Jan 17th, 12:00 AM

Tell Me and I Forget, Involve Me and I Learn: Design and Evaluation of a Multimodal Conversational Agent for Supporting Distance Learning

Julian Schuir
Universität Osnabrück, Germany, julian.schuir@uni-osnabrueck.de

Eduard Anton
Universität Osnabrück, Germany, eduard.anton@uni-osnabrueck.de

Marian Eleks
Universität Osnabrück, Germany, meleks@uni-osnabrueck.de

Frank Teuteberg
Universität Osnabrück, Germany, frank.teuteberg@uos.de

Tell Me and I Forget, Involve Me and I Learn: Design and Evaluation of a Multimodal Conversational Agent for Supporting Distance Learning

Julian Schuir, Eduard Anton, Marian Eleks, and Frank Teuteberg

University of Osnabrück, Osnabrück, Germany
{julian.schuir, eduard.anton, meleks, frank.teuteberg}@uni-osnabrueck.de

Abstract. The COVID-19 pandemic has shifted children's learning routines from schools to their own homes, necessitating learning support solutions. This paper reports on a design science research project that combines augmented reality with a conversational agent to assist schoolchildren in learning complex subjects by providing verbal descriptions and interactive animations. Drawing on the theoretical foundations of multimedia learning, we derive three design principles to resolve seven issues associated with distance learning. The instantiated artifact augments text-based learning resources and facilitates learning in a context-sensitive manner through multimodal output. The proof-of-concept evaluation with 11 experienced teachers and researchers in the field of didactics confirms the usefulness of these design principles and suggests refinements of the artifact.

Keywords: distance education, augmented reality, conversational agents

1 Introduction

The COVID-19 pandemic has severely affected education systems worldwide, displacing 168 million children from school for at least a year [1]. Although governments and schools attempt to ensure a regular education through distance learning measures (e.g., online sessions), this goal is rarely achieved. In Germany, for example, only 6% of schoolchildren were able to attend a daily online teaching session during early lockdown periods. Thus, the responsibility often falls on parents who are homeschooling to explain the remote learning materials to their children or to assist with learning issues [2]. A study with 1,000 German households reveals the consequences of transferring the responsibility for education to parents and schoolchildren; not only is it a major psychological issue and time burden, the ability to cope with this responsibility significantly depends on the socio-economic background of the household. Therefore, the pandemic situation particularly disadvantages children of non-academics or low-achieving children [3].

This educational issue indicates a need to support schoolchildren in meaningful learning while unburdening parents throughout the ongoing pandemic. Meaningful learning refers to a "deep understanding of the material, which includes attending to important aspects of the presented material, mentally organizing it into a coherent

17th International Conference on Wirtschaftsinformatik,
February 2022, Nürnberg, Germany

cognitive structure, and integrating it with relevant existing knowledge" [4, p. 43]. Modern approaches to support schoolchildren in meaningful learning often include multimedia comprising videos, images, and text delivered on various devices, platforms, and applications. However, this type of multimedia presentation is laborious to produce and inflexible in terms of adapting the learning content. Furthermore, it provides no guarantee of learning success as it often overwhelms learners' cognitive capacity [5]. Nevertheless, Mayer [4], [6] demonstrates that multimedia learning can provide significant benefits without overloading the cognitive capacity as long as verbal narratives are concise and combined with visual animations.

Since the experiments of Mayer in the 1990s [e.g., 6], technology has become increasingly sophisticated in its ability to augment learning materials with verbal narratives and representations by leveraging advancements in natural language processing (NLP) and three-dimensional (3D) visualizations. Technologies such as conversational agents (CAs) can be employed as text- or speech-based personal assistants to enable dialogue with the learner, and provide advice [7-8], while augmented reality (AR) provides the capability to visualize 3D models to facilitate the understanding of complex phenomena [9-10]. Deploying CAs in pedagogical settings is not a new concept; for instance, Hobert [11] designed a chatbot-based learning system for students to support programming instructions. Pedagogical CAs are valuable for distance learning as they encourage schoolchildren to actively engage with the learning material at a personalized level of support that is scalable irrespective of the number of learners or the presence of teachers [8], [11-12]. AR has also been used to visualize learning content, for instance, in natural sciences [13-14]. A literature review by Diegmann et al. [15] reveals that AR increases learner motivation, stimulates learner engagement, and enhances user satisfaction. As such, AR is particularly suitable for student-centered learning by allowing to "explore knowledge and solve problems autonomously" [15, p. 1548]. Given these capabilities of CAs and AR, we propose that combining both technologies in the sense of a multimodal CA is well-suited to provide schoolchildren with context-sensitive support in distance learning. However, to the best of our knowledge, pairing CAs with AR to augment existing learning resources remains a research gap. Based on this research gap and the issues associated with distance learning, we formulate the following research question:

How can we develop a multimodal CA that supports schoolchildren in distance learning for comprehension issues of learning materials?

Following a design science research (DSR) approach [16], we address this research question by developing a multimodal CA to provide schoolchildren with explanations and visualizations for meaningful learning while reducing media disruption between learning materials and digital resources. In this paper, we focus on a use case from a middle school biology curriculum related to the anatomy of the human body.

2 Related Work

COVID-19 has forced most educational institutions to quickly replace the face-to-face sessions with alternative digital solutions [17]. However, as Rodriguez-Abitia [18] remarks, “[t]ransitioning from a face-to-face learning environment to distance education involves more than conducting full classes via videoconferencing tools.” Confirming this notion, Rai and Selnes [19] emphasize that classroom learning is only one of several learning activities, and each task must be aligned with the employed technology as there is no universal pedagogical strategy. Nevertheless, so far most proposed interventions described in information systems (IS) research to ease the COVID-19-induced educational situation have a platform-dependent online focus or relate to some form of blended learning aimed at higher education [18], [20-22]. Notwithstanding, digital courses and learning platforms often reinforce the digital divide due to infrastructural and skill shortages in adopting online offerings [22]. Therefore, printed learning and reading materials remain an essential strategy for reaching schoolchildren despite infrastructural, sociodemographic, or learning disadvantages [23]. Nonetheless, engaging with the reading materials entails a considerable amount of personal responsibility for schoolchildren and their parents to ensure that texts are read and comprehended [3]. Therefore, supporting schoolchildren while they read textbooks remains a major challenge for educators. This study responds to this issue and calls for additional research on how technology can enhance specific learning activities by combining AR and CAs [19], [24].

According to Milgram and Kishino [25, p. 3], AR constitutes a subset of mixed reality environments, “in which real and virtual world objects are presented together within a single display” by superimposing visual 3D in real-time [26]. Solutions for the visual augmentation of text-based material are scarce. For example, Ibáñez et al. [27] used AR to visualize electromagnetic phenomena to support experimental learning and increase learning effectiveness. In another example, Di Serio et al. [28] promoted motivation and satisfaction levels by employing AR solutions in a visual arts context. Despite the additional information provided by visual stimuli, the use of AR-based learning solutions does not guarantee more effective learning [19]. Engaging with a visual 3D representation often requires prior knowledge, even when the model itself is intended to educate. This interdependence constitutes the representation dilemma [29]. CAs can solve this dilemma as they are able to verbally or textually summarize a 3D model and its associated context [8]. CAs are “both text-based and voice-based automated dialog systems that can interact with a human user via natural language and answer questions on specific topics” [30, p. 2]. CAs are widely used as general-purpose assistants, for example, to control household appliances, or are employed in domain-specific areas such as customer service or educational settings [31]. Depending on their technological configuration and their targeted group, CAs can support all facets of the learning process (e.g., preparation, reflection) [12]. For instance, CAs can be embedded as smart personal tutors in e-learning systems to explain tasks and answer learners’ questions [11], [32]. Their capability to recognize users’ intent from textual or auditory input enables the output of appropriate static or dynamic responses in visual, textual, auditory, or multimodal form [12]. Yet, the visual output of CAs mostly remains limited to the visualization of 2D avatars rather than educational content itself [33].

Drawing on the cognitive theory of multimedia learning [34], we consider that learners use an auditory and a visual channel to process information. Each of these channels has a limited capacity to select and integrate information during the active learning process [34]. Against this background, we propose that providing sparse and carefully considered auditory and visual stimuli through multimodal CAs holds promising potential to support meaningful learning by allowing schoolchildren to avoid investing cognitive capacity to log into an educational platform, watch a video, and learn on a screen. Instead, we can leverage the principle of temporal contiguity, where the narration and animation remain close to the text without dropping from the cognitive process, which promotes meaningful learning [29].

3 Research Approach

To respond to our research question, we employed the DSR approach proposed by Peffers et al. [35]. This approach provides a framework to solve real-world problems while contributing to the body of knowledge by iteratively refining the design knowledge [16]. Figure 1 illustrates the research approach comprising six main stages.

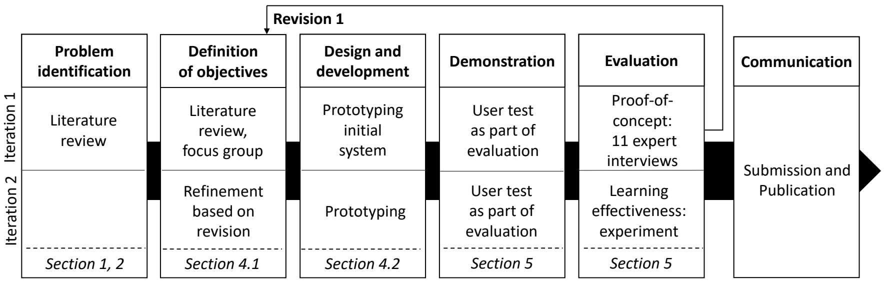

Figure 1. Design science research approach based on Peffers et al. [35]

The first step of a DSR project is problem identification [35]. We therefore began by conducting a systematic literature review according to the guidelines of vom Brocke et al. [36] to identify the issues surrounding distance learning for schoolchildren during the COVID-19 pandemic. To this end, we searched for both practical and academic publications published since January 2020, employing keywords such as “distance learning,” “COVID-19 education,” and “home schooling.” Subsequently, to align the solution space with the problem space, we conducted another literature review in the fields of human-computer interaction, computer science, and IS to identify justificatory design knowledge [37]. For this purpose, we queried Scopus, EbscoHost, Google Scholar, and ResearchGate by applying the search string *((AR OR “augmented reality” OR “mixed reality” OR MR OR “conversational agent” OR CA) AND (“education” OR pedagoc* OR didactic*)).* To narrow the focus of the literature search, we restricted the search to the fields title, abstract, and keywords, where possible. To be eligible for inclusion in the sample, the literature must a) be written in English, b) examine design approaches for AR or CAs in educational settings, and c) include a qualitative or quantitative research approach. Studies not fulfilling these criteria were removed from the sample while screening titles, abstracts, and full-texts. To ensure objectivity, two researchers independently executed this process. The literature search yielded a total of 37 papers and six applications with relevance to our project.

In the second step, we organized the issues outlined for distance learning with the aid of Webster and Watson's concept-centric approach [38]. To derive the meta-requirements (MRs), a focus group of seven IS researchers, including experts from the domain of CAs, carefully debated the literature review findings [39]. At the beginning of this discussion, one researcher presented the concept matrix before collectively deriving the MRs. Based on the MRs, we deduced design principles (DPs) by incorporating the justificatory knowledge, drawing on the syntax of Gregor et al. [37].

In the third step, the DPs were instantiated by developing an information technology (IT) artifact, which constitutes one of the main contributions of DSR projects along with constructs, methods, and models [40]. To verify the problem-solving ability of our concept, we pursued a human-risk and effectiveness strategy [41]: After the initial design phase, we conducted 11 interviews with experienced teachers and researchers to assess the utility of our DPs before proceeding to the second iteration. The second evaluation will comprise an experiment to investigate the learning effectiveness.

4 Artifact Description

4.1 Issues, Meta-Requirements, and Design Principles

At the beginning of the COVID-19-induced lockdown periods, there was a shortage of engaging online courses (**Issue 1**), so only a small percentage of German schoolchildren received one learning unit per day [2]. To accommodate the short-term transition to distance learning, teachers instead distributed text-based learning resources (e.g., textbooks and worksheets) to their schoolchildren [42]. Prior research has demonstrated that schoolchildren often prefer paper-based reading materials to digital learning [43-44]. Nevertheless, educators often complement paper-based materials with multimedia e-learning solutions that are heavily computer-dependent, requiring schoolchildren to learn textual content on a screen [22], [45]. Such e-learning solutions pose the risk of cognitive overload [5]. This cognitive overload (**Issue 2**) is caused by the fact that schoolchildren must cognitively process both essential and incidental information [4]. As a result, it is especially detrimental to schoolchildren who already suffer from learning problems. These schoolchildren in particular need the support of their teachers or family members [2]. However, in distance education, the lack of teacher and parent support (**Issue 3**) raises difficulties as schoolchildren do not receive the individual support that is provided in schools in a face-to-face setting [46].

Given the issues outlined above, the artifact should assist schoolchildren in a context-sensitive manner with individualized support (**MR1**) [47]. To discover an effective way to support schoolchildren, we draw on Mayer's [34] multimedia learning theory by providing both auditory and animated visual media (**MR2**). While multimedia learning constitutes an effective measure for explaining cause-and-effect relationships, extraneous material (i.e., embellished content) must be avoided to prevent oversaturating the cognitive system [4], [34]. To this end, Mayer and Moreno [4] propose the weeding principle, which describes a load-reducing technique in which verbal explanations are kept concise and supported by brief, animated visualizations. Mayer and Moreno [4] demonstrated that schoolchildren who received weeding content learned more effectively than schoolchildren who received more complex material. Therefore, the artifact should employ the weeding strategy to assist schoolchildren (**MR3**) [4]. Although there are multimedia applications for explaining complex phenomena (e.g., SchulAR [48]), prior studies have reported a contextual misalignment between these applications and existing curricula (**Issue 4**) [47]. Moreover, existing e-learning applications lack flexibility due to missing content administration capacities (**Issue 5**) [18]. To align with existing curricula and resources (e.g., exercise sheets; **MR4**), the artifact must additionally support content customization (**MR5**). Considering MR1–5, we propose **DP1** concerning individual multimedia support:

To enable schoolchildren to obtain individual support that aligns with existing learning resources, provide the system with capabilities to administer multimedia content consisting of concise and compact verbal descriptions and 3D animations because weeding facilitates the comprehension of complex relationships and avoids cognitive overload [4].

At the user level, the low digital literacy of schoolchildren problematizes the use of e-learning formats (**Issue 6**) [2]. Therefore, we aim to meet the users' capabilities by providing an intuitive interaction (**MR6**). Xu and Warschauer [49] have demonstrated that speech recognition facilitates machine-learner interactions as voice-based interfaces require lower digital literacy than, for instance, computers. Hence, our artifact must understand varying user intents in natural language (**MR7**) [50]. The application must further respond to the intents through suitable visual and auditory output channels [4]. Compared to two-dimensional interfaces (e.g., laptops), AR technologies provide visual learning content more tangibly by enabling animated 3D models to be explored from different perspectives. In educational settings, the rich visualization capabilities of AR facilitate the understanding of complex, otherwise invisible phenomena—for instance, by spatially illustrating anatomic structures [51]. We hence aim to integrate an AR-based interface that visualizes the schoolchildren's intents in an animated fashion (**MR8**). Drawing on the weeding concept [4], our artifact must additionally include capabilities to complement the visualization through speech-based output delivery in the sense of multimodal output to address the auditory information processing channel (**MR9**). We therefore suggest **DP2** for multimodal user interface (UI) design:

To enable schoolchildren with low digital literacy to intuitively interact with the system and to support weeding, provide the system with a CA that automatically detects natural language intents and visualizes these intents via AR while explaining them auditorily because voice-based input simplifies the interaction with the system while multimodal output promotes understanding of the learning content [49-51].

Another issue that challenges distance learning is the shortage of access to modern technical infrastructure (**Issue 7**). An OECD report from 2020 indicates that a significant share of schoolchildren have restricted access to computers, tablets, and smartphones, especially among low-income households. In larger families, for example, children share devices [2]. Consequently, our artifact must demonstrate high device compatibility to ensure support for as many schoolchildren as possible (**MR10**). In view of the plurality of schoolchildren, the underlying architecture also needs to be scalable, allowing different users to access the services simultaneously (**MR11**). Given these MRs, we propose **DP3**:

To enable multiple schoolchildren with restricted technological resources to access the application independent of their devices in distance learning, provide the system with a centralized infrastructure that can be accessed via web browsers because this enables simultaneous access for multiple schoolchildren with a variety of devices.

In summary, we identified seven issues that were translated into 11 MRs. Based on these MRs, we derived three central DPs concerning the provision of multimedia learning support (**DP1**), the multimodal UI (**DP2**), and the system infrastructure (**DP3**). Figure 2 summarizes the issues, MRs, and DPs.

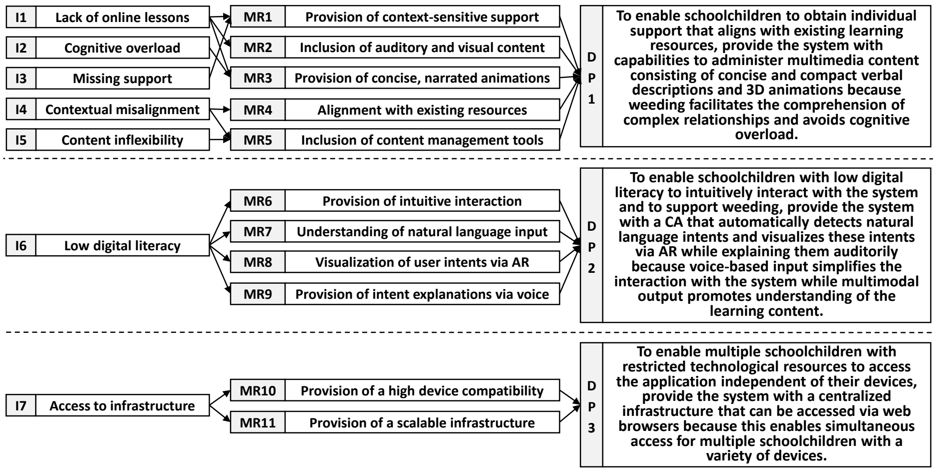

Figure 2. Issues, meta-requirements, and design principles following Gregor et al. [37]

4.2 Application

The three derived DPs guided the development of the ExplainAR web application. To illustrate the artifact, this section begins by describing the application's main features before elaborating on the technical instantiation of the DPs.

ExplainAR is a multimodal CA that processes speech-based follow-up questions from schoolchildren and answers them by combining verbal explanations with AR-based 3D visualizations. Teachers can pre-define the multimedia content consisting of short verbal explanations, 3D models, and keywords in a web portal. As the teacher is responsible for the content administration, the web-based CA can act as a tutor for different categories of knowledge (e.g., factual or conceptual) of different subject areas targeted at schoolchildren. As such, the CA addresses different facets of learning (e.g., repeat or reflection) [12]. Preparing biology lessons, teachers can supplement their worksheets with a marker provided by the web portal. When a schoolchild is experiencing comprehension difficulties related to the learning content, he or she can access the ExplainAR web application. Once the application is launched, the schoolchild views a live stream of the camera. When the schoolchild targets the marker with the camera, a virtual question mark is displayed, and the microphone is activated. The CA processes the natural language queries of the schoolchild, recognizes the user's intent, displays the corresponding 3D model (e.g., spleen) in the field of view, and verbally describes the function of the respective organ.

Multimedia learning support (DP1). To enable content administration, we prototypically implemented a web portal that includes a collection of 3D models to choose from and a text editor through which any text can be entered to describe the model or related content. The selection of 3D models can be retrieved from external databases that can be connected via an open interface. In addition, the educator defines sample questions and phrases related to the 3D model. The CA engine leverages the sample questions to further train the model and recognize the corresponding intent. The teacher-created content consists of a) a 3D model, b) an explanation in text form, c) sample phrases and questions, and d) keywords, that are stored in a database.

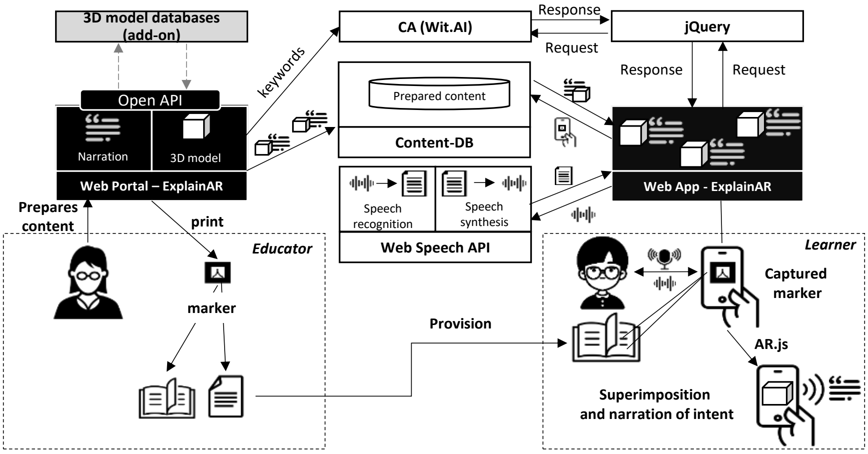

Figure 3. System infrastructure

Multimodal UI and infrastructure (DP2 and DP3). The database and the web application are hosted on a central server (cf. Figure 3). Upon launching ExplainAR via a web browser, the device camera automatically opens. The schoolchild moves the camera frame over the marker and can then ask a question such as "What is the function of the liver?" or make a request such as "Show me where the liver is located in the body." The speech is converted to text by a web speech API [52] via natural language understanding and NLP and returned to the application, which in turn passes the text to the CA (operationalized by Wit.AI [53]). The CA identifies the schoolchild's intent and returns it to the application. Based on the identified intent, a specific model-text combination is returned from the database, and the 3D model is superimposed onto the field of view of the used device powered by AR.js [54] and the educator's prepared narration is transcribed into speech and played through the device's speakers using the Web Speech API [52]. The auditory and visual stimuli complement the paper-based textual information. To prevent the conversation from ending abruptly if the user's input is not understood, we opted for a conversation recovery strategy by combining social cues with follow-up requests (cf. Figure 4). The CA apologizes for not understanding the user input and kindly requests to repeat the request [55].

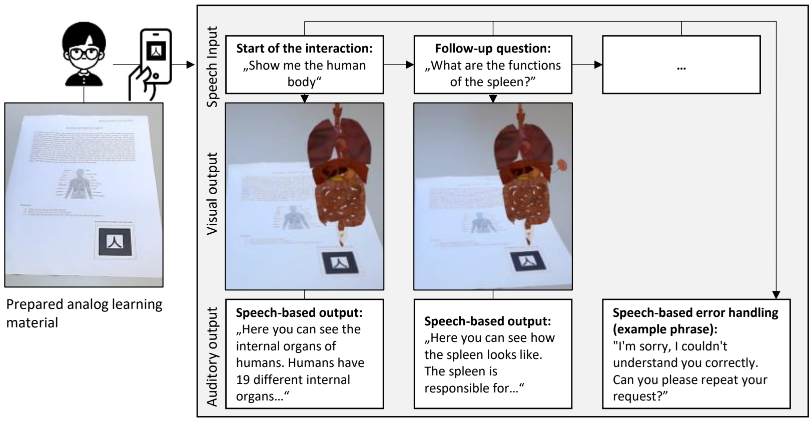

Figure 4. Learner-agent interaction

5 Evaluation

Setup and sample. To evaluate whether our concept can solve the issues presented in Section 4.1, the FEDS framework of Venable et al. [41] led us to a human risk and effectiveness strategy. We conducted a formative and artificial evaluation of the initial instantiation of our design knowledge to evaluate the utility of our concept for teachers and schoolchildren and identify refinements. We focused on comparing our system to the use of two-dimensional e-learning resources (i.e., web-based apps such as Anton [56] and mobile apps such as SchulAR [48]) that are tailored to the curricula and serve to promote active learning. We interviewed professionals with an educational background, as they a) are responsible for providing didactic guidance to schoolchildren, b) have extensive experience with the administration of digital learning resources, and c) constitute a target user group of our artifact.

In total, we conducted 11 semi-structured interviews, including seven secondary school teachers, and four graduate researchers from the field of biology didactics. Eight participants were female, and three were male. Prior to the interviews, respondents received an explanation of the artifact and completed two tasks. First, we asked them to imagine that they had previously attended an online lecture on human organs and were required to complete homework on this topic. As the participants had failed to understand the functions and structure of the liver and spleen, the respondents should ask the agent three exemplary questions concerning the functions and structure of these organs. Upon completion, we proceeded to the interviews, guided by the following three questions: 1) "What are the strengths and weaknesses of the concept?", 2) "What are the benefits of the concept compared to existing apps?", 3) "How can the concept be improved?" The interviews took place in July 2021 and lasted 42 minutes on average. We conducted a qualitative content analysis of the transcripts [57] by applying open coding regarding the perceived benefits and refinements of the concept. Table 1 summarizes the perceived benefits of employing the artifact in distance learning and provides representative interview statements.

Table 1. Perceived benefits of ExplainAR compared to existing solutions.

DP	Code Cluster	Quotes (Examples)
1	Learner-centered communication (1)	*"You have to test existing apps yourself before each lesson, because every teacher uses different, partly fairly complex, terminology and existing apps do not allow for customization of such content. Your concept allows me to tailor the delivery of the learning content to the capabilities of the target audience."* *"I believe that the voice assistant is very important because it allows for a learner-centered explanation of what is being learned."*
	Improved control of learning objectives (2)	*"The customization is one of the most important features as it allows teachers to maintain control of the learning objectives, which is often not the case with existing solutions."*
	Unburdening teachers and parents (3)	*"I would use the artifact because I can manage the learning content myself, relieving the lessons of questions that would otherwise be asked three times in class as the children can work at their own pace and query the CA."*
2	Support of disseminating complex phenomena (4)	*"In natural science, there are highly condensed technical texts where one technical term is explained with the next technical term, which is often too complex for schoolchildren. That's why multimodal learning approaches are so important to clarify the text … and your bot even brings the auditory channel to the book."* *"The 3D visualization allows disseminating complicated, dangerous, or otherwise difficult to demonstrate phenomena such as the anatomy of the eye."*
	Increased learner Engagement (5)	*"By additionally animating the learning content, schoolchildren certainly can focus much better because something is happening."*
	Enhanced usability (6)	*"We experienced that the technology is usually the first thing to go wrong. Some students can't even handle the keyboard. The QR code appears to me to be quite nice … as the app opens directly and the child can simply access the information by voice."*
3	High device compatibility (7)	*"Compared to existing solutions like MergeCube, the present concept features the advantage of not requiring additional and dedicated hardware."*

Perceived benefits. All the teachers rated the concept as very useful and agreed that they could imagine supplementing their distance learning routines with ExplainAR. They appreciated the idea of content reduction and delivering the information through the multimodal CA by employing auditory and visual stimuli. Compared to existing solutions such as Anton [56] and SchulAR [48], respondents highlighted seven benefits: 1) The content administration allows teachers to tailor the communication of the learning content to the prior knowledge and cognitive abilities of the schoolchildren (e.g., children with learning difficulties). 2) Teachers can define learning objectives, as they can tailor the content to their courses. 3) Enabled by content self-management, repetitive questions can be anticipated, reducing the supervision efforts of teachers and parents. 4) 3D visualization via AR facilitates the presentation of subjects that are difficult to present on paper or in 2D (e.g., spatial relationships). 5) Interactive visualization via AR and short auditory stimuli provided by the CA increase learner engagement. 6) The interaction via speech allows bypassing digital illiteracy of learners. 7) Implementing the multimodal CA as a web application ensures compatibility with a wide range of devices.

Refinements. Nevertheless, the teachers emphasized five refinements. To further leverage the potential of the weeding strategy, two participants recommended using enhanced animated 3D models to visualize sequences in 3D. For instance, one didactic expert mentioned that this enhancement would allow to disseminate more complex phenomena and increase learners' attention: "*With animated 3D sequences, we literally* *could show chemical reactions to students without them needing a lab at home*." 2) Interestingly, nine respondents suggested implementing a feature that alerts teachers to the complexity and length of texts created in the web portal to avoid cognitive overstimulation in the auditory channel. For instance, one teacher stated that "*schoolchildren stop listening to teachers after three sentences when learning new facts."* 3) To extend the application range of the artifact (e.g., to foreign language teaching), seven teachers proposed to enable multi-language interaction for migrant schoolchildren. 4) Moreover, the interviews revealed the necessity to annotate individual components of a 3D model (e.g., liver spatially represented in front of the human body) while the CA explains the respective learning content. The respondents mentioned that this extension would be valuable for foreign language teaching "*to improve pronunciation and spelling skills*." 5) Furthermore, two teachers suggested implementing a handover recovery strategy that forwards the schoolchildren's questions to the teacher if the CA does not understand the user's intent. This forwarding enables teachers to identify and discuss frequently asked questions during online sessions. Next, these refinements will be integrated before evaluating the learning effectiveness of the artifact with two classes.

6 Discussion, Limitations, and Future Research

As a result of the pandemic, home schooling has become the "new normal" for many schoolchildren, which presents challenges for learning success, especially for children of non-academics or schoolchildren who suffer from a lack of parental support and low digital literacy [3]. Informed by Mayer's theory of multimedia learning [34], this research argues that multimodal CAs can resolve these issues by providing short verbal explanations and interactive visualizations of complex phenomena. In this paper, we developed three DPs and instantiated them by implementing an IT artifact. The results of the proof-of-concept evaluation with 11 teachers and researchers confirmed the utility of our design knowledge. ExplainAR was framed as a beneficial supplement to support schoolchildren and relieve teachers in distance learning settings. Furthermore, the evaluation findings for each DP enabled us to revise our design considerations.

Main implications and contributions. Given the short-term demand for digital education opportunities, our research is expected to be valuable for practitioners and researchers for several reasons. First, our research informs policymakers, teachers, and parents about the potential of multimodal CAs to facilitate distance learning. The interviewed teachers showed a high level of interest in our artifact, indicating that they had previously lacked such multimodal solutions. Thus, our results suggest that the combination of CAs and AR is a promising addition to traditional e-learning offerings or an alternative to video conferencing to avoid so-called "zoombies" [45]. For example, educators are currently experimenting with measures such as recording short video lectures that explain a complex issue [21]. Our artifact adopts this concept of didactic reduction but simplifies the self-creation of multimedia content. Second, our findings inform software developers and providers to exploit recent advancements in the realms of CAs and AR. While analog learning materials have already been complimented by AR, to the best of our knowledge, no existing applications combine CAs with AR in a pedagogical context. As such, the design knowledge may guide future implementations in the realms of pedagogical CAs and AR-based learning applications. For instance, our study should encourage providers of applications such as SchulAR [48] to improve their applications by integrating content management capabilities to enable more effective learning scenarios for schoolchildren and teachers.

Apart from its practical relevance, this study provides threefold scientific contributions. First, we contribute three preliminary DPs and the instantiation itself to the DSR knowledge base for multimodal CAs. According to Gregor and Hevner [58], the artifact itself encompasses a level 1 contribution (i.e., a situated implementation) that reflects preliminary design knowledge. To advance the maturity of the DPs, empirically verify the problem-solving ability of the artifact, and ultimately contribute a level 2 contribution in terms of a nascent design theory, we intend to evaluate the artifact as part of an experimental study in the next evaluation. Thus, our research extends prior works by scholars such as Hobert [11], who developed DPs for pedagogical CAs. In our project, DP1 and DP2 should be emphasized, as they enable the extension of speech-based interaction in educational settings through multimodal output based on AR technologies. The preliminary evaluation indicates that this multimodality improves on the shortcomings of purely visualizing technologies, such as the representation dilemma [29], and of purely text-based robots, such as the challenge of comprehending more complex concepts [32]. Therefore, we respond to recent calls by the IS community to leverage its key competencies in designing IS to mitigate the severe consequences of the pandemic [59] and to determine how technology can enhance specific learning activities [19], [24]. Second, our study contributes to the emerging research field in the area of pedagogical CAs by introducing a multimodal CA interface. While multimodal pedagogical CAs already exist, their visual aspects focus rather to the embodiment of CAs [12]. Our CA extends this definition of multimodality by focusing on 3D visualization of learning materials directly over the worksheet and textbooks, complemented by audio-based output of teacher-prepared texts. Third, we contribute to the literature on the cognitive theory of multimedia learning by implementing the weeding strategy as an effective didactical concept for personalized e-learning scenarios [34]. To the best of our knowledge, this is the first research to implement this concept on a multimodal CA.

Limitations and future research. Our research has several limitations that suggest avenues for future research. First, the literature review we used to derive the issues and meta-requirements drew on a limited sample. Future research should include interviews with stakeholders such as schoolchildren and parents to enrich the design knowledge further to provide a practical perspective on the requirements. Second, thus far, we have only evaluated an initial instantiation of the DPs and exclusively interviewed German participants. Due to COVID-19 and the ongoing restrictions, the artifact has not been tested with the target user group (i.e., schoolchildren) in a real-life setting. Therefore, we plan to evaluate the artifact in terms of its effectiveness and cognitive load in a field experiment with two classes in the future. Since teachers are the second major target group of our artifact besides students, the content management system will need to be evaluated with a larger sample in the future as well.

References

1. UNICEF: COVID-19 and School Closures: One Year of Education Disruption. (2021)
2. OECD: Strengthening online learning when schools are closed - The role of families and teachers in supporting students during the COVID-19 crisis. https://read.oecd-ilibrary.org/view/?ref=136_136615-o13x4bkowa&title=Strengthening-online-learning-when-schools-are-closed (Accessed: 22.08.2021)
3. Wößmann, L., Freundl, V., Grewenig, E., Lergetporer, P., Werner, K., Zierow, L.: Bildung in der Coronakrise: Wie haben die Schulkinder die Zeit der Schulschließungen verbracht, und welche Bildungsmaßnahmen befürworten die Deutschen?, ifo Schnelld. 73(9), 25–39 (2020)
4. Mayer, R. E., Moreno, R.: Nine ways to reduce cognitive load in multimedia learning. Educ. Psychol. 38(1), 43–52 (2003)
5. Sweller, J., Chandler, P.: Evidence for Cognitive Load Theory. Cogn. Instr. 8(4), 351–362 (1991)
6. Mayer, R. E.: Multimedia learning: Are we asking the right questions?. Educ. Psychol. 32 (1), 1–19 (1997)
7. Hobert, S., Meyer von Wolff, R.: Say Hello to Your New Automated Tutor – A Structured Literature Review on Pedagogical Conversational Agents, In: 14. Internationale Tagung Wirtschaftsinformatik, Siegen, pp. 301–314 (2019)
8. Winkler, R., Söllner, M., Leimeister, J.M.: Enhancing problem-solving skills with smart personal assistant technolog. Comput. Educ. 165 (2021)
9. Sommerauer, P., Müller, O.: Augmented reality for teaching and learning - A literature review on theoretical and empirical foundations, In: Twenty-Sixth European Conference on Information Systems, Portsmouth (2018)
10. Harborth, D.: Augmented Reality in Information Systems Research: A Systematic Literature Review, In: Twenty-third Americas Conference on Information Systems, Boston (2017)
11. Hobert, S.: Say Hello to 'Coding Tutor'! Design and Evaluation of a Chatbot-based Learning System Supporting Students to Learn to Programming Instructions, In: International Conference on Information Systems (ICIS), Munich (2019)
12. Weber, F., Wambsganss, T., Rüttimann, D., Söllner, M.: Pedagogical Agents for Interactive Learning: A Taxonomy of Conversational Agents in Education. In: Forty-Second International Conference on Information Systems, Austin (2021)
13. Sommerauer, P., Müller, O.: Augmented reality in informal learning environments: A field experiment in a mathematics exhibition. Comput. Educ. 97, 59–68 (2014)
14. Gopalan, V., Bakar, J.A.A., Zulkifli, A.N.: A brief review of augmented reality science learning, In: AIP Conference Proceedings (2017)
15. Diegmann, P., Schmidt-Kraepelin, M., Eynden, S., Basten, D.: Benefits of Augmented Reality in Educational Environments - A Systematic Literature Review. In: Proceedings der 12. Internationalen Tagung Wirtschaftsinformatik (WI 2015), pp. 1542–1556 (2015)
16. Hevner, A. R., March, S. T., Park, J., Ram, S.: Design science in information systems research. MIS Q 28(1), 75–105 (2004)
17. Ouh, E.L, Gan, B.K.S.: Effective Digital Learning Practices for IS Design Courses during COVID-19. In: Twenty-Seventh Americas Conference on Information Systems (2021)
18. Rodriguez-Abitia, G.: Coping with COVID-19 in Mexico: Actions for Educational Inclusion. Commun. Assoc. Inf. Syst. (48) 93–101 (2021)
19. Rai, R.S., Selnes, F.: Conceptualizing task-technology fit and the effect on adoption – A case study of a digital textbook service. Inf. Manag. 56(8) (2019)

20. Zha, S., He, W.: Pandemic Pedagogy in Online Hands-on Learning for IT/IS Courses. Commun. Assoc. Inf. Syst. 48, 80–87 (2021)
21. Mavengere, N., Henriksen-Bulmer, J., Passmore, D., Mayes, H., Fakorede, O., Coles, M., & Atfield-Cutts, S.: Applying Innovative Technologies and Practices in the Rapid Shift to Remote Learning. Commun. Assoc. Inf. Syst. 48(1), 185–195 (2021)
22. Gan, I., Sun, R.: Digital Divide and Digital Barriers in Distance Education during COVID-19. In: Proceedings of the 54th Hawaii International Conference on System Sciences (2021)
23. UNICEF: Guidance on Distance Learning Modalities to Reach All Children and Youth During School Closures (2020)
24. Huang, R., Tlili, A., Wang, H., Shi, Y., Bonk, C. J., Yang, J., & Burgos, D.: Emergence of the Online-Merge-Offline (OMO) Learning Wave in the Post-COVID-19 Era: A Pilot Study. Sustainability 13(6) (2021)
25. Milgram, P., Kishino, F.: Taxonomy of mixed reality visual displays. IEICE Trans. Inf. Syst. 77(12), 1321–1329 (1994)
26. Azuma, R. T.: A survey of augmented reality. Presence Teleoperators Vir Env 6(4), 355–385 (1997)
27. Ibáñez, M.B., Di Serio, Á., Villarán, D., & Kloos, C.D.: Experimenting with electromagnetism using augmented reality: Impact on flow student experience and educational effectiveness. Comput. Educ. 71, 1–13 (2014)
28. Di Serio, Á., Ibáñez, M.B., & Kloos, C.D.: Impact of an augmented reality system on students' motivation for a visual art course. Comput. Educ. 68, 586–596 (2013)
29. Keller, S., Rumann, S., Habig, S.: Cognitive Load Implications for Augmented Reality Supported Chemistry Learning. Information 12(3) (2021)
30. Laumer, S., Maier, C., & Gubler, F. T.: Chatbot acceptance in healthcare: Explaining user adoption of conversational agents for disease diagnosis. In: 27th European Conference on Information Systems, pp. 1–18 (2019)
31. Anton, E., Oesterreich, T. D., Schuir, J., Teuteberg, F.: Painting A Holistic Picture of Trust in and Adoption of Conversational Agents: A Meta-Analytic Structural Equation Modeling Approach. In: Proceedings of the 55th Hawaii International Conference on System Sciences (2022)
32. Hobert, S.: How are you, chatbot? Evaluating chatbots in educational settings - Results of a literature review. In: Pinkwart, N., Konert, J, DELFI 2019. pp. 259–270 (2019)
33. Nunamaker, J. F., Derrick, D. C., Elkins, A. C., Burgoon, J. K., & Patton, M. W.: Embodied conversational agent-based kiosk for automated interviewing. J. Manag. Inf. Syst. 28(1), 17–48 (2011)
34. Mayer, R.E.: Cognitive Theory of Multimedia Learning, In: The Cambridge Handbook of Multimedia Learning, R. Mayer, Ed. Cambridge: Cambridge University Press, pp. 43–71 (2010)
35. Peffers, K., Tuunanen, T., Rothenberger, M. A., Chatterjee, S.: A design science research methodology for information systems research. J. Manag. Inf. Syst. 24(3), 45–77 (2007)
36. Vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., Cleven, A.: Reconstructing the giant: On the importance of rigour in documenting the literature search process. In: 17th European Conference on Information Systems, Verona (2009)
37. Gregor, S., Kruse, L. C., Seidel, S.: The Anatomy of a Design Principle. J. Assoc. Inf. Syst. 21(6) (2020)
38. Webster, J., Watson, R.T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. MIS Q 26(2), xiii–xxiii (2002)
39. Morgan, D. L.: The focus group guidebook, vol. 1. Sage publications (1997)

40. March, S. T., Smith, G. F.: Design and natural science research on information technology. Decis. Support Syst. 15(4) 251–266 (1995)
41. Venable, J., Pries-Heje, J., Baskerville, R.: FEDS: A Framework for Evaluation in Design Science Research. Eur. J. Inf. Syst. 25(1), 77–89 (2016)
42. Mavengere, N., Henriksen-Bulmer, J., Passmore, D., Mayes, H., Fakorede, O., Coles, M., Atfield-Cutts, S: Applying innovative technologies and practices in the rapid shift to remote learnin., Commun. Assoc. Inf. Syst. 48(1), 185–195 (2021)
43. Keller, A.: In Print or On Screen? Investigating the Reading Habits of Undergraduate Students Using Photo-Diaries and Photo-Interviews, Libri 62(1), 1–18 (2012)
44. Olsen, A. N., Kleivset, B., Langseth, H.: E-Book Readers in Higher Education. SAGE Open 3(2) (2013)
45. Toney, S., Light, J., Urbaczewski, A.: Fighting Zoom Fatigue: Keeping the Zoombies at Bay. Commun. Assoc. Inf. Syst. 48, 40–46 (2021)
46. Schober, B., Marko, L., Spiel, C.: Lernen unter COVID-19-Bedingungen: Erste Ergebnisse, Univesität Wien (2020)
47. Graesser, A., McNamara, D.: Self-regulated learning in learning environments with pedagogical agents that interact in natural language. Educ. Psychol. 45(4), 234–244 (2010)
48. KIDS interactive GmbH, SchulAR. https://www.kids-interactive.de/ (Accessed: Aug. 10, 2021).
49. Xu, Y., Warschauer, M.: Young children's reading and learning with conversational agents. In: Conf. Hum. Factors Comput. Syst. - Proc., pp. 1–8 (2019)
50. Diederich, S., Brendel, A.B., Kolbe, L.M.: Designing Anthropomorphic Enterprise Conversational Agents. Bus. Inf. Syst. Eng., 62(3), 193–209 (2019)
51. Kamphuis, C., Barsom, E., Schijven, M., & Christoph, N.: Augmented reality in medical education?. Perspect. Med. Educ., 3(4), 300–311 (2014)
52. Mozilla Developers, Web Speech API - Web APIs | MDN, MDN Web Docs (2021)
53. Wit.ai Inc: wit.ai https://wit.ai/. (Accessed: 10.08.2021)
54. Etienne, J.: AR.js Documentation, https://ar-js-org.github.io/AR.js-Docs/. (Accessed: 10.08.2021)
55. Benner, D., Elshan, E., Schöbel, S., Janson, A.: What do you mean? A Review on Recovery Strategies to Overcome Conversational Breakdowns of Conversational Agents. In: Forty-Second International Conference on Information Systems, Austin (2021)
56. Solocode GmbH, Anton. https://anton.app/en_us/ (Accessed: 10.08.2021)
57. Mayring, P.: Qualitative Content Analysis. Forum: Qualitative Social Research, 1 (2), (2000)
58. Gregor, S., Hevner, A. R.: Positioning and presenting design science research for maximum impact. MIS Q 37(2) (2013)
59. Thomas, O., Hagen, S., Frank, U., Recker, J., Wessel, L., Kammler, F., Zarvic, N., Timm, I.: Global Crises and the Role of BISE. Bus. Inf. Syst. Eng. 62(4), 385–396 (2020)