawka, Wünsche, and Thompson (2023)
			NASA-TLX (Hart und Staveland, 1988), heart rate variability	Zhao et al. (2023)

Note: Study designs comparing multiple AR displays are marked with *; Two papers (Sasikumar et al. 2021; Baumeister et al. 2017) containing two studies are considered in the analysis.

Appendix 4

Overview of Cognitive Load/Workload, Learning, and Performance Effects in Studies Examining Learner-Treatment Interactions ($N=4$), Clustered by Display Type

AR display	Learner characteristic	Design dimension	Effects on cognitive load	Effects on learning/ performance	Reference
Optical see-through	Confidence	Virtual instructor: present vs. absent	Workload: Only overconfident students learning with a virtual instructor showed a strong negative correlation ($p=0.02$, $r=-0.79$) between workload and learning performance	Learning performance: Sign. interaction effect ($p<0.01$), showing that only underconfident students performed better with a virtual instructor than without	Kim et al. (2023)
	Spatial abilities	Spatial visualisation: stereoscopic 3D vs. monoscopic 3D	Workload: No interaction of visual–spatial abilities (MRT scores) with the type of spatial visualisation	Knowledge: No sign. interaction of visual–spatial abilities (MRT score) with the type of spatial visualisation concerning knowledge test scores	Bogomolova et al. (2023)
Video see-through	Spatial abilities	Spatial visualisation: 3D vs. 2D virtual model	Cognitive load: No sign. interaction effect of spatial abilities and spatial visualisation on GCL or ECL	Knowledge: Sign. interaction effect ($p=0.03$, $\omega^2=0.02$), showing that only for high ($p<0.01$) and average ($p=0.02$) spatial abilities (MRT score) spatial relations knowledge is increased with 3D compared to 2D visualisation	Krüger, Palzer, and Bodemer (2022)
	Prior experience	Other: Visual representation of instructions for playing the piano without vs. with auditory metronome	CL: Only for inexperienced piano players, CL is increased with a metronome in comparison to no metronome ($p=0.04$) during songs #4, #5, #6 Eye movement: Only for experienced piano players, with metronome eye movement is decreased in song #1 ($p=0.03$), song #2 ($p=0.04$) and song #3 ($p=0.03$); focus switches are reduced in song #3 ($p=0.03$) and increased in song #6 ($p=0.02$); focus time is increased in song #3 ($p=0.03$) and reduced in song #6 ($p=0.02$); focus depth is increased in song #4 ($p=0.02$)	Errors: Only for inexperienced piano players, missed end counts increased with a metronome in song #4 and decreased in song #5; Only for experienced piano players, start time errors are reduced with metronome during song #1 ($p=0.01$), song #2 ($p=0.03$) and song #3 ($p=0.03$), and extra end counts are reduced during song #1 ($p=0.04$ and $p=0.05$) and increased during song #5 ($p=0.04$)	Lange-Nawka, Wünsche, and Thompson (2023)