



A collaborative AR application for education: from architecture design to user evaluation

Stefano Masneri¹ · Ana Domínguez² · Guillermo Pacho² · Mikel Zorrilla² · Mikel Larrañaga¹ · Ana Arruarte¹

Received: 28 June 2023 / Accepted: 3 January 2024 / Published online: 13 February 2024
© The Author(s) 2024

Abstract

Augmented reality applications can be used in an educational context to facilitate learning. In particular, augmented reality has been successfully used as a tool to boost students' engagement and to improve their understanding of complex topics. Despite this, augmented reality usage is still not common in schools and it still offers mostly individual experiences, lacking collaboration capabilities which are of paramount importance in a learning environment. This work presents an application called *ARoundTheWorld*, a multiplatform augmented reality application for education. It is based on a software architecture, designed with the help of secondary school teachers, that provides interoperability, multi-user support, integration with learning management systems and data analytics capabilities, thus simplifying the development of collaborative augmented reality learning experiences. The application has been tested by 44 students and 3 teachers from 3 different educational institutions to evaluate the usability as well as the impact of collaboration functionalities in the students' engagement. Qualitative and quantitative results show that the application fulfils all the design objectives identified by teachers as key elements for augmented reality educational applications. Furthermore, the application was positively evaluated by the students and it succeeded in promoting collaborative behaviour. These results show that *ARoundTheWorld*, and other applications built using the same architecture, could be easily developed and successfully integrated into existing schools curricula.

Keywords Augmented reality · Multi-user interactions · Collaborative learning · User evaluation · Data analysis

1 Introduction

Augmented reality (AR) is a technology which overlays virtual objects (augmented components) into the real world. These virtual objects then appear to coexist in the same space as objects in the real world (Akçayır and Akçayır 2017; Azuma et al. 2001). Another widely used definition is from Milgram and Kishino (1994) where, in a continuum ranging from a purely virtual environment to a completely real one, AR is positioned close to the real environment and the users perceive the real world with an additional layer of virtuality. Even though the technology was originally used as a tool to help assembly workers at Boeing by showing them virtual labels through the use of custom made glasses (Caudell and Mizell 1992), AR has quickly attracted research and industry attention in many different areas such as gaming (Das et al. 2017), cultural heritage (Vlachos et al. 2022), customer engagement (McLean and Wilson 2019), manufacturing (Ong et al. 2008) or education (Avila-Garzon et al. 2021; Garzón et al. 2019). Especially, in the last ten years, there has been a surge of research related to the usage

✉ Stefano Masneri
smasneri001@ikasle.ehu.eus

Ana Domínguez
adominguez@vicomtech.org

Guillermo Pacho
gpacho@vicomtech.org

Mikel Zorrilla
mzorrilla@vicomtech.org

Mikel Larrañaga
mikel.larranaga@ehu.eus

Ana Arruarte
a.arruarte@ehu.eus

¹ Computer Languages and Systems Department, University of the Basque Country UPV/EHU, San Sebastian, Spain

² Fundación Vicomtech, Basque Research and Technology Alliance (BRTA), San Sebastian, Spain

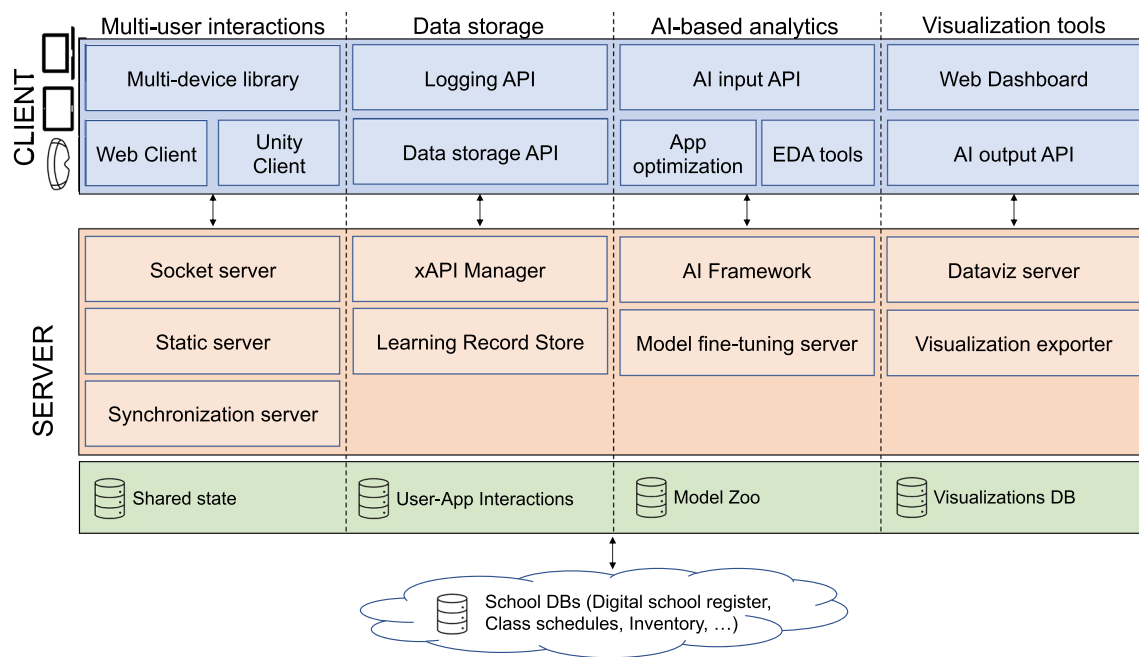


Fig. 1 The *cleAR* architecture as defined in Masneri et al. (2023)

of AR in education and its impact on student engagement and academic results (Bressler and Bodzin 2013; Sirakaya and Cakmak 2018; Chang et al. 2022). This can largely be attributed to the increased availability of AR-ready devices as well as the familiarity of today's students with technology-enhanced learning (TEL).

Despite huge improvements in both hardware and software which led to an increased offer of AR applications for mobile devices and AR headsets, as well as extensive research on the usage of AR in education (Dinis et al. 2017; Akçayır and Akçayır 2017; Chen et al. 2017; Ibáñez and Delgado-Kloos 2018; Pellas et al. 2019; Masneri et al. 2022b), the usage of AR in primary and secondary schools is still not common (Commission et al. 2023). In a previous work (Masneri et al. 2023), the main reasons behind this were identified, namely the limited collaboration capabilities of existing apps, the inability to create new content and the difficulty of adapting to existing school curricula. In the same work, the *cleAR* (see Fig. 1) architecture was presented. This is an interoperable architecture that enables the creation of multi-user AR applications, simplifies the development process and allows the stakeholders to add new content to existing applications, track user progress and integrate application data into the learning management system (LMS) used by the teachers.

In Masneri et al. (2023), 47 teachers were asked to fill in a questionnaire to help identify the requirements of the architecture and were also involved in several interviews to define the scope of the solutions. Ultimately, four design objectives (DOs) were identified:

- **Interoperability (DO1):** The architecture enables the creation of applications that can run on multiple devices—tablets, smartphones, head-mounted displays (HMDs) or laptops—and provide APIs for development on multiple platforms.
- **Multi-user capabilities (DO2):** As collaboration is a key requirement, the architecture provides tools to support multi-user functionalities, both for remote and in-class collaboration.
- **Data analytics (DO3):** The architecture enables long-term data storage as well as tools for automatic data analysis and visualization, by providing an API to access standard dataviz and machine learning libraries.
- **Easy to develop (DO4):** The applications relying on the architecture can be developed quickly and easily.

In the previous work, three proof-of-concepts (PoCs) were developed with the aim of performing a conceptual evaluation of the architecture and to demonstrate that it fulfils the aforementioned DOs.

In this new manuscript, we build upon such previous work and introduce *ARoundTheWorld*, a multiplatform collaborative AR geography game. The primary goal of the application is to demonstrate the potential of the *cleAR* architecture for creating interoperable applications that can be integrated into school curricula.¹ Additionally, the game

¹ The application is open source and the code can be accessed at this [repository](#).

aims to enhance student engagement through the collaborative functionalities provided by *cleAR*. The application has been developed incorporating feedback from the teachers of a Basque primary and secondary school association² and it has been evaluated after being tested with 44 students. Once the test was complete, teachers were interviewed while students were asked to fill a short questionnaire about the *ARoundTheWorld* User Interface (UI), the User eXperience (UX) it offered, as well as its effectiveness as a tool for raising the engagement of the students and enabling collaboration between them. To perform a quantitative evaluation about *ARoundTheWorld* collaboration capabilities, the app collected data—in the form of xAPI statements (Clarke et al. 2020)—about its usage, the number of interactions between students and the performance of the students.

The choice of Geography as the application domain is motivated by the fact that geographical exploration is an integral part of child development (Catling 1993), and the use of maps helps students improve spatial thinking skills (Collins 2018). The application is structured as a quiz where students take turns to answer geography questions. If a student is struggling to answer a question, other students can interact in the augmented space and provide hints to the active user. *ARoundTheWorld* works both as a mobile and a web application, is easily extensible and provides several logging and tracking mechanisms, which can be easily integrated into the school's LMS to enable automatic tracking of the progress of the students.

The main contributions of this work can be summarized as:

- A complete description of the application and the feasibility of implementing the different components of the *cleAR* architecture to fulfil all the design objectives.
- A qualitative evaluation of the technology integrated in the application, based on the questionnaires filled in by the 44 students and the interviews with their teachers.
- An analysis of the data collected by the application during the user study, with a focus on the effects of collaboration capabilities on the quiz results and the engagement of the students.

The number of users who participated in the study is not representative of the whole student population. The aim of this study is not to demonstrate the positive effects of AR in school settings, but rather to validate the architecture presented in a real school setting, and not only through the development of PoCs.

The rest of the paper is structured as follows: Sect. 2 covers the relevant state of the art regarding collaborative AR

applications for education. Section 3 describes the implementation details of *ARoundTheWorld*, while Sect. 4 outlines our methodological framework and the evaluation process. In Sect. 5, we describe the results obtained from the student questionnaires and teacher interviews and the quantitative analysis of the data collected through the application. Finally, Sect. 6 presents the conclusions and suggests future research lines.

2 Related work

While the first mention of collaborative AR is more than 20 years old (Billinghurst and Kato 2002), there are only a few works presenting such applications in educational contexts, mainly due to the difficulty of including multi-user capabilities in AR apps. A systematic review of AR applications used in education (Iqbal et al. 2022) mentions that collaborative learning in AR represents a critical research direction, but so far very few studies provide collaborative functionalities in an AR environment (Pan et al. 2021; Choi et al. 2017). The work from Cai et al. (2017) presents an application that makes use of a Kinect camera to extract 3D information from the scene and display virtual magnetic induction lines. As the students move around the room the objects simulating the magnets, the system updates in real time the representation of the magnetic field. In Takahashi et al. (2018), the authors designed a large scale AR and projection system, modifying the gymnasium of the school, to create a learning game for children with autism spectrum disorder (ASD), which was designed to keep their attention focused on the content provided. Laviolo et al. (2018) presented a markerless application for learning how an artificial neural network works, where the students can manually tweak the values of the network parameters and see how it affects the ability of the network to classify images.

As it heavily relies on visual representation of data, several technologies have been exploited to make the teaching of geography more effective and engaging. In the context of AR, Palaigeorgiou et al. (2018) used a projector to create tangible 3D maps with which up to three students could interact at the same time. Xefteris and Palaigeorgiou (2019) extended the concept of tangible maps by including the usage of programmable robots to guide the students through a virtual journey.

As far as evaluating the effectiveness of AR applications for education, the vast majority of the studies highlight a positive (albeit limited) effect derived from using the technology. Chang et al. (2022) performed a meta-analysis of 134 studies which suggests that AR benefits all the learning outcomes evaluated, with the largest effect being on students performance. A systematic review of 45 studies (da Silva et al. 2019) reaches the same conclusions, but highlights the

² ikastola.eus.

many differences in the evaluation protocols, which complicate the statistical analysis of AR effectiveness across different applications.

AR applications are often implemented as serious games, in which using gamification concepts the students can more easily learn and retain concepts that would otherwise not interest them. Oh et al. (2017) described a game-based simulation where the users can study the properties of light such as reflection and refraction. López-Faican and Jaen (2020) created a multiplayer game in which children can improve their communication skills by practising in an AR environment, while Çelik and Yangın Ersanlı (2022) described a gamified AR app used in a Content and Language Integrated class.

Several publications focus on the importance of effective UIs and UXs in enhancing student engagement. A systematic review of the literature analysed 49 studies (Law and Heintz 2021) and identified a lack of knowledge about usability and user experience frameworks, suggesting that there is a disconnect between Human–Computer Interaction (HCI) and TEL communities, as well as a lack of AR-specific UX evaluation metrics. The work of Thamrongrat and Law (2019) evaluated the learning effect for teaching 3D geometry using an AR application compared to traditional learning, as well as the User Engagement (UE) of the students using the app compared to the ones in the control group. Another study (Alrashidi et al. 2017) compared the effectiveness of learning software debugging concepts using an AR application versus a non-AR approach.

Applications used in schools usually generate data that are stored on the school LMS. A standard that is recently gaining traction for collecting data about learners' activities is eXperience API (xAPI),³ an open-source software specification that makes it possible to collect data about a wide range of learning experiences. This is achieved by sending each activity that needs to be recorded to a learning record store (LRS) in a consistent and secure format (Clarke et al. 2020). The activities are collected as statements stored as JSON objects. Statements can be tuned to a specific use case by defining a vocabulary of valid statements. Secretan et al. (2019) described a system where xAPI is used to perform learning analytics in an AR environment, while Wu et al. (2020) used xAPI to collect data for a 3D design course.

Despite the impressive amount of literature about the use of AR technology in educational environments, to the best of our knowledge, there is no solution that incorporates all the requirements identified by the teachers. These requirements aim to make AR apps a useful learning tool by providing collaboration capabilities, customisation, interoperability, analytics, LMS integration and ease of use.

3 Collaborative AR application

In this section, we summarize the application implemented using *cleAR*, the architecture presented in Masneri et al. (2023) and how the developed application fulfils the DOs presented in Sect. 1. The application, called *ARoundTheWorld*, is a collaborative geography quiz in which students answer a set of questions prepared by the teacher. Once started, the application sends a question to the first student (for example, “Where is Kyoto?”). The student answers by placing a pin on the 3D globe of the Earth shown in the augmented space. Other students can collaborate with the active user in two ways. They can suggest to her in which continent the answer is located and—once the user has placed the pin but has not confirmed her choice yet—by sending a “thumbs up” or “thumbs down” feedback.⁴ Once a student answers, the application sends a new question to the next user and repeats the process until all the questions have been answered. Figure 2 shows the application workflow, highlighting the interactions of teacher and students.

The application considers three types of users (“teacher”, “players” and “watchers”), depending on their role and their means of interacting with other users. The first role is that of the students participating in the quiz—the *player*—described above. Another role is that of the *teacher*, who controls the overall status of the app through a web-based interface. The final role—the *watcher*—is that of the students who are not actively participating in the quiz (that is, they are not answering any questions). They can watch what other students are doing and suggest to them the correct answer. This role was designed to let students without an AR capable device engage with the players by checking what they are doing and collaborate with them by suggesting the correct answer.

The application is designed to require minimal supervision from the teacher to let him or her interact as much as possible with the students. As shown in Fig. 3, the teacher interface consists of four parts:

- A 3D representation of the augmented content as viewed by the active user (that is, the student who is answering the current question).
- The list of users connected to the app, together with the current score of the players and the last question they answered.
- The suggestions sent to the student currently answering the question.
- A dashboard (accessible in a separate tab) with charts of the scores achieved by each student across different sets of questions.

³ <https://xapi.com/overview/>.

⁴ A video description of the application is available [here](#).

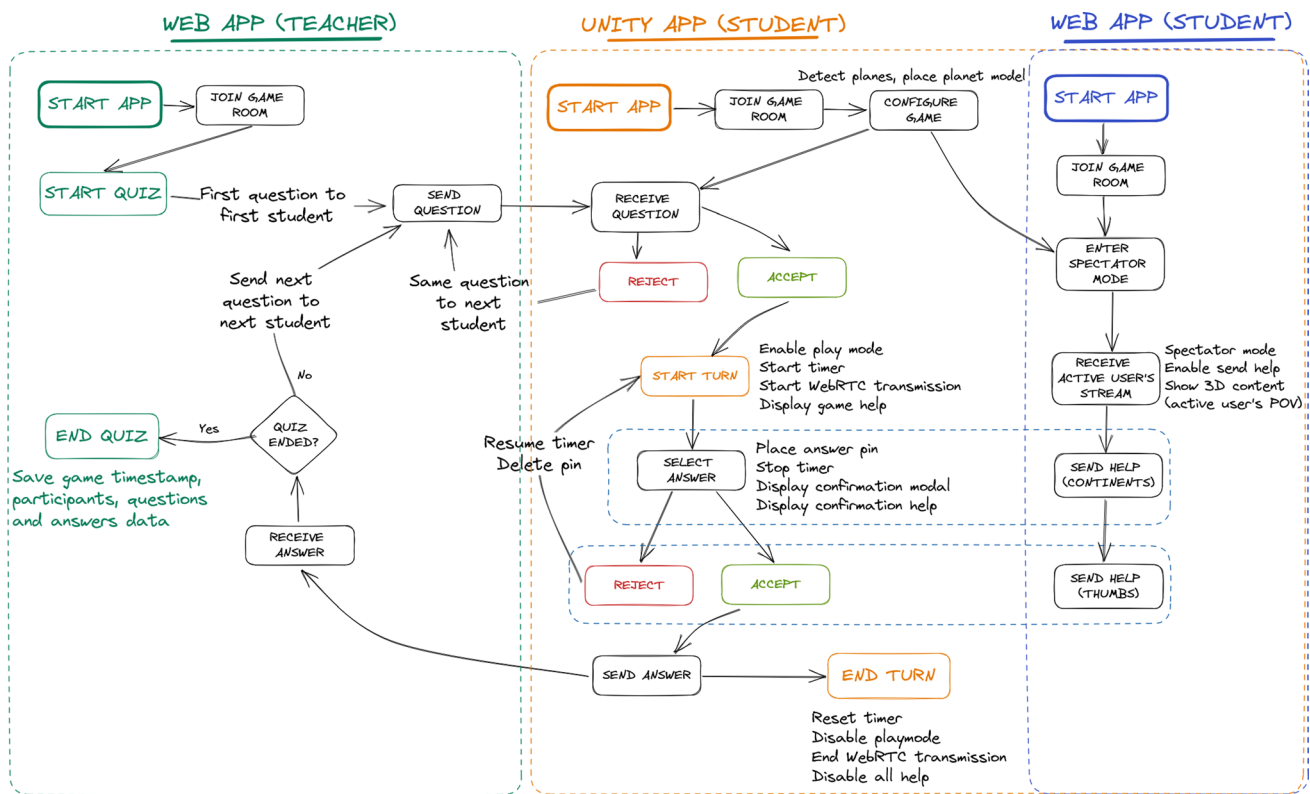


Fig. 2 Workflow of the ARoundTheWorld application

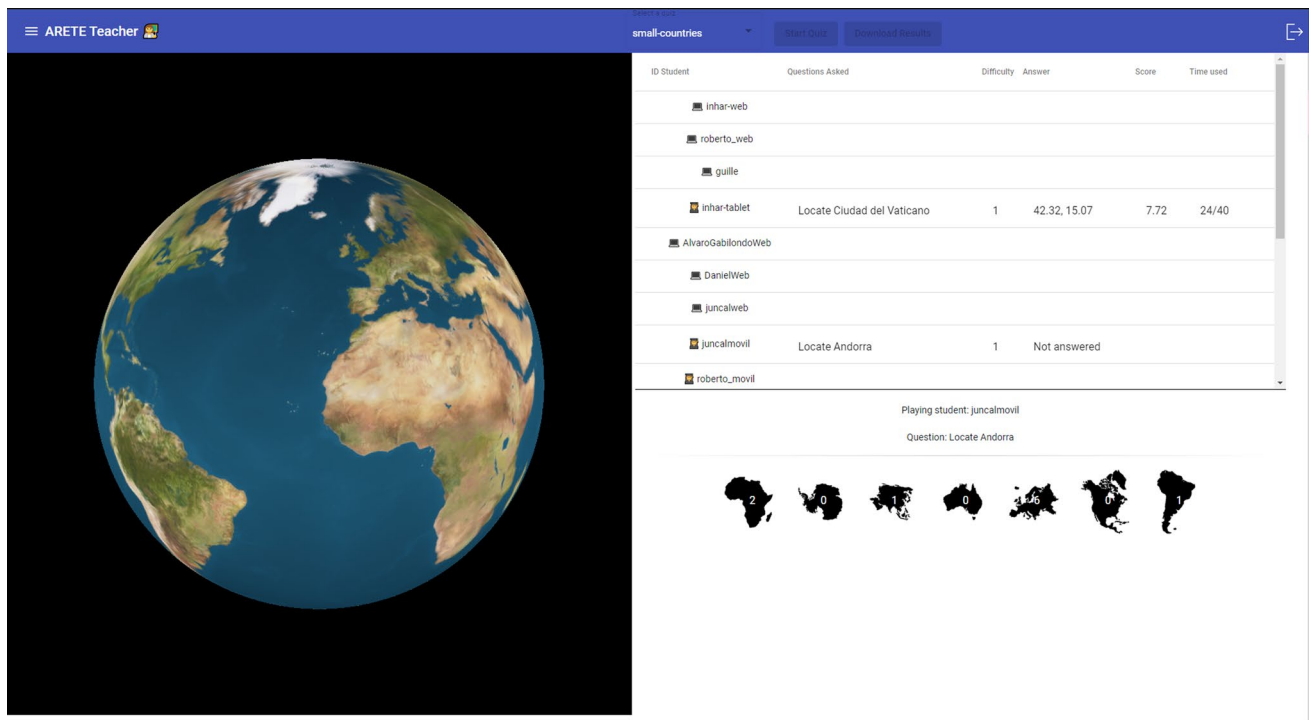


Fig. 3 The web-based teacher interface of the ARoundTheWorld application

The teachers who filled in the questionnaire described in Masneri et al. (2023) mentioned that one of the factors limiting the usage of AR apps in schools is the lack of customisation capabilities. In this respect, *ARoundTheWorld* provides an additional web interface from which the teacher can create new sets of questions. To minimize the amount of work required by the teacher, the coordinates of each location are computed automatically using the Wikimedia API⁵ and the questions are stored as JSON files which are directly added to the application. The interface of the watchers is web based, too, and has a look and feel similar to the teacher interface.

For the application to successfully achieve interoperability (DO1), several types of hardware as well as software libraries need to be supported. In the aforementioned survey, the teachers reported a wide spectrum of devices available in their schools. Chromebooks and Android tablets were the most commonly used but other options included laptops, PCs, smartphones (both Android and iOS based) and iPads. Furthermore, while none of the teachers reported using AR headsets such as HoloLens, we believe that such devices provide the best AR learning experience for users. For this reason, our application supports Microsoft Mixed Reality Toolkit and is fully compatible with HoloLens devices. The application for mobile and tablet devices has been developed using Unity 2020.3, and the AR functionalities are provided by the AR foundation framework. The web application has been built using Typescript and Three.js (to enable 3D content to be displayed in the browser). All the logging data and the statements collected during app usage are stored in a Mongo database in the Learning Locker instance deployed on AWS. Porting to HoloLens and iOS devices is achieved through, respectively, Unity integration with Microsoft MR Toolkit and by exporting Unity's project file to an XCode environment.

The application supports multi-user capabilities (DO2) by relying on the functionalities provided by the *cleAR* architecture, which provides a library for sharing 1-to-1 or broadcast message passing (Masneri et al. 2022a) with minimal changes to the existing code base. When a student is asked to answer a question, she becomes the active user. She shares the camera position (which determines her view of the 3D globe) as well as the position of the pin, once placed, with the other users. The other students will then see on their devices the 3D globe in the same way the active user does. For users on a mobile device, this happens directly in the augmented space, while users using a PC will see the globe in a virtual 3D environment on a `<canvas>` element. At the same time, suggestions from users are shared in a broadcast fashion, so that every student knows about the suggestions

sent by others. Finally, the teacher interface shares information about the current question, the score obtained by the active user after receiving her answer and the cumulative score of each user. The information is shared 30 times per second and it allows a smooth UX for every participant. The bandwidth usage is low since only basic data types such as strings and numbers are shared between users and the delay is below 15 milliseconds on both WiFi and mobile networks. A previous approach tried to combine message passing and the transmission of the screen of the active user, using WebRTC, to the students using a PC to better simulate the AR experience (Matsumoto et al. 2023). Unfortunately, such a solution has proven not to be scalable. Due to poor Unity support for WebRTC servers such as Janus, the application suffered delays which severely impacted the performance. With more than five users, the UX was severely affected, and the app became unusable when more than ten users were connected to the same session.

To comply with the data analytics design objective (DO3), the application enables data collection through the storage of eXperience API (xAPI) statements on a Learning Locker⁶ instance. Learning Locker is the standard data repository for storing learning activity statements generated by xAPI. eXperience API is a web service that enables the secure sending and storing of learning experiences to an LRS. xAPI statements use JSON format and at their core they are formed by the triplet *Actor-Verb-Object*. The Actor represents the person performing a specific action (the Verb), while the Object could be another person or an xAPI activity on which the actor acts upon. xAPI statements can optionally include additional information such as Timestamps, Context or Results, to provide more detailed information. Storing statements across each session enables the application to keep track of user activity and to store additional logging messages that simplify application debugging. Learning Locker provides basic analytics and plotting capabilities through a web interface, as well as filtering and exporting the data in CSV format. These functionalities have been extended through the development of a Python package⁷ that includes methods to perform advanced data exploration and plotting, as well as running common machine learning models on xAPI statements data. One of the aims of the package is to simplify data analysis as much as possible, enabling teachers without development skills to extract valuable information from the collected data. For this reason, the package directly integrates GPT-4 (OpenAI 2023; Osmulski 2023), so that users with a valid OpenAI Key can use natural language to debug or generate code when needed. The package has been used to analyse the data

⁵ https://www.mediawiki.org/wiki/Wikimedia_REST_API.

⁶ <https://learninglocker.atlassian.net/wiki/spaces/DOCS/overview>.

⁷ https://stocastico.github.io/xapi_analysis/.

Table 1 Details of the demographics and number of devices used across each test

Participants	Gender		Players		Watchers	Total
	Males	Females	Tablets	Smartphones		
14 year-old	9	8	4	5	8	17
17 year-old	16	1	3	6	8	17
19 year-old	6	4	3	6	1	10
Total	31	13	10	17	17	44

collected during the evaluation of the app and the results will be presented in Sect. 5.

Finally, to demonstrate how the aforementioned architecture enables developers to easily create multi-user applications (DO4), we asked the developer of *ARoundTheWorld* if and how the architecture helped him in the development process. The developer mentioned that the architecture API enabled him to extend the application to enable multiple users in a transparent way. He could avoid having to deal with low-level networking issues or having to implement platform specific methods. While it was not possible to estimate the amount of lines of code or hours saved by its usage, the developer said that he was satisfied by the capabilities of the architecture and would use it again for future projects. Nevertheless, in order to enable teachers to create an ecosystem of collaborative AR applications for education, the developer mentioned that the availability of authoring tools to easily create applications on top of the *cleAR* architectural design would be desirable.

4 Evaluation

Our study aims to investigate how collaborative AR solutions may benefit the learning experience, and what, if any, are the usability issues of multi-user applications that can be used on different devices such as tablets, mobile phones or laptops. In the literature, there is no agreement on how to conduct evaluation of AR-based educational apps. The survey from Santos et al. (2013) analyses 87 AR applications and the evaluation protocols included interviews, observing and coding overt behaviour and expert reviews. Of those who used questionnaires, the majority crafted their own. Among the works that used established questionnaires, some relied on the ISONORM Usability questionnaire (Prümper 1999), Technology Acceptance Model (Davis and Venkatesh 1996), Constructivist Multimedia Learning Environment Survey (Maor 1999), Instructional Material Material Survey (Keller 1987), Intrinsic Motivation Inventory (Ryan and Deci 2000). The number of participants in the evaluation of AR applications for education varies a lot depending on the study. A systematic review of the literature conducted by Masneri et al. (2020) on articles published between 2015 and 2020 shows that

this number varies between 2 and 290 participants, while another survey by Santos et al. (2013) analysed studies where the number of participants ranged from 4 to 419.

In this work, a questionnaire which adapts and extends the Positive System Usability Scale (P-SUS) (Brooke et al. 1996; Sauro and Lewis 2011), with a few additional questions added to specifically evaluate the collaborative capabilities of the application, was developed and used. The questions, presented in Appendix A, were grouped into four classes depending on what they were evaluating: the interest of the application as an educational tool, the usability of the app, its collaboration capabilities and its functionality. Additionally, the participants could provide free-form feedback about the overall experience and whether they would recommend it to other students. Finally, we also conducted an interview with the teachers to collect their feedback about the learning experience, how collaboration may impact the involvement of the students, how AR apps could be used to evaluate the students knowledge of a subject and how they would take advantage of the data collected through the application.

At the beginning of the evaluation, the participants were briefed about the experiment and its purpose and were asked to sign a consent form. The questionnaires were anonymous but had an ID associated, so that during data analysis we were able to associate the answers to the questionnaires with the data collected from each device through the xAPI statements. The evaluation involved 44 students from three schools in San Sebastian between March and May 2023. Each experiment involved students of different ages (14, 17 and 19 years old) and their corresponding teachers. None of the participants had previous experience with AR applications, but they were familiar with the hardware devices (smartphones, tablets and PCs) used during the evaluation. In each school, the students were split into two groups: the first one represented the *players*, tasked with answering the quiz questions using the application on mobile or tablet devices, while the second group represented the *watchers*, who used a laptop or PC to see how the students answered the quiz and provided suggestions along the way. Table 1 shows the number of participants in each experiment, as well as details about the type of devices used when interacting with the application.

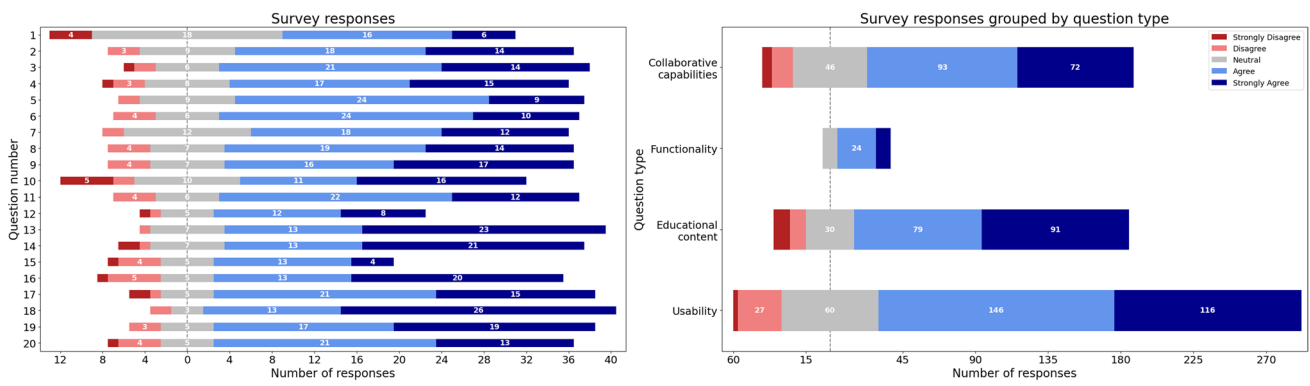


Fig. 4 Left: survey results on each question. Right: survey results grouped by question type

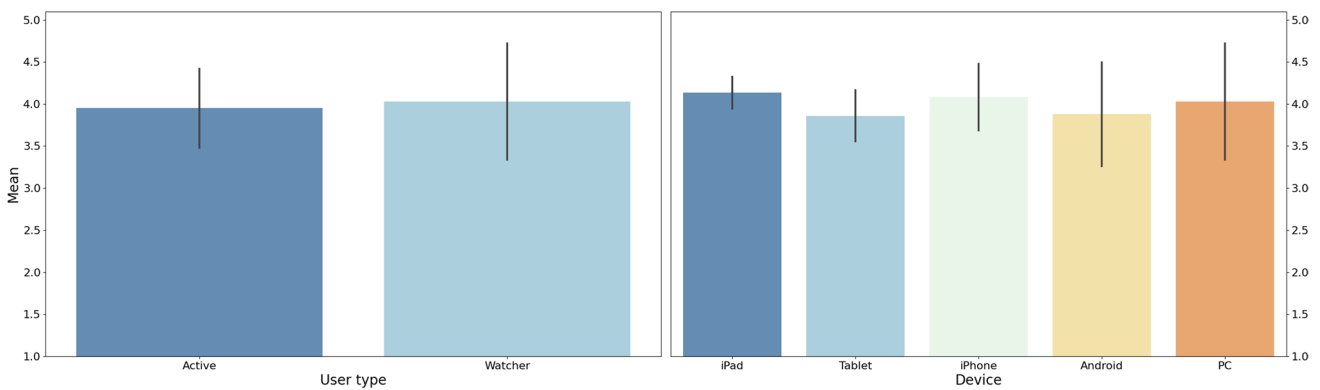


Fig. 5 Mean score of the questionnaire answers. The vertical bars represent the standard deviation

Once each participant had a device assigned, they were asked to connect to the application by selecting the session ID representing the experiment and the user ID which would be used as the *Actor* value for the xAPI statements generated while using the application. After a short Q&A session to clarify doubts about the app usage the teacher started the quiz and the students would then take turns to answer two sets of questions. Once the quiz was over, the teacher stopped the data collection and was able to check the score of each student and to export the data. After logging out of the session, the students filled in the questionnaires while we conducted the post-study interview with the teacher.

5 Results and discussion

In this Section, we present and discuss the results of the survey as well as the data collected from the app. An extended version, which also includes a subgroup analysis of the data split for different age group, is available in the data

repository⁸. The quantitative results from the responses to the post-intervention survey (shown in Appendix A) are summarised in Fig. 4. They are shown as stacked bar charts (Friedman and Amoo 1999), where the chart on the left refers to the answers to each question and the chart on the right to the groups of questions mentioned in Sect. 4. From the figure, it can be appreciated that the application was very well received by the students and that every question except the first one was answered positively (“Agree” or “Strongly agree”) more than 60% of the time. The average rating for each question ranges from 3.45 to 4.43, with limited variability across answers, with the standard deviation being below 1 for most of the questions. The plot on the right in Fig. 4 shows similar results: the students assigned the highest score to questions related to usability, especially questions 2 (“I found the application to be simple”) and 3 (“I thought the application was easy to use”). The questions related to functionality received a high score as well, while the ones relating to the educational content of the app show the highest variability: those questions received the highest

⁸ https://github.com/Stocastico/Evaluation_paper

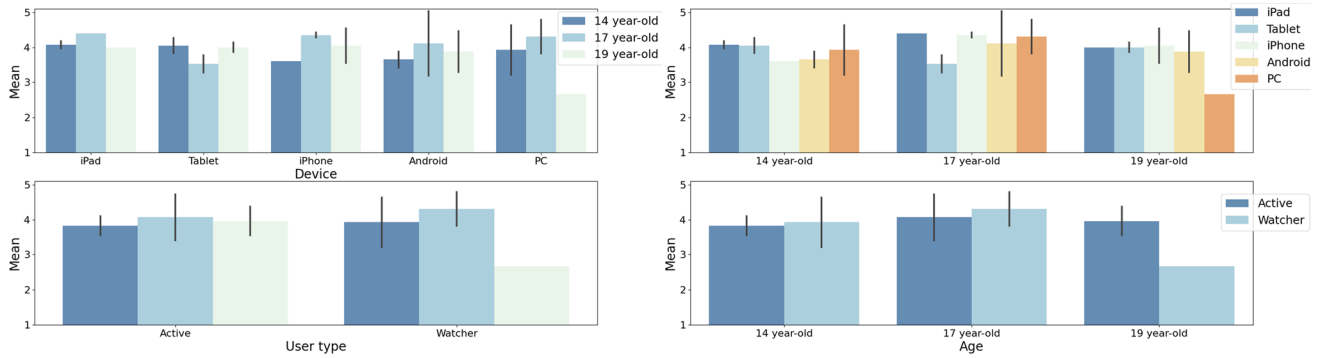


Fig. 6 Survey results split by user type. Left: Average question score by device used and student role. Right: average question score by students age

amount of negative answers while also receiving the highest score more than 40% of the time.

The bar plots of Fig. 5 are used to identify differences in the answers of the students based on their role when using the app and the device they used. Somewhat surprisingly the *watchers* gave a slightly higher mean score, albeit with a much higher variability in the answer. As for the device type, the users on an Apple device (iPhone or iPad) gave a higher score compared to students using an Android device or a PC, but the differences are not statistically relevant.

Figure 6 shows the mean score and the standard deviation for the questionnaire answers grouped by user. We split the students by age, their role when using the app and the device they used. From this visualization, an outlier can be easily identified, represented by the only student in the 19-year-olds’ group who used a PC and was the only non-active user in that session. The reason for the lower score, as identified by the comments provided by the student in the questionnaire, was that the experience for him did not feel particularly immersive nor collaborative, as his role was fairly different from that of his classmates.

Since *ARoundTheWorld* collected data in the form of xAPI statements, an analysis of the data received was conducted in order to detect whether there was a correlation between the score in the questionnaire and the number of statements collected by the application for each user. The actions registered by the app include both interaction between students, such as the suggestions sent, and the interactions of a user with the app. As shown in Fig. 7, there is a high variability in how much the students interacted. It is clear, though, that the *players* (the students using a mobile device and interacting with the augmented content) were much more involved in using *ARoundTheWorld*. This is probably because their role was much more interactive and immersive.

An interesting aspect to analyse is whether there is any correlation between the number of interactions of each student and the answers they have given to the survey questions.

Two statistical approaches were followed. First, a correlation analysis was performed to check whether there was a relation between the number of interactions and the average scores given to the questions by the students, by calculating the Pearson correlation coefficient. The second approach is that of hypothesis testing. This checks whether the answers given by students with a high number of interactions are significantly different from the answers given by students who had a low number of interactions. A two-sample T test assuming equal variances has been used for testing. In both cases, a significance level of $p < 0.05$ was established. Since the interactions between the *watchers* (students on a PC) and *players* (students on a mobile device) are significantly different, we also performed the analysis for these specific subsets of the data as well.

Unfortunately, the analysis performed is not conclusive. None of the tests returned a p value below the significance level, and the correlations identified (most notably between the interactions of students on a PC and their answers to the survey) are not statistically significant.

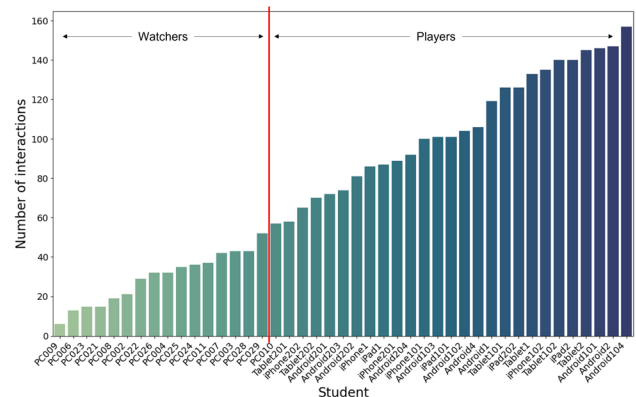


Fig. 7 Interactions with the application for each student (identified by the device used)

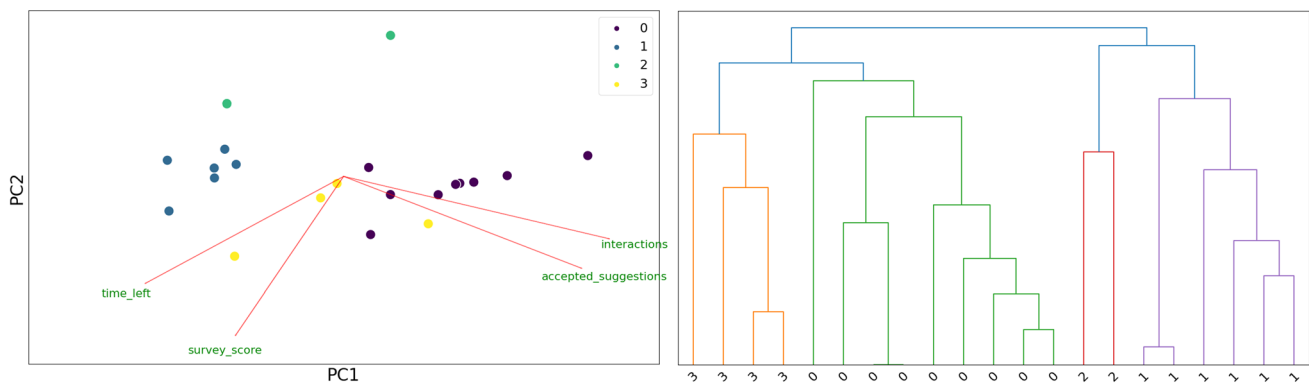


Fig. 8 Left: Clustering of the active users data on the PCA space. Right: the dendrogram representing the hierarchical clustering

A similar analysis was conducted to inspect whether there were correlations between the score obtained in the quiz and the answers in the survey. In this case as well, no statistically significant correlation was found. This was expected since the app was designed to ask questions of varying difficulties, but the difficulty level of the questions received by a student did not change during the test. As expected, students who received easier questions achieved a higher score and there is indeed a significant correlation between these two variables.

Another correlation we wanted to analyse was the one between the mean score obtained in the survey and the number of suggestions sent by the user. The statements about suggestions are an interesting variable because for each question, every user (besides the *active player*, the one answering the current question) was able to send two suggestions. For this reason, many such statements were collected during the trial. To encourage students to provide suggestions, the application assigned points for each correct suggestion. In this case, the analysis showed a significant positive correlation ($r = .37$, $p = .044$), meaning that the most engaged students were the ones that gave a higher score in the survey.

Finally, we wanted to perform a clustering analysis of the data, to check if we could identify distinct groups of users. In this case, we focused only on the students who used a mobile device, since they provided a greater number of features to work with. We considered as variables of interest the average time left per question, the number of suggestion accepted, the total number of interactions and the mean value of the answers in the survey. A dimensionality reduction using PCA (Jolliffe 2002), shown on the left in Fig. 8, revealed that the first two principal components explained more than 70% of the variance in the data. Additionally, a biplot analysis indicated that the most relevant variable for the first

principal component was the number of user interactions, while for the second one it was the results of the survey.

As the number of data points is small, we used a hierarchical clustering algorithm (Caliński and Harabasz 1974). A Silhouette score (Rousseeuw 1987) computed for cuts between 2 and 6 suggested that the optimal number of clusters in this case was either 2 or 4. The clustering results (shown in Fig. 8, right) identified one big group of students characterized by having a particularly high number of interactions and another one having a higher score in the survey answer. The other two clusters were harder to characterize. In one case, we could not clearly identify a common feature in the data, while in the other the cluster only contained two members, and the intra-cluster variable suggests that those data points are probably outliers.

After running the trials in school, we also conducted a post-intervention interview with the teachers. The three teachers seemed very intrigued by the possibility of easily being able to use AR in school without having to resort to any specific hardware. They especially valued the fact that the collaborative features of *ARoundTheWorld* encouraged the students to work together to provide the answer, either through the features of the application or simply by talking to each other. Another relevant point for the teachers was the possibility of adding new content on their own, as well as the fact that they could export the results to the school LMS. The teachers were more sceptical about the AI features provided by the backend. They mentioned that the vast majority of their colleagues do not have sufficient knowledge to perform the analysis on their own. They would rather prefer using a PowerBI or Tableau interface to visualize data and extract basic reports. The teachers also mentioned that the role of the *watchers* was too passive and that in longer experiments these students might lose interest. They suggested enabling the role of active user when using a PC, even

if that means not using AR components but a browser-based 3D graphics library.

6 Conclusions and future work

In this work, we presented *ARoundTheWorld*, a multiplatform AR application which implements collaborative capabilities and gamification concepts. The application is based on a Geography quiz and it fulfils the design objectives identified in Masneri et al. (2023). The evaluation, conducted with 44 students and 3 teachers, and the analysis of the xAPI statements showed that students evaluated very positively the application. Additionally, we measured a small but statistically significant correlation between the ratings in the questionnaire and the engagement of the students. Furthermore, post-study interviews with the teachers identified the collaborative capabilities and the possibility of personalizing the app content as being key factors for a sustained usage of AR apps. In fact, one of the teachers suggested the possibility of adding more collaborative features, such as a chat system or speech-based interactions to make the application more immersive and more appealing when used in a distributed setting.

Future work directions include testing the application with a larger sample size, in schools spread over a wider geographical area. This would reduce the noise in the data collected while also allowing a more statistically robust analysis of the data. Another line of work is the creation

of a software suite that simplifies the analysis of the data for teachers without a software development background, as well as an authoring tool to further simplify the development of collaborative AR applications. While this is not a new idea (Rajaram and Nebeling 2022; Thanyadit et al. 2022), accomplishing a widespread diffusion of AR applications in education requires software solutions that allow quick and easy development, content personalisation after the release of the app and the creation of a central repository for storage and sharing of assets, plugins and applications. We believe that this work represents a first step towards this goal.

Appendix A: Questionnaire

Students were asked to fill a 20-item subjective questionnaire, using a Likert scale from 1 to 5 to assess their agreement with each sentence. The items in the questionnaire belong to four different clusters, depending on the aspect to evaluate (collaborative aspects, app usability, functionality, interest as an educational tool). Questions Q12 and Q15 were filled only by students who used the mobile applications, as the ones using the web interface could not receive suggestions but only provide them. Q18 and Q20 were framed slightly differently for users on a mobile device or using a PC: for the former group, the question referred to the usage of augmented reality, while for the latter it was about the inclusion of 3D elements in the application (Table 2).

Table 2 List of questions in the post-intervention questionnaire

	Question
Q1	I think that I would like to use the application frequently
Q2	I found the application to be simple
Q3	I thought the application was easy to use
Q4	I think that I could use the application without the support of a technical person
Q5	I found the various functions in the application were well integrated
Q6	I would imagine that most people would learn to use the application very quickly
Q7	I found the application very intuitive
Q8	I felt very confident using the application
Q9	I could use the application without having to learn anything new
Q10	I would like to use the application during a test
Q11	Being able to provide suggestions made me feel more involved
Q12	Receiving suggestions made me more confident when answering a question
Q13	At all times I have been able to understand what the person who had to respond to the exercise was doing
Q14	I find it more interesting to solve the exercises through the application than through a web page or in writing
Q15	Suggestions from my classmates have helped me when answering the exercise
Q16	The device used has allowed me to use the application easily
Q17	I would like to use the application to learn new concepts
Q18	Being able to use augmented reality/ 3D elements makes the application more entertaining
Q19	There are several ways to collaborate with my classmates through the application
Q20	Thanks to augmented reality / 3D elements I have felt immersed in the learning activity.

Acknowledgements The research in Vicomtech has been supported by European Union's Horizon 2020 research and innovation programme under grant agreement No 856533, project ARETE (Augmented Reality Interactive Educational System). The research in the UPV/EHU has been partially supported by the ADIAN grant IT-1437-22 from the Basque Government and grant PID2021-127777OB-C21 of MCIU/AEI/FEDER, UE from the Spanish Government. The authors would like to thank Nahia Ugarte and Pello Bereziartua as well as their students for helping us perform the user evaluations at their schools. The consent forms were prepared with the help and approval of the Ethics commission of the University of the Basque Country (M10_2022_362).

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Data Availability Statement A cleaned and anonymised version of the datasets generated during the current study and the corresponding data analysis are available at the [following repository](#), released on an Open Data Commons Attribution License (ODC-BY).

Declarations

Conflict of interest The authors declare that they have no competing interests with regards to the submitted work.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Akçayır M, Akçayır G (2017) Advantages and challenges associated with augmented reality for education: a systematic review of the literature. *Educ Res Rev* 20:1–11. <https://doi.org/10.1016/j.edurev.2016.11.002>
- Alrashidi M, Gardner M, Callaghan V (2017) Evaluating the use of pedagogical virtual machine with augmented reality to support learning embedded computing activity. In: Proceedings of the 9th international conference on computer and automation engineering, pp 44–50. <https://doi.org/10.1145/3057039.3057088>
- Avila-Garzon C, Bacca-Acosta J, Duarte J et al (2021) Augmented reality in education: an overview of twenty-five years of research. *Contemp Educ Technol* 13(3):21. <https://doi.org/10.30935/cedtech/10865>
- Azuma R, Bailiot Y, Behringer R et al (2001) Recent advances in augmented reality. *IEEE Comput Graphics Appl* 21(6):34–47. <https://doi.org/10.1109/38.963459>
- Billinghurst M, Kato H (2002) Collaborative augmented reality. *Commun ACM* 45(7):64–70. <https://doi.org/10.1145/514236.514265>
- Bressler DM, Bodzin AM (2013) A mixed methods assessment of students' flow experiences during a mobile augmented reality science game. *J Comput Assist Learn* 29(6):505–517. <https://doi.org/10.1111/jcal.12008>
- Brooke J et al (1996) SUS—a quick and dirty usability scale. *Usabil Evaluat Ind* 189(194):4–7
- Cai S, Chiang FK, Sun Y et al (2017) Applications of augmented reality-based natural interactive learning in magnetic field instruction. *Interact Learn Environ* 25(6):778–791. <https://doi.org/10.1080/10494820.2016.1181094>
- Caliński T, Harabasz J (1974) A dendrite method for cluster analysis. *Commun Stat* 3(1):1–27. <https://doi.org/10.1080/03610927408827101>
- Catling S (1993) The whole world in our hands. *Geography* 78(4):340–358
- Caudell TP, Mizell DW (1992) Augmented reality: an application of heads-up display technology to manual manufacturing processes. In: Proceedings of the twenty-fifth hawaii international conference on system sciences, pp 659–669
- Çelik F, Yangın E, Eranlı C (2022) The use of augmented reality in a gamified CLIL lesson and students' achievements and attitudes: a quasi-experimental study. *Smart Learn Environ* 9(1):30. <https://doi.org/10.1186/s40561-022-00211-z>
- Chang HY, Binali T, Liang JC et al (2022) Ten years of augmented reality in education: a meta-analysis of (quasi-) experimental studies to investigate the impact. *Comput Educ* 191(104):641. <https://doi.org/10.1016/j.compedu.2022.104641>
- Chen P, Liu X, Cheng W, et al (2017) A review of using augmented reality in education from 2011 to 2016. In: *Innovations in Smart Learning*. Springer Singapore, Singapore, pp 13–18. https://doi.org/10.1007/978-981-10-2419-1_2
- Choi J, Yoon B, Jung C, et al (2017) Arclassnote: augmented reality based remote education solution with tag recognition and shared hand-written note. In: 2017 IEEE international symposium on mixed and augmented reality (ISMAR-Adjunct). IEEE, pp 303–309. <https://doi.org/10.1109/ISMAR-Adjunct.2017.94>
- Clarke B, et al (2020) xAPI-Spec. <https://github.com/adlnet/xAPI-Spec>. Accessed 2023-03-01
- Collins L (2018) The impact of paper versus digital map technology on students' spatial thinking skill acquisition. *J Geogr* 117(4):137–152. <https://doi.org/10.1080/00221341.2017.1374990>
- Commission E, Directorate-General for Communications Networks C, Technology, et al (2023) Extended reality-opportunities, success stories and challenges (health, education)-final report. Publications Office of the European Union. <https://doi.org/10.2759/121671>
- da Silva MM, Teixeira JMX, Cavalcante PS et al (2019) Perspectives on how to evaluate augmented reality technology tools for education: a systematic review. *J Braz Comput Soc* 25:1–18. <https://doi.org/10.1186/s13173-019-0084-8>
- Das P, Zhu M, McLaughlin L et al (2017) Augmented reality video games: new possibilities and implications for children and adolescents. *Multimodal Technol Interact* 1(2):8. <https://doi.org/10.3390/mti1020008>
- Davis FD, Venkatesh V (1996) A critical assessment of potential measurement biases in the technology acceptance model: three experiments. *Int J Hum Comput Stud* 45(1):19–45. <https://doi.org/10.1006/ijhc.1996.0040>
- Dinis FM, Guimarães AS, Carvalho BR, et al (2017) Virtual and augmented reality game-based applications to civil engineering education. In: 2017 IEEE global engineering education conference (EDUCON), pp 1683–1688. <https://doi.org/10.1109/EDUCON.2017.7943075>
- Friedman HH, Amoo T (1999) Rating the rating scales. In: Friedman HH, Amoo T (1999) Rating the Rating Scales. *Journal of Marketing Management*, Winter, pp 114–123. <https://ssrn.com/abstract=2333648>
- Garzón J, Pavón J, Baldiris S (2019) Systematic review and meta-analysis of augmented reality in educational settings. *Virtual Real* 23(4):447–459. <https://doi.org/10.1007/s10055-019-00379-9>

- Ibáñez MB, Delgado-Kloos C (2018) Augmented reality for stem learning: a systematic review. *Comput Educ* 123:109–123
- Iqbal MZ, Mangina E, Campbell AG (2022) Current challenges and future research directions in augmented reality for education. *Multimodal Technol Interact* 6(9):75. <https://doi.org/10.3390/mti6090075>
- Jolliffe IT (2002) Principal component analysis for special types of data. Springer, Berlin. https://doi.org/10.1007/0-387-22440-8_13
- Keller JM (1987) Development and use of the arcs model of instructional design. *J Instr Dev* 10(3):2–10. <https://doi.org/10.1007/s10055-023-00764-5>
- Laviole J, Thevin L, Albouys-Perrois J, et al (2018) Nectar: multi-user spatial augmented reality for everyone: three live demonstrations of educative applications. In: Proceedings of the virtual reality international conference-laval virtual, pp 1–6. <https://doi.org/10.1145/3234253.3234317>
- Law ELC, Heintz M (2021) Augmented reality applications for K-12 education: a systematic review from the usability and user experience perspective. *Int J Child Comput Interact* 30:100–321. <https://doi.org/10.1016/j.ijcci.2021.100321>
- López-Faicán L, Jaen J (2020) Emofindar: evaluation of a mobile multiplayer augmented reality game for primary school children. *Comput Educ* 149(103):814. <https://doi.org/10.1016/j.compedu.2020.103814>
- Maor D (1999) A teacher professional development program on using a constructivist multimedia learning environment. *Learn Environ Res* 2(3):307–330. <https://doi.org/10.1023/A:1009915305353>
- Masneri S, Dominguez A, Wild F, et al (2020a) Work-in-progress-ARETE-an interactive educational system using augmented reality. In: 2020 6th International conference of the immersive learning research network (iLRN). IEEE, pp 283–286. <https://doi.org/10.23919/iLRN47897.2020.9155186>
- Masneri S, Domínguez A, Sanz M et al (2022b) Collaborative multi-user augmented reality solutions in the classroom. In: Auer ME, Hortsch H, Michler O et al (eds) Mobility for smart cities and regional development-challenges for higher education. Springer, Cham, pp 1004–1011. https://doi.org/10.1007/978-3-030-93907-6_106
- Masneri S, Domínguez A, Zorrilla M et al (2022c) Interactive, collaborative and multi-user augmented reality applications in primary and secondary education: a systematic review. *JUCS J Univ Comput Sci* 28(6):564–590
- Masneri S, Domínguez A, Sanz M et al (2023) cleAR: an interoperable architecture for multi-user ar-based school curricula. *Virtual Real* 27:1813–1825. <https://doi.org/10.1007/s10055-023-00764-5>
- Matsumoto K, Kannan T, Tanuwijaya S, et al (2023) Unity render streaming. <https://github.com/Unity-Technologies/UnityRenderStreaming>
- McLean G, Wilson A (2019) Shopping in the digital world: examining customer engagement through augmented reality mobile applications. *Comput Hum Behav* 101:210–224. <https://doi.org/10.1016/j.chb.2019.07.002>
- Milgram P, Kishino F (1994) A taxonomy of mixed reality visual displays. *IEICE Trans Inf Syst* 77(12):1321–1329
- Oh S, So HJ, Gaydos M (2017) Hybrid augmented reality for participatory learning: the hidden efficacy of multi-user game-based simulation. *IEEE Trans Learn Technol* 11(1):115–127. <https://doi.org/10.1109/TLT.2017.2750673>
- Ong SK, Yuan M, Nee AY (2008) Augmented reality applications in manufacturing: a survey. *Int J Prod Res* 46(10):2707–2742. <https://doi.org/10.1080/00207540601064773>
- OpenAI (2023) GPT-4 technical report. [arXiv:2303.08774](https://arxiv.org/abs/2303.08774)
- Osmulski R (2023) Ask AI-library for talking to ChatGPT from within Jupyter Notebook. https://github.com/radekosmulski/ask_ai
- Palaigeorgiou G, Karakostas A, Skenteridou K (2018) Touching and traveling on 3d augmented tangible maps for learning geography. *Interact Technol Smart Educ* 15(3):279–290. <https://doi.org/10.1108/ITSE-12-2017-0066>
- Pan X, Zheng M, Xu X et al (2021) Knowing your student: targeted teaching decision support through asymmetric mixed reality collaborative learning. *IEEE Access* 9:164742–164751. <https://doi.org/10.1109/ACCESS.2021.3134589>
- Pellas N, Fotaris P, Kazanidis I et al (2019) Augmenting the learning experience in primary and secondary school education: a systematic review of recent trends in augmented reality game-based learning. *Virtual Real* 23(4):329–346. <https://doi.org/10.1007/s10055-018-0347-2>
- Prümper J (1999) Test it: Isonorm 9241/10. In: HCI (1), 1028–1032
- Rajaram S, Nebeling M (2022) Paper trail: an immersive authoring system for augmented reality instructional experiences. In: Proceedings of the 2022 CHI conference on human factors in computing systems, pp 1–16. <https://doi.org/10.1145/3491102.3517486>
- Rousseeuw PJ (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 20:53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Ryan RM, Deci EL (2000) Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *Am Psychol* 55(1):68. <https://doi.org/10.1037/0003-066X.55.1.68>
- Santos MEC, Chen A, Taketomi T et al (2013) Augmented reality learning experiences: survey of prototype design and evaluation. *IEEE Trans Learn Technol* 7(1):38–56. <https://doi.org/10.1109/TLT.2013.37>
- Sauro J, Lewis JR (2011) When designing usability questionnaires, does it hurt to be positive? In: Proceedings of the SIGCHI conference on human factors in computing systems, pp 2215–2224. <https://doi.org/10.1145/1978942.1979266>
- Secretan J, Wild F, Guest W (2019) Learning analytics in augmented reality: blueprint for an AR / xAPI framework. In: 2019 IEEE international conference on engineering, technology and education (TALE), pp 1–6. <https://doi.org/10.1109/TALE48000.2019.9225931>
- Sirakaya M, Cakmak EK (2018) The effect of augmented reality use on achievement, misconception and course engagement. *Contemp Educ Technol* 9(3):297–314. <https://doi.org/10.30935/cet.444119>
- Takahashi I, Oki M, Bourreau B, et al (2018) An empathic design approach to an augmented gymnasium in a special needs school setting. *Int J Des* 12(3)
- Thamrongrat P, Law ELC (2019) Design and evaluation of an augmented reality app for learning geometric shapes in 3D. In: IFIP conference on human-computer interaction. Springer, pp 364–385. <https://doi.org/10.1007/978-3-030-29390-1>
- Thanyadit S, Heintz M, Law ELC et al (2022) How easy is it to extend reality? a usability study of authoring toolkits. *Human Interact Emerg Technol (IHET 2022): Artif Intell Future Appl* 68:9. <https://doi.org/10.54941/ahfe1002707>
- Vlachos A, Perifanou M, Economides AA (2022) A review of ontologies for augmented reality cultural heritage applications. *J Cultur Heritage Manag Sustain Dev* (ahead-of-print). <https://doi.org/10.1108/JCHMSD-06-2021-0110>
- Wu Y, Guo S, Zhu L (2020) Design and implementation of data collection mechanism for 3D design course based on xAPI standard. *Interact Learn Environ* 28(5):602–619. <https://doi.org/10.1080/10494820.2019.1696842>
- Xeferis S, Palaigeorgiou G (2019) Mixing educational robotics, tangibles and mixed reality environments for the interdisciplinary learning of geography and history. *Int J Eng Pedagog* 9(2):82–98. <https://doi.org/10.3991/ijep.v9i2.9950>