



Impact and analysis of a collaborative augmented reality educational environment

Naveed Ahmed¹ · Mohammed Lataifeh¹

Received: 31 January 2023 / Revised: 11 April 2023 / Accepted: 11 April 2023
© Beijing Normal University 2023

Abstract

This work presents “LeARn,” a new network-based collaborative learning environment that employs augmented reality to transform a real-world surface into a virtual lab. To demonstrate the system, a scenario with a virtual chemistry lab is presented. In this demo, any real-world surface is augmented by virtual lab equipment utilized in a chemistry experiment. The virtual lab is hosted by the instructor, and all the students can join the lab only using their mobile phones or tablets. Each member can interact with the lab equipment, which can be visualized in real-time by the instructor or fellow students. The system allows for real-time communication that fosters a truly collaborative environment. A comprehensive user study involving 72 participants (36 female, and 36 male) is conducted to evaluate the usability and learning impact of the proposed system. The participants were first year and second year undergraduate students (17 to 19 years of age) with an equal distribution over the two academic years. The user study employs both direct and indirect observations, combined with multiple surveys to collect both quantitative and qualitative data. The data analysis is performed from multiple viewpoints, and the collaborative AR learning environment is evaluated in terms of its usability, effectiveness of its collaboration functionality, and its impact on learning. Finally, the results also evaluate the feasibility of employing the system together with face-to-face learning.

Keywords Collaborative learning · Education · Augmented reality · Learning impact · User interaction

✉ Naveed Ahmed
nahmed@sharjah.ac.ae
Mohammed Lataifeh
mlataifeh@sharjah.ac.ae

¹ Department of Computer Science, University of Sharjah, Sharjah, United Arab Emirates

Introduction

The ever-evolving technological landscape is having a profound impact on all aspects of the educational ecosystem (Tamim et al., 2011). The magnitude of the impact is the same at the micro level from the individual learning or teaching experience to the macro level of the overall learning environment for all stakeholders. The advent of Covid-19 pandemic had a devastating impact on the learning process and resulted in a hasty transition toward full-time remote learning (Rodrigues et al., 2020). Several tools and technologies were adopted to facilitate this remote learning process. These tools and technologies are generally geared toward the dissemination of learning content from the instructor to students, therefore, they often result in a lower engagement level as the experience does not achieve the same level of interactivity as a face-to-face classroom (Lucas et al., 2020). In addition, most of the employed technology is not designed to foster and encourage collaboration (Zhou et al., 2021).

Social interaction and collaboration are some of the most useful qualities of an educational environment that result in a higher level of engagement and an increased learning impact (Garcia-Sanjuan et al., 2018). Collaboration requires people to be in the same physical space, where the learning process involves both communication and visualization. While communication allows for a more traditional learning mechanism in the form of instructions, in a collaborative environment, feedback from peers plays an equally significant role. Similarly, visualization plays an equally vital role during collaboration. Not only one learns from the demonstration from the instructor, but visualizing peers and learning from their work is a crucial part of the learning process (Laal & Ghodsi, 2012).

While most remote learning utilizes bi-directional audio and visual transmissions over desktops and smartphones, some of the recent approaches also employ extended reality (XR) for a novel experience. XR which encompasses virtual reality (VR), augmented reality (AR), and mixed reality (MR) has proven to be an effective platform for remote learning (Pantelidis, 2010). VR has been demonstrated to be a viable platform in a learning environment that can lead to higher students' engagement (Roussos et al., 1999; Winn, 1993). One of the downsides of VR is the requirement of a somewhat cumbersome head-mounted display (HMD) that is ergonomically not feasible for a longer period of use. In contrast, the entry barrier for AR is quite lower and any camera-based device, such as a smartphone can be employed to create a compelling AR experience (Azuma, 1997).

AR experiences are increasingly employed in all levels of educational environments and are having a transformative effect on the overall learning process (Akçayır & Akçayır, 2017). These experiences span a number of fields, e.g., astronomy (Agarwal & Thakur, 2014), anatomy (Walker et al., 2017), STEM (Ibáñez & Delgado-Kloos, 2018), or entertainment (Mota et al., 2016, 2018). In addition, some of the recent AR systems are designed with a multi-user environment (Fidan & Tuncel, 2019; Ibáñez & Delgado-Kloos, 2018; Mota et al., 2016, 2018), but none of these methods evaluate their systems in terms of their

effectiveness of collaboration and their impact on the learning process. One of the recent works (López-Faicán & Jaen, 2020), performed a detailed user study on a multi-user AR game in a school environment. Although the study is quite thorough, the experiment setup is extremely simplistic, and the collaboration element is very limited as the actions performed by the user are very independent of each other due to the gamification of the environment which also encourages competitiveness. Therefore, it is important to deploy AR in a truly collaborative learning environment, where the tasks require a higher level of complexity and evaluate its effectiveness on the overall learning process.

In this work, we present “LeARN,” a collaborative AR learning environment that transforms a real-world environment into a virtual lab and allows multiple users to join and collaboratively perform an experiment. Students not only visualize their fellow students experimenting in real-time, but they can also communicate and help others as they do in a face-to-face environment. The complexity of the experiment is very high, as we developed a virtual qualitative inorganic analysis experiment that requires several delicate steps that in general take a lot of time and precision in a real-world environment. We demonstrate that even such a complex experiment can be effectively performed from the safety of a personalized space in a virtual environment while also keeping the characteristics of a face-to-face collaborative environment. Our system is quite easy to deploy and has a very low entry barrier, as it only requires a smartphone for both the AR environment and communication. To the best of our knowledge, this is a completely novel system, and no other system currently exists that provides a network-based AR collaboration with voice support for multiple users.

We perform an extensive user study to evaluate the usability, effectiveness, and learning impact of the proposed system. The user study enrolls participants from first and second-year academic levels. The effectiveness of collaboration is also extensively analyzed using a group-based analysis. The multi-parameter and multi-level collaboration analysis shows that students’ familiarity with each other coupled with their level of seniority affects the collaboration significantly.

A deeper insight is achieved by analyzing the data from a gender perspective, in both individual and group-based analysis. There have been several studies that analyze gender differences in the effectiveness, attitude, and adoption of educational technologies (Harsh et al., 2012; Ottemo et al., 2020; Punter et al., 2017; Reyhav & McHaney, 2017; Rodríguez-Ardura & Meseguer-Artola, 2021; Tondeur et al., 2016; Zhou & Xu, 2007). In one of the earlier studies (Zhou & Xu, 2007), the authors claimed that female students displayed a lower level of confidence and experience when using computers in educational settings. In Tondeur et al. (2016), authors argue that women have less favorable attitudes toward computers but are not disadvantaged toward their use as an educational tool. In another study (Punter et al., 2017), the authors argue that girls have an advantage in the more information-oriented dimensions that require sharing, evaluating, and reflecting processes. They also highlight the results where boys performed better when applying technical functionality. Additionally, analyzing gender differences in a collaborative learning environment using mobile technology (Reyhav & McHaney, 2017), demonstrated that female students tend to benefit more

from video-based material, whereas male students derive higher satisfaction from text-based learning. Indeed, the most recent work exploring gender perspective (Ottemo et al., 2020) argued for further gender contextualization toward gender-inclusive programs, as the flow (Csikszentmihalyi & Csikszentmihalyi, 1992) experiences have proven fundamentally distinct for male and female students (Rodríguez-Ardura & Meseguer-Artola, 2021).

In light of the above studies that explore the impact of gender differences in technology in education, we found it imperative to organize our user study to analyze the proposed system from the point of view of each gender. Therefore, our user study incorporates an even number of female and male students. In addition, the groups are formed to include female, male, and mixed-gender students. The collected quantitative data from both individual and collaborative sessions allows us to analyze gender differences, if any, in the use of AR technology in an educational environment.

Finally, the effectiveness and the learning impact of the collaborative AR educational environment are also extensively analyzed, which is a requirement if remote learning tools are to be employed in an educational setting (Garzón & Acevedo, 2019; Khan et al., 2019). Hence, the multi-level analysis in this work along with a user study comparing AR-based remote learning with face-to-face learning indicates that AR-based remote learning provides an enticing alternative to face-to-face learning under special circumstances.

The main contributions of our work are:

1. Design a novel collaborative learning AR environment that allows multi-user interaction in real-time.
2. Develop a complex virtual chemical experiment in an AR environment that requires a number of steps and fosters collaboration.
3. Perform a comprehensive user study to evaluate the effectiveness and learning impact of a collaborative augmented reality learning environment.
4. Analyze the effectiveness of collaboration for AR-based remote learning by employing students from diverse levels of academic experience and degrees of familiarity with each other.
5. Statistical data analysis from multiple viewpoints to correctly identify and classify parameters that influence collaboration and learning in a collaborative AR environment.
6. A comparison of AR-based remote learning with face-to-face learning to identify the scenarios where AR-based remote learning can be an effective substitute for the face-to-face experience.

In the following sections, we will first present the design and development of the collaborative augmented reality system. Afterward, the implemented chemistry experiment is presented in detail. Followed by a walkthrough of the system demo. In the following section, the user study is detailed with the hypothesis and experiment setup. Finally, the results are presented together with a detailed discussion, followed by conclusions.

System design and implementation

Our system had two main challenges to overcome. The first challenge was to develop a collaborative AR environment, the second was to design a complex virtual experiment that can truly validate the effectiveness of an augmented collaborative environment.

Designing a collaborative AR system

To design a collaborative AR system, we had to use the tools and technologies that allowed platform-independent AR and connectivity tools. Instead of implementing a custom solution, we employed Unity (Unity, 2022) which provides device-independent solutions for both AR and networking. Unity's AR Foundation Framework (UnityAR, 2022) includes core features from all the major AR platforms: ARKit, ARCore, MagicLeap, and HoloLens. In addition, Unity's XR Interaction Toolkit (UnityXR, 2022) was used to add interactivity to the AR application.

Unity's networking component: Photon Unity Networking (PUN) (UnityPUN, 2022) was adopted to implement online connectivity. PUN allows real-time interaction between multiple devices over the network and is also employed for real-time voice chat (UnityVoice, 2022). PUN is also cross-platform and complements AR Foundation Framework in allowing us to deploy the system on a range of devices for the user study. We implemented a network lobby where the instructor hosts the session, and students are able to join the session by connecting to the server.

Designing the experiment

We decided to implement a sufficiently complex experiment from the undergraduate-level Chemistry I course. As the course is taken by all of the Science and Engineering students either in their first or second year, it provides us with a rich set of students that could be used for user testing. A survey of students and faculty members recommended qualitative analysis of anions, as it is an experiment with an intermediate level of difficulty, and a collaborative environment has a significantly positive effect on students' performance. The experiment is used to find out the identity of anions in an unknown mixture. The experiments require the precise addition of a fixed number of drops of an unknown mixture to a test tube. This is followed by the precise addition of additional chemical solutions, depending on the test. The experiment also requires the use of litmus paper to check for acidity, the use of a centrifuge to separate precipitates from a solution, and the heating of test tubes using a water bath. We implemented the testing of the four common anions. In this test, an unknown solution is given to the students, and they need to follow the pre-defined steps to find out the composition of the unknown solution. It requires handling of the test tubes, dropper, water bath, centrifuge, and PH papers. The required equipment and the four tests are shown in Fig. 1.

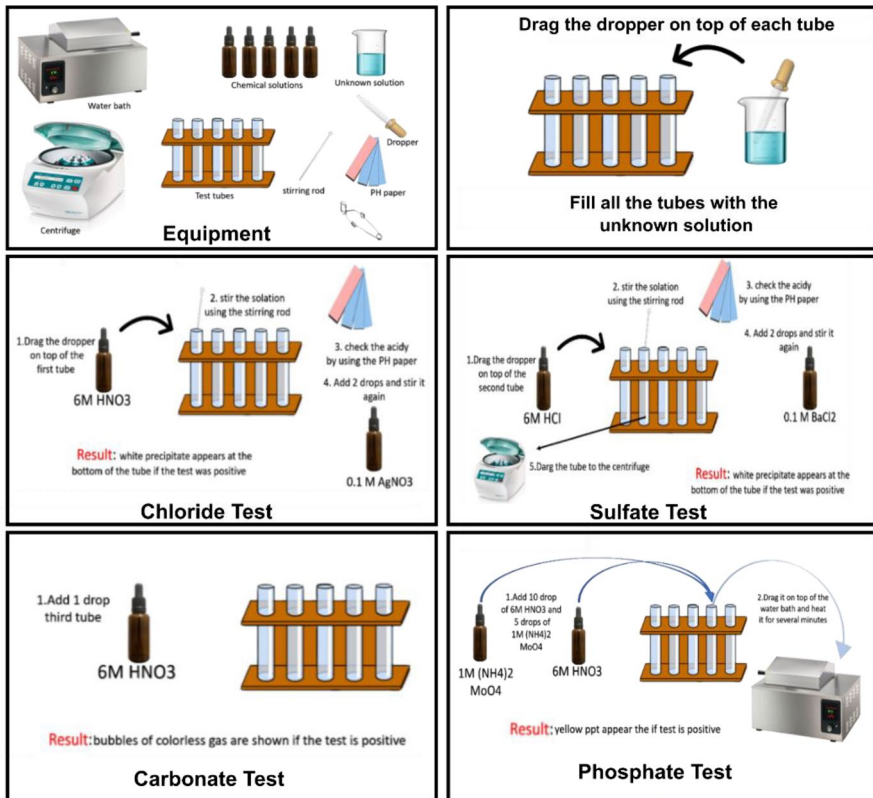


Fig. 1 An overview of the experiment showing the equipment and two of the tests

System demo

The system requires a user login at the start (Fig. 2). The successful login takes the user to the lobby where they either need to create a session or join an existing session as shown in. Once the session starts, the user needs to point the camera toward any surface, and once the surface is identified, it is populated by the experiment setup. The user has the option to zoom in and out using the provided ‘+’ and ‘-’ buttons, as shown in Fig. 2.

The system runs in a synchronous network environment, where the instructor can do the demonstration, or ask a specific student to perform some steps. As everyone is watching the same experiment and communicating over the network it is easy to perform the steps in a collaborative manner, where each student can take turns or guide their fellow experimenting students. Such intervention is made easy for anyone in the group to immediately correct the mistakes or perform a step if it is proving difficult for one of their peers. Figure 3 shows a session simultaneously running on two devices over the network.

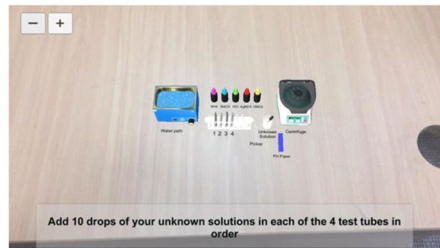


Fig. 2 Login screen (top), and the experiment setup (bottom) showing all the equipment on a real-world planar surface at different zoom levels

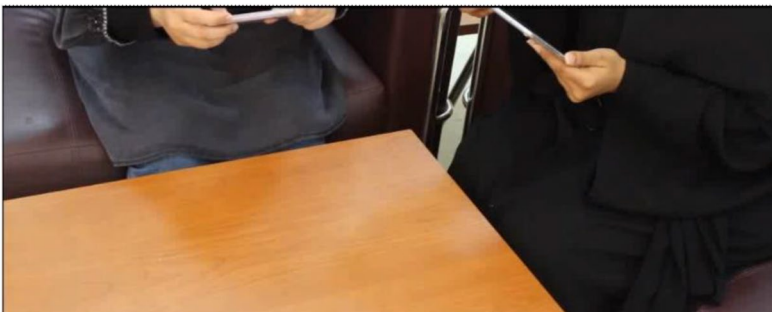
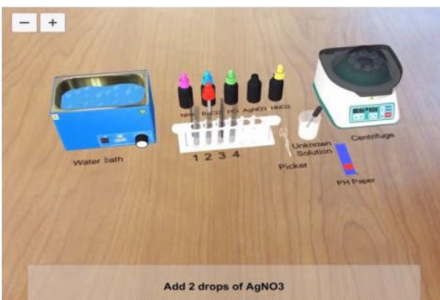


Fig. 3 An experiment running synchronously on two devices (bottom). The red dropper is used to add AgNO₃ by one of the users (top left) and can be observed by other users (top right)

User study

In the user study, the system was evaluated in terms of its usability, User Experience (UX), effectiveness, collaboration, and impact on learning. The following usability parameters were measured: easy to use, effective to use with good utility, and easy to learn. In terms of the UX, several positive and negative user experience factors are measured. Additionally, a survey was conducted to evaluate if the participants are satisfied with real-time collaboration components, and their usefulness during the experiment. Quantitative data in the form of experiment duration for individual and collaborative sessions were recorded. In addition, direct observation of collaborative sessions was used to identify the number of instances of collaboration between participants. Finally, an additional survey was carried out to measure the effectiveness of the system as a learning tool, and its impact on learning.

In order to test the system, we enlisted the help of 72 participants (17–19 years of age). The number of male and female participants is equal, and they are divided into 12 groups of 6 students each. Half of the students are from the first year, while the other half are from the second year. As the lab is a part of the introductory course, most of the students tend to take the course in the first or second year. In addition, three of the first-year and three of the second-year groups were randomly assigned, while the other three for each category were open to enrolment by choice. This allowed students who are familiar with each other to join a particular group and allowed us to gauge the magnitude and effectiveness of cooperation of students with different levels of familiarity with each other.

In order to correctly analyze the data in terms of gender performance, for both first and second-year categories, two groups were exclusive for female students, two for male students, and two for mixed genders. Table 1 shows the composition of 12 groups according to the above criteria.

As can be seen in Table 1, it allows us to collect statistically significant data in terms of the student's academic level, gender, and degree of familiarity with each other. This allowed us to validate the effectiveness of the system both individually, and collaboratively. The higher number of participants allowed us to run the collaborative experiment with a larger group. For each session, the instructor explained and demonstrated the experiment to the group. The students were sent the experiment sheet in advance that lists all the steps for a particular test. The instructor then asked each student to individually perform one of the four experiments. In the second phase, the instructor asked the group to perform a different experiment

Table 1 Groups formation and distribution

Academic level	Group formation	Gender distribution		
		Female	Male	Mixed
First year	Randomly assigned	Female	Male	Mixed
	Self-enrolled	Female	Male	Mixed
Second year	Randomly assigned	Female	Male	Mixed
	Self-enrolled	Female	Male	Mixed

collaboratively. Students were encouraged to engage and collaborate during the group experiment.

For the quantitative analysis, the session times for individual and group sessions were recorded. One passive observer joined every group session without any interference to log the number of collaboration instances. Collaboration was identified if students engaged with each other verbally to discuss the experiment or helped their fellow students. In some cases, students interacted with the system on behalf of their fellows. Following the conclusion of each session, students were asked to fill in two questionnaires. The first questionnaire comprising 12 questions was used to evaluate the usability of the app and its impact on learning. The second questionnaire comprising 4 questions was used to evaluate the app's suitability for collaboration. Students also provided open-ended feedback about the UX and other issues, which was used for the qualitative analysis. The flow of the experiment can be seen in Fig. 4. The survey questions, along with all the quantitative and qualitative results are presented in the next section.

Results and discussion

As explained in the previous section, we collected the usage data for the app for both individual and collaborative use. The individual data comprises the time taken to complete the experiment. We analyzed the overall data from multiple viewpoints in terms of students' academic level and gender. This allowed us to find out if there is any parameter that has a significant impact on the usage. In general, it is a mobile-based application, and all students are familiar with the system, and our null hypothesis was that there should not be any significant difference in terms of conducting the experiment with respect to their gender and academic level in both individual and collaborative sessions. For the collaborative sessions, we also evaluated the group enrollment scheme and group composition. Once again, the null hypothesis states that none of these factors should have any impact on participants' performance.

Individual session results

As explained in the user study, the individual sessions were comprised of each participant performing one of the experiments individually. The time taken to complete the experiment was automatically logged. The students are classified as per their gender and academic level. To evaluate the statistically significant results we define null hypotheses that students' gender and academic level have no influence on their experiment completion time. In principle, we can use a t-test when comparing two groups, and single-factor Analysis of Variance (ANOVA) can be employed when comparing three or more groups. We employed single-factor ANOVA for every analysis for the sake of a consistent presentation, as single-factor ANOVA is

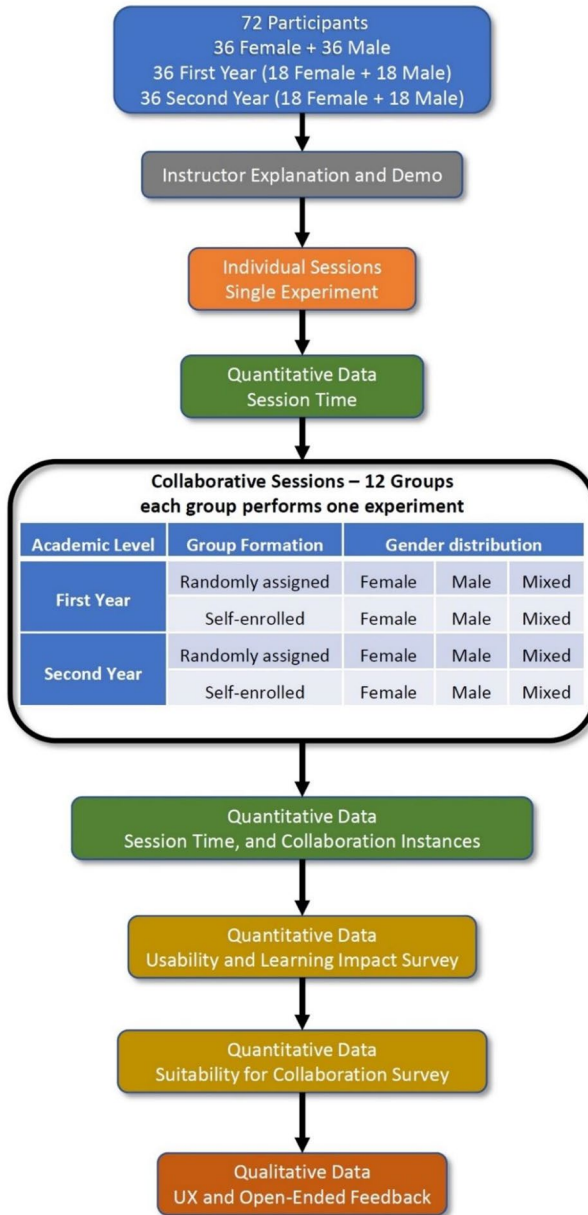


Fig. 4 Experiment flow showing all the steps from individual and collaborative sessions to the final data collection

statistically equivalent to a t test when performed on two groups, and as the t -test is a special case of ANOVA the computed p -value is identical from both tests. Please note that we used $\alpha = 0.05$ for ANOVA, and the null hypothesis is refuted if the p -value is less than α .

Table 2 Single-factor ANOVA individual sessions (academic level)

Variable	Groups	<i>N</i>	Mean	SD	<i>F</i> (1, 70)	η^2
Academic level	First year	36	359.31	25.53	0.12	0.002
	Second year	36	357.39	22.27		

Table 3 Single-factor ANOVA individual sessions (gender)

Variable	Groups	<i>N</i>	Mean	SD	<i>F</i> (1, 70)	η^2
Gender	Female	36	355.56	25.27	0.99	0.014
	Male	36	361.14	22.24		

Individual session—academic level

We had 72 participants in total divided into two groups of 36 students each of first and second-year students. Individual session times were analyzed using the single-factor ANOVA. The results can be seen in Table 2.

As can be seen in Table 2, there is no statistically significant effect of academic level on participants' experiment completion time, $F(1, 70)=0.12$, and $p=0.74$. Therefore, we cannot refute the null hypothesis. It shows that both academic levels performed at the same level of efficiency and productivity.

Individual session results—gender

The 72 participants comprised 36 female and 36 male students. This allowed us to analyze the data from the gender point of view using single-factor ANOVA. The results can be seen in Table 3.

As can be seen in Table 3, there is no significant effect of gender on participants' experiment completion time, $F(1, 70)=0.99$, and $p=0.32$. Therefore, we cannot refute the null hypothesis. It shows that both genders performed at the same level of efficiency and productivity.

Individual session results—gender and academic level

The academic level and gender were not statistically significant factors when considered separately. We also wanted to verify if there is any correlation between the academic level and gender at a finer scale. Both 36 male and 36 female participants are further subdivided into two groups of 18 participants each of first and second-year academic level. We analyzed the quantitative data for each gender with respect to the academic level. The results of male and female participants with respect to academic levels can be seen in Table 4.

As can be seen in Table 4, there is no significant effect of academic level within each gender on participants' individual experiment completion time, $F(3,68)=0.39$, and $p=0.76$. Therefore, we cannot refute the null hypothesis. It shows even within

Table 4 Single-factor ANOVA individual sessions (gender)

Variables	Groups	<i>N</i>	Mean	SD	<i>F</i> (3, 68)	η^2
Gender and academic level	First year (female)	18	355.67	26.03	0.39	0.017
	Second year (female)	18	355.44	25.25		
	First year (male)	18	362.94	25.22		
	Second year (male)	18	359.33	19.38		

both genders, participants at both academic levels performed at the same level of efficiency and productivity.

Individual session results—summary

In summary, after analyzing individual results from academic level, gender, and both gender and academic level viewpoints, it can be stated that none of the factors had any influence on the efficiency and productivity of participants. The null hypothesis that academic level and gender have no impact on individual completion time could not be refuted. In contrast to earlier works on gender differences when it comes to technology (Punter et al., 2017; Reychav & McHaney, 2017; Tondeur et al., 2016; Zhou & Xu, 2007), our study did not identify any statistically significant difference between the two genders. As identified by (Reychav & McHaney, 2017), the gender differences in adopting mobile technologies are already minimal. Given the familiarity of students with smartphone interfaces and using various AR applications, our results are not unexpected, particularly so, as such familiarity elicit the basic construct that forms the desired flow in e-learning environments (Rodríguez-Ardura & Meseguer-Artola, 2021).

Collaborative session results

In contrast to the individual sessions, where no significant difference is observed in terms of any parameters, we observed some significant differences in the group-based collaborative sessions data for various parameters. It is to be noted that collaborative sessions are analyzed not only in terms of gender, and academic level, but also in relation to random and self-enrolled group formation. Once again, our null hypothesis states that none of the factors should have any impact on the participant's performance and collaboration in the groups. As explained in the user study section, in addition to the experiment completion time, we manually logged the number of interactions between students during the experiment to quantify the collaborations. The interactions are classified as when a participant helped the other participant either verbally or performed actions on their behalf. This had an overall impact on how quickly students complete an experiment in a collaborative session. Thus, the influence of each factor on both the completion time and collaboration is analyzed. We present the results for each factor and its combinations in the following sections.

Table 5 Single-factor ANOVA collaborative sessions completion time (academic level)

Variable	Groups	<i>N</i>	Mean	SD	<i>F</i> (1, 10)	η^2
Academic level	First year	6	331.00	8.74	0.26	0.026
	Second year	6	328.50	8.12		

Table 6 Single-factor ANOVA collaborative sessions interactions (academic level)

Variable	Groups	<i>N</i>	Mean	SD	<i>F</i> (1, 10)	η^2
Academic level	First year	6	12.33	6.95	0.13	0.012
	Second year	6	13.67	6.02		

Collaborative sessions—academic level

In the first analysis, the student's performance in a collaborative session was evaluated from the point of view of their academic level. As explained in the User Study section, the students were divided into 12 groups, where six groups comprised first-year students, and six groups comprised second-year students. The collected quantitative data includes the groups' experiment completion time and the number of interactions. We analyzed this quantitative data from the point of view of academic levels. The results for the collaborative sessions with respect to their academic level in terms of the completion time can be seen in Table 5, whereas the results for the number of interactions can be seen in Table 6.

As can be seen in Table 5, there is no significant effect of academic level on participants' collaborative experiment completion time, $F(1, 10)=0.26$, and $p=0.61$. Therefore, we cannot refute the null hypothesis. It shows that participants at both academic levels performed at the same level of efficiency and productivity in a collaborative session.

As can be seen in Table 6, there is no significant effect of academic level on participants' interaction during a collaborative session, $F(1, 10)=0.13$, $p=0.73$. Therefore, we cannot refute the null hypothesis. It shows that participants at both academic levels interacted with each other at a similar level.

Collaborative sessions—collaborative group formation

In the first analysis, the student's performance in a collaborative session was evaluated from the point of view of collaborative group formation. The students were divided into 12 groups, where 6 groups were randomly formed (3 from first-year students, and 3 from second-year students), and 6 groups were formed via self-enrollment (3 from first-year students, and 3 from second-year students). The collected quantitative data includes the groups' experiment completion time and the number of interactions. We analyzed this quantitative data from the point of view of group formation. The results for the collaborative sessions with respect to their collaborative group formation in terms of the completion time can be seen in Table 7, whereas the results for the number of interactions can be seen in Table 8.

Table 7 Single-factor ANOVA collaborative sessions completion time (collaborative group formation)

Variable	Groups	<i>N</i>	Mean	SD	<i>F</i> (1, 10)	η^2
Group formation	Random	6	336.67	4.55	36.76	0.786
	Self-enrolled	6	322.83	3.25		

Table 8 Single-factor ANOVA collaborative sessions interactions (collaborative group formation)

Variable	Groups	<i>N</i>	Mean	SD	<i>F</i> (1, 10)	η^2
Group formation	Random	6	7.33	1.97	90.31	0.900
	Self-enrolled	6	18.67	2.16		

Table 9 Single-factor ANOVA collaborative sessions completion time (academic level and collaborative group formation)

Variables	Groups	<i>N</i>	Mean	SD	<i>F</i> (3, 8)	η^2
Academic level and group formation	Random (first year)	3	338.33	3.51	11.72	0.815
	Random (second year)	3	335.00	5.57		
	Self-enrolled (first year)	3	323.67	4.16		
	Self-enrolled (second year)	3	322.00	2.65		

As can be seen in Table 7, there is a significant effect of collaborative group formation on the participants' experiment completion time, $F(1, 10) = 36.76$, and $p = 1.21E-04$. Therefore, the null hypothesis is refuted, and the alternative hypothesis is supported. The results show that students performed significantly better in a self-enrolled group where they were familiar with their group members.

As can be seen in Table 8, there is a significant effect of collaborative group formation on participants' interaction during a collaborative session, $F(1, 10) = 90.31$, and $p = 2.53E-06$. Therefore, the null hypothesis is refuted, and the alternative hypothesis is supported. The results show that students showed a very high level of interaction in a self-enrolled group where they were familiar with their group members.

Collaborative sessions—academic level and collaborative group formation

The academic level analysis showed that academic level is not a statistically significant factor when it comes to completion time and the number of interactions in a collaborative session. On the other hand, collaborative group formation has a statistically significant impact on both metrics. Thus, we analyzed the quantitative data both in terms of academic levels and group formations. This allows for a fine-grain analysis of all possible combinations of both factors. The results for the collaborative sessions with respect to both academic levels and collaborative group formations in terms of the completion time can be seen in Table 9, whereas the results for the number of interactions can be seen in Table 10.

Table 10 Single-factor ANOVA collaborative sessions interactions (academic level and collaborative group formation)

Variables	Groups	<i>N</i>	Mean	SD	<i>F</i> (3, 8)	η^2
Academic level and group formation	Random (first year)	3	6.67	1.15	29.72	0.918
	Random (second year)	3	8.33	2.08		
	Self-enrolled (first year)	3	18.33	3.21		
	Self-enrolled (second year)	3	19.00	1.00		

As can be seen in Table 9, there is a significant effect of some combination of academic level and collaborative group formation on participants' interaction during a collaborative session, $F(3, 8) = 11.72$, and $p = 2.68E-03$. Therefore, the null hypothesis is refuted, and the alternative hypothesis is supported. As we are now comparing more than two groups, Tukey's HSD test is employed to identify the combination of factors that have a statistically significant difference.

Tukey's HSD test compares the absolute difference between the means of each combination of factors, and the difference is classified as statistically significant if the difference is greater than the HSD value. Tukey's HSD test for multiple comparisons found that mean value of completion time between Random (First Year) and Random (Second Year) do not have a statistically significant difference. Similarly, there is no statistically significant difference between the mean value of completion time between Self-Enrolled (First Year) and Self-Enrolled (Second Year). In contrast, all other combinations of academic level and group formation have a statistically significant difference between their mean value of completion time. The results show that self-enrolled groups at both academic levels perform at a similar level of higher efficiency whereas random groups perform at statistically significantly lower levels of efficiency. The collaborative interaction analysis at the same fine-grain level confirms this result as shown in Table 10.

As can be seen in Table 10, there is a significant effect of academic level and collaborative group formation on participants' interaction during a collaborative session, $F(3,8) = 29.72$, and $p = 1.09E-04$. Therefore, the null hypothesis is refuted, and the alternative hypothesis is supported. As we are now comparing more than two groups, Tukey's HSD test is employed to identify the combination of factors that have a statistically significant difference.

Tukey's HSD test for multiple comparisons found that mean value of interactions between Random (First Year) and Random (Second Year) do not have a statistically significant difference. Similarly, there is no statistically significant difference between the mean value of interactions between Self-Enrolled (First Year) and Self-Enrolled (Second Year). In contrast both self-enrolled groups at the first and second-year levels have a significantly higher number of interactions with no statistically significant difference between their number of interactions. In contrast, all other combinations of academic level and group formation have a statistically significant difference between their mean value of interactions. The results show that group formation is the single most significant factor when it comes to achieving a higher level of efficiency and productivity in a collaborative session. We believe this is due to

Table 11 Single-factor ANOVA collaborative sessions completion time (collaborative group gender composition)

Variables	Groups	<i>N</i>	Mean	SD	<i>F</i> (2, 9)	η^2
Group composition	Female	4	331.50	11.27	0.18	0.039
	Male	4	327.75	6.18		
	Mixed	4	330.00	8.29		

Table 12 Single-factor ANOVA collaborative sessions interactions (collaborative group gender composition)

Variables	Groups	<i>N</i>	Mean	SD	<i>F</i> (2, 9)	η^2
Group composition	Female	4	14.00	7.87	0.09	0.019
	Male	4	13.00	6.58		
	Mixed	4	12.00	5.89		

students' familiarity with their peers in the group, as it contributes to mediating the influence of gender and academic levels and builds a shared goal, which in turn unifies the group flow experience (Zhou et al., 2021).

Collaborative sessions—gender

The quantitative data was also analyzed from the viewpoint of genders. As there are 12 groups, and as shown in Table 1, 4 groups comprising female students, 4 groups comprising male students, and 4 evenly mixed groups of female and male students. The results for the collaborative sessions with respect to their collaborative groups' gender composition in terms of the completion time can be seen in Table 11, whereas the results for the number of interactions can be seen in Table 12.

As can be seen in Table 11, there is no significant effect of gender on participants' collaborative experiment completion time, $F(2, 9)=0.18$, $p=0.84$. Therefore, we cannot refute the null hypothesis. It shows that all collaborative groups irrespective of their gender composition performed at the same level of efficiency and productivity in a collaborative session.

As can be seen in Table 12, there is no significant effect of academic level on participants' interaction during a collaborative session, $F(2, 9)=0.09$, and $p=0.92$. Therefore, we cannot refute the null hypothesis. It shows that the level of interaction was statistically similar in all groups when analyzed from the point of view of genders.

Collaborative sessions—summary

The collaborative session results clearly follow a similar pattern to the individual session results where both the academic level and gender did not result in any statistically significant impact on participants' collaborative performance. The single significant factor for collaborative sessions was group formation. These results are partially corroborated by the earlier research into students' collaboration studies (López-Faican & Jaen, 2020; Zhou et al., 2021). We mitigated the possible

shortcomings of Zhou et al. (2021) by providing clear instructions before the experiments. Also, the findings by Zhou et al. (2021) show that students regulate their behaviors in group discussions based on the composition of groups. Our results clearly show that students in self-enrolled groups had a higher level of interaction, and consequently, a higher level of efficiency and productivity. This shows that a higher degree of familiarity results in a higher level of collaboration in a virtual experiment conducted in an augmented reality educational environment. In contrast to López-Faicán and Jaen (2020) our results are much more fine-grained with a deeper collaborative analysis from multiple viewpoints.

Usability, impact on learning and collaboration surveys

In addition to the quantitative data obtained through the system logs, and direct observations, we also conducted two surveys to evaluate the usability, impact on learning, and effectiveness of collaboration. While an AR-based learning system may attract positive feedback due to its novelty rather than its effectiveness in performing its core task of improving the overall learning experience. Nonetheless, it is equally important to get an understanding of the new system which can replace face-to-face learning, and how well it prepares students if they need to perform similar tasks in a real-world environment. Our survey comprising 12 questions can be seen below in Table 13. The students answered each question on a five-level Likert scale. The results of the survey can be seen in Fig. 5.

It can be seen in Fig. 5 that other than Q#6 and Q#9 the majority of students reacted very positively to the app and its learning impact. The only two points where students differed in opinion were Q#6 which asked if the virtual lab can completely replace face-to-face lab, and most of the students felt that it should not be used as a

Table 13 Usability and impact on learning survey questions

Question
1. The instructions to perform the experiment using the app were clear
2. I was able to recognize all the required equipment in the app
3. It was easy to perform all the steps of the experiment using the app
4. I could clearly observe the results of each step of the experiment
5. I understood the concept of qualitative analysis of anions after the experiment
6. The virtual lab can completely replace face-to-face lab
7. The virtual lab would be more suitable as a companion to the face-to-face lab
8. Compared to a face-to-face lab, using the virtual equipment is easier
9. Using the virtual lab prepares me in handling the equipment in the face-to-face lab
10. Performing the experiment in virtual lab better prepares me for the same experiment in the face-to-face lab
11. The virtual lab taught me all the concepts of the experiment same as I would expect from a face-to-face lab
12. I believe that performing the experiment in the virtual has contributed positively towards my learning

Usability and Impact on Learning Survey

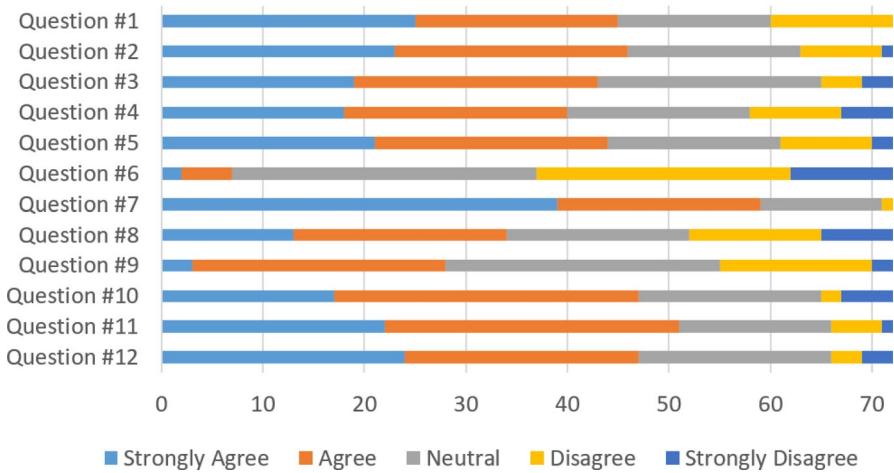


Fig. 5 Usability and Impact on Learning survey results showing positive results for both metrics

complete replacement, hence, the qualitative survey showed that a real-world face-to-face experience is still a preferred mode of learning for most students. On the other hand, an overwhelmingly positive response to Q#7 clearly demonstrates that students would like to see the remote learning experience as a compliment to face-to-face learning because in cases where a student cannot attend in person, a viable collaborative solution to perform the lab is still available. Additionally, the response to Q#9 shows that handling virtual objects is not a substitute for handling the same objects in a physical space. This is a valid observation, though the response to Q#2, Q#8, and Q#10 clearly shows that the virtual lab has its merit and eliminates sources of frustrations of the face-to-face labs, and students learn about the equipment and their use in a safer setting, which helps them to better prepare for a face-to-face environment. Finally, the responses from Q#11 and Q#12 clearly show that in terms of learning, students found the virtual lab to provide them with a similar level of a learning experience as a face-to-face lab, and they have achieved the same level of understanding as they would have from the face-to-face learning. In addition, the majority of students found the virtual lab a positive learning experience with a meaningful impact on their learning.

Furthermore, we surveyed all participants to gauge the effectiveness of collaboration while using the application. The survey comprised four questions rated over a 5-level Likert scale. The results of the survey can be seen in Fig. 6, where most of the responses were extremely positive despite some concerns about the availability of collaboration tools and if the system truly encourages collaboration. These will be discussed in the limitations section.

The findings of the surveys are in agreement with the recent studies that show the benefits of incorporating collaborative learning or AR in a learning environment (Fidan & Tuncel, 2019; Garcia-Sanjuan et al., 2018; Garzón & Acevedo, 2019;

Collaboration Survey

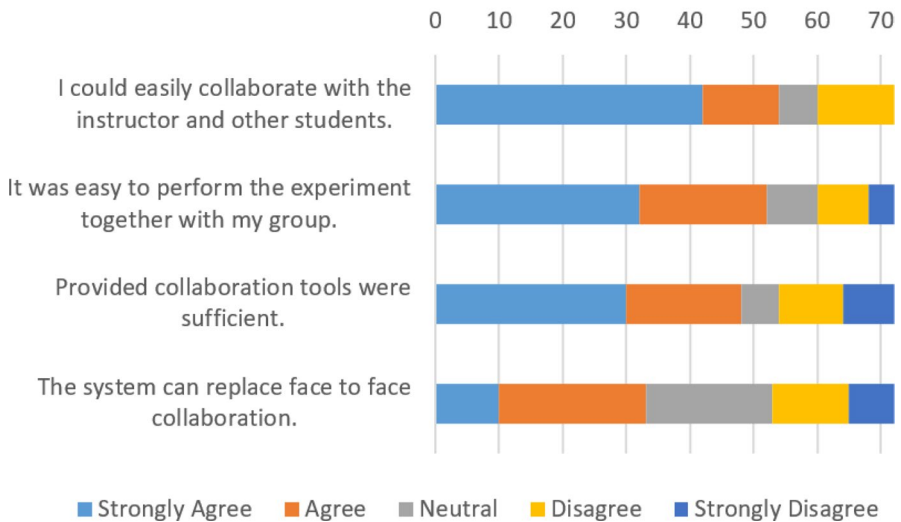


Fig. 6 Collaboration survey results showing a very high level of satisfaction

López-Faicán & Jaen, 2020). Our results show that an interactive AR learning environment can be used as an effective tool for both individual and collaborative learning. In contrast to the previous methods, our work has a very complex collaborative AR lab-based experiment scenario. Our user study is performed on multiple levels, and both individual and group-based sessions are analyzed. We have performed a statistical data analysis from multiple viewpoints and also compared AR-based remote learning to face-to-face learning and analyzed its impact on learning.

Limitations and future research

Some of the limitations discussed earlier are directly related to the application, i.e., lack of video chat, and overly cumbersome manipulation of test tubes using the picker. The other limitation is the deployment of the app on only one mobile device. In recent years, the advances in AR have resulted in the availability of several wearable smart glasses, e.g., Magic Leap, Nreal Air. A more natural deployment solution for the proposed app would be to use smart glasses as it will allow natural user interaction using gestures. The user study can also be expanded to compare mobile-based with glasses-based AR. In addition, the app can also be deployed as a complete virtual solution for a virtual reality environment. This will also be explored in future work across different AR/VR platforms.

In terms of the provided collaboration tools, we found through the open-ended feedback that participants also expect to have audio and video communication features. We did not provide a video communication feature as it felt counterintuitive for an AR system because it would take up important screen space. In addition, we anticipate that a video screen overlay would increase the information overload and

reduce students' focus, at least within the proposed context provided in this work. We plan to explore this feature in future versions of the app. Additionally, the response to the final question clearly shows that similar to the impact on the learning survey, participants do not feel that the system can completely replace a face-to-face system. This is an expected result as being the general expressed preference of student to use face-to-face learning with remote learning as a complementary tool. Nevertheless, both the usage data, and survey results clearly show that even without a video communication feature that system is capable of a high level of collaboration between users, and they were able to conduct experiments effectively within a group.

Finally, we received very positive feedback regarding the UX of the system. The overall experience of joining the sessions was highlighted as positive for its lack of complexity. The negative feedback was the use and handling of the test tube. We implemented the experiment exactly as one would perform in the real lab where a test tube picker is used to handle a test tube, but for a mobile app, students expected a simple drag and drop over an intermediate input mechanism. This is a fair criticism, and we would like to have it addressed in future work.

Our user study can also be improved by implementing additional data logging to acquire additional quantitative data regarding the app's usage. So far there is no analysis of the error rate that can provide a very precise measure of the ease-of-use parameter. Additionally, the study is limited to first and second-year students, which is due to the specific nature of the lab experiment that is offered in the introductory course. For a comprehensive study, a comparative analysis of students from all academic levels should be made.

Despite some of the limitations, our work not only demonstrates the effectiveness of an AR-based collaborative learning environment as a remote learning tool and its impact on learning. The study also provides insightful details on eclectic properties and features to be considered while developing applications for similar contexts. A new initiative is already being drafted to add new features to the system and implement collaborative learning in different areas of science. We also plan to implement the system in VR and design a collaborative AR and VR environment for a generalized XR-based learning environment.

Conclusion

This work presented "LeARn," a collaborative AR educational environment to replace face-to-face learning. We demonstrated this system using a virtual chemistry lab experiment. The system allows real-time synchronous interaction over the network, and voice communication to create a collaborative environment. We performed an extensive user study to evaluate the effectiveness and learning impact of the proposed system. We analyzed the usage of a virtual lab in terms of its individual and collaborative use. The individual use was analyzed in terms of the participant's gender and academic level. For individual sessions, the quantitative data in the form of experiment duration was recorded. The collaborative effectiveness was evaluated in terms of group work comprising female, male, and mixed-gender

groups. The groups were also organized with respect to the academic level of students, along with their random or self-enrolled composition. For the group sessions, the quantitative data in the form of experiment duration and the number of collaborative instances were recorded. In addition, a comprehensive survey was conducted to evaluate the application's usability, impact on learning, and collaboration support.

We employed single-factor ANOVA to identify statistically significant differences between various parameters for both individual and collaborative sessions. The individual results were analyzed from multiple viewpoints. The results did not show any statistically significant difference from the point of view of genders or academic levels. A deeper analysis of each gender at different academic levels also did not show any statistically significant difference. The collaborative session results follow a similar pattern noted in the individual session results where both the academic level and gender did not result in any statistically significant impact on the participant's collaborative performance. The single significant factor for collaborative sessions was group formation. Students in self-enrolled groups had a higher level of interaction, and consequently a higher level of efficiency and productivity. This shows that a higher degree of familiarity results in a higher level of collaboration in a virtual experiment conducted in an augmented reality educational environment.

We also conducted multiple surveys to evaluate the usability, UX, collaborative support, and learning impact of the application. The survey results show that the system has a very high level of usability and supports collaboration. The UX survey was overall positive but highlighted two issues related to the manipulation of test tubes and the lack of video chat. Finally, the learning impact survey clearly demonstrated the effectiveness of the application as a learning environment and justifies its usage as a complimentary tool together with face-to-face learning. In the future, we would like to deploy the virtual lab on different types of AR devices. We would also like to implement the same experiment in a VR environment and compare results of both. Having the app deployed across different platform and devices, will allow further evaluation to the important insights gathered from this study to further advance users' efficiency and productivity within the new paradigms.

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Declarations

Conflict of interest The authors declare no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Ethical statement We hereby declare that this manuscript is the result of our independent creation under the reviewers' comments. This manuscript does not contain any research achievements that have been published or written by other individuals or groups. We are the only authors of this manuscript. The legal responsibility of this statement shall be borne by us.

References

- Agarwal, C., & Thakur, N. (2014). The evolution and future scope of augmented reality. *International Journal of Computer Science Issues (IJCSI)*, 11(6), 59.
- Akçayır, M., & Akçayır, G. (2017). Advantages and challenges associated with augmented reality for education: A systematic review of the literature. *Educational Research Review*, 20, 1–11. <https://doi.org/10.1016/j.edurev.2016.11.002>
- Azuma, R. T. (1997). A survey of augmented reality. *Presence: Teleoperators and Virtual Environments*, 6(4), 355–385. <https://doi.org/10.1162/pres.1997.6.4.355>
- Csikszentmihalyi, M., & Csikszentmihalyi, I. (1992). *Optimal experience: Psychological studies of flow in consciousness*. Cambridge University Press.
- Fidan, M., & Tuncel, M. (2019). Integrating augmented reality into problem based learning: The effects on learning achievement and attitude in physics education. *Computers & Education*, 142, 103635. <https://doi.org/10.1016/j.compedu.2019.103635>
- García-Sanjuan, F., Jurdi, S., Jaen, J., & Nacher, V. (2018). Evaluating a tactile and a tangible multi-tablet gamified quiz system for collaborative learning in primary education. *Computers & Education*, 123, 65–84. <https://doi.org/10.1016/j.compedu.2018.04.011>
- Garzón, J., & Acevedo, J. (2019). Meta-analysis of the impact of augmented reality on students' learning gains. *Educational Research Review*, 27, 244–260. <https://doi.org/10.1016/j.edurev.2019.04.001>
- Harsh, J. A., Maltese, A. V., & Tai, R. H. (2012). A perspective of gender differences in chemistry and physics undergraduate research experiences. *Journal of Chemical Education*, 89(11), 1364–1370. <https://doi.org/10.1021/ed200581m>
- Ibáñez, M.-B., & Delgado-Kloos, C. (2018). Augmented reality for STEM learning: A systematic review. *Computers & Education*, 123, 109–123. <https://doi.org/10.1016/j.compedu.2018.05.002>
- Khan, T., Johnston, K., & Ophoff, J. (2019). The impact of an augmented reality application on learning motivation of students. *Advances in Human-Computer Interaction*, 2019, 7208494. <https://doi.org/10.1155/2019/7208494>
- Laal, M., & Ghodsi, S. M. (2012). Benefits of collaborative learning. *Procedia—Social and Behavioral Sciences*, 31, 486–490. <https://doi.org/10.1016/j.sbspro.2011.12.091>
- López-Faicán, L., & Jaen, J. (2020). EmoFindAR: Evaluation of a mobile multiplayer augmented reality game for primary school children. *Computers & Education*, 149, 103814. <https://doi.org/10.1016/j.compedu.2020.103814>
- Lucas, M., Nelson, J., & Sims, D. (2020). *Schools' responses to COVID-19: Pupil engagement in remote learning*. National Foundation for Educational Research.
- Mota, J. M., Ruiz-Rube, I., Doderó, J. M., & Arnedillo-Sánchez, I. (2018). Augmented reality mobile app development for all. *Computers & Electrical Engineering*, 65, 250–260. <https://doi.org/10.1016/j.compeleceng.2017.08.025>
- Mota, J. M., Ruiz-Rube, I., Doderó, J. M., & Figueiredo, M. (2016). Visual environment for designing interactive learning scenarios with augmented reality. In *12th International conference on mobile learning 2016*. International Association for Development of the Information Society.
- Ottemo, A., Berge, M., & Silfver, E. (2020). Contextualizing technology: Between gender pluralization and class reproduction. *Science Education*, 104(4), 693–713. <https://doi.org/10.1002/sce.21576>
- Pantelidis, V. S. (2010). Reasons to use virtual reality in education and training courses and a model to determine when to use virtual reality. *Themes in Science and Technology Education*, 2(1–2), 59–70.
- Punter, R. A., Meelissen, M. R., & Glas, C. A. (2017). Gender differences in computer and information literacy: An exploration of the performances of girls and boys in ICILS 2013. *European Educational Research Journal*, 16(6), 762–780. <https://doi.org/10.1177/1474904116672468>
- Reychav, I., & McHaney, R. (2017). The relationship between gender and mobile technology use in collaborative learning settings: An empirical investigation. *Computers & Education*, 113, 61–74. <https://doi.org/10.1016/j.compedu.2017.05.005>
- Rodrigues, M., Franco, M., & Silva, R. (2020). COVID-19 and disruption in management and education academics: Bibliometric mapping and analysis. *Sustainability*. <https://doi.org/10.3390/su12187362>
- Rodríguez-Ardura, I., & Meseguer-Artola, A. (2021). Flow experiences in personalised e-learning environments and the role of gender and academic performance. *Interactive Learning Environments*, 29(1), 59–82. <https://doi.org/10.1080/10494820.2019.1572628>

- Roussos, M., Johnson, A., Moher, T., Leigh, J., Vasilakis, C., & Barnes, C. (1999). Learning and building together in an immersive virtual world. *Presence: Teleoperators & Virtual Environments*, 8(3), 247–263. <https://doi.org/10.1162/105474699566215>
- Tamim, R. M., Bernard, R. M., Borokhovski, E., Abrami, P. C., & Schmid, R. F. (2011). What forty years of research says about the impact of technology on learning: A second-order meta-analysis and validation study. *Review of Educational Research*, 81(1), 4–28. <https://doi.org/10.3102/0034654310393361>
- Tondeur, J., Van de Velde, S., Vermeersch, H., & Van Houtte, M. (2016). Gender differences in the ICT profile of university students: A quantitative analysis. *DiGeSt: Journal of Diversity and Gender Studies*, 3(1), 57. <https://doi.org/10.11116/jdivegendstud.3.1.0057>
- Unity. (2022). Unity: The leading platform for creating interactive, real-time content. <https://unity.com/>
- UnityAR. (2022). Unity AR Foundation. <https://unity.com/unity/features/arfoundation>
- UnityPUN. (2022). Photon Unity Networking. <https://www.photonengine.com/pun>
- UnityVoice. (2022). Photon Unity Voice Chat. <https://www.photonengine.com/voice>
- UnityXR. (2022). Unity XR Interaction Toolkit. <https://shorturl.at/ovx13>
- Walker, Z., McMahon, D. D., Rosenblatt, K., & Arner, T. (2017). Beyond Pokémon: Augmented reality is a universal design for learning tool. *SAGE Open*, 7(4), 2158244017737815. <https://doi.org/10.1177/2158244017737815>
- Winn, W. (1993). A conceptual basis for educational applications of virtual reality. Technical Publication R-93-9. Human Interface Technology Laboratory of the Washington Technology Center, University of Washington.
- Zhou, G., & Xu, J. (2007). Adoption of educational technology: How does gender matter? *The International Journal of Teaching and Learning in Higher Education*, 19, 140–153.
- Zhou, Q., Suraworachet, W., Pozdniakov, S., Martinez-Maldonado, R., Bartindale, T., Chen, P., Richardson, D., & Cukurova, M. (2021). Investigating students' experiences with collaboration analytics for remote group meetings. In *Artificial intelligence in education: 22nd international conference, AIED 2021*, Utrecht, The Netherlands, 14–18 June 2021, Proceedings, Part I (pp. 472–485). https://doi.org/10.1007/978-3-030-78292-4_38

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Naveed Ahmed is an associate professor at the Department of Computer Science, University of Sharjah. He received his Ph.D. in computer science from the University of Saarland (Max-Planck-Institute for Informatics), Germany, in 2009. He worked as a research and development engineer at Autodesk in Cambridge, United Kingdom, for 2 years. He is currently working as an associate professor at the Department of Computer Science, University of Sharjah. His research interests include Human Computer Interaction, Extended Reality, 3D animation, and multi-view video-based modeling and rendering.

Mohammed Lataifeh is an assistant professor at the Department of Computer Science, University of Sharjah. He received his Ph.D. in design and Information Technology from the De Montfort University (Institute of Creative Technologies), UK, in 2015. He worked for several institutions as solution consultant (ERP systems, E-commerce, Graphics and Multimedia). He is currently working as an assistant professor at the Department of Computer Science, University of Sharjah. His research interests include enterprise systems, graphics, and HCI within mixed reality environments.