Annals of the Academy of Romanian Scientists
Series on Science and Technology of Information
ISSN 2066 - 8562 Volume **4**, Number **1**/**2011**

USABILITY OF AN AUGMENTED REALITY LEARNING SCENARIO: A MIXED METHODS EVALUATION APPROACH

Costin PRIBEANU[1], Dragos Daniel IORDACHE[2]

Abstract. *Augmented reality (AR) technologies provide teachers with new opportunities to enhance students' motivation to learn. The mix of real and virtual objects requires specific interaction techniques thus making the design for usability a difficult task. Individual usability evaluation methods have specific strengths and weaknesses that suggest a mixed methods research approach. This paper presents a triangulation of quantitative and qualitative data that increases confidence in the evaluation results, provides a broader view on the strengths and weaknesses of the learning scenario and enables a deeper understanding of usability problems.*

Keywords: usability evaluation, mixed methods research, triangulation**,** augmented reality, learning scenario, educational platform

1. Introduction

The mix of real and virtual objects is challenging the developers of Augmented Reality (AR) systems to design novel interaction techniques which are mainly driven by the possibilities to interact with real objects and augment them with useful information. Educational systems based on the augmented reality (AR) technology are creating a new kind of user learning experience by bringing real life objects into a computer environment which in turn could better support a learning-by-doing approach to education [1, 3, 12, 17]. However, designing for usability is not an easy task in the AR field given the particularities of interaction techniques and the lack of specific user-centered design methods [9, 22].

This paper is reporting on a mixed methods research approach to the usability evaluation of an AR-based learning scenario developed in the framework of the ARiSE (Augmented Reality for School Environments) research project. Six research partners (Fraunhofer IAIS, Technical University Prague, Siauliai University, ICI Bucharest, Brighton University, and Across Limited) and two school partners (one for Germany and another from Lithuania) participated in this FP6 (Framework Programme 6) project. The ICT research in FP6 focused on generating new technologies integrated into day-to-day life with ease-to-use

[1]Senior Researcher I, PhD, National Institute for Research and Development in Informatics, pribeanu@ici.ro.
[2]Researcher, PhD student, National Institute for Research and Development in Informatics, iordache@ici.ro.

human-computer interfaces. In this respect, augmented reality proved to be useful for guiding and teaching situations. Also, the usability of AR systems is a key feature for the successful integration of AR-based applications in primary and secondary schools.

The main objective of the ARiSE project was to test the pedagogical effectiveness of introducing AR in schools and creating remote collaboration between classes around AR display systems. ARiSE developed a new technology, the Augmented Reality Teaching Platform (ARTP) in three stages. In each stage a research prototype (software application) was developed that implemented a learning scenario based on a different interaction paradigm.

The first prototype implemented a Biology learning scenario for secondary schools. The interaction paradigm is "3D process visualization" and was targeted at enhancing the students" understanding and motivation to learn the human digestive system. The second prototype implemented a Chemistry scenario. The interaction paradigm is "building with guidance" and was targeted at understanding the periodic table of Chemical elements, the structure of atoms / molecules, and chemical reactions. The third prototype developed an application for content management. The interaction paradigm was "telepresence and remote collaboration" and aimed at teaching students how to create new content and present it in a collaborative framework.

In order to get a fast feedback from both teachers and students, each prototype was tested with users during the ARiSE Summer School which was held yearly. All project partners participated at the summer school. Two groups of students and two teachers from German and Lithuanian partner schools and three groups of students accompanied by a total of 4 teachers from 3 general (basic) schools from the host partner were invited.

A first version of the Biology scenario was developed in 2006 and proved to be unsatisfactory with respect to the design of interaction techniques [18]. An improved version was tested again in 2007. The usability results were useful and revealed several strengths and weaknesses of the implemented scenario [4, 13, and 20]. The results were interesting but the confidence in our findings was low due to the small number of users and the special context created by the summer school. Therefore we repeated the user testing on the final version of the prototype in 2008 with a larger number of more representative users.

This paper aims at presenting a mixed methods research approach to the usability evaluation of the Biology application based on a closer integration of quantitative and qualitative data and evaluation techniques. The purpose is to get a broader view on usability aspects including the relationship between various factors that are influencing the acceptance of this new technology, such as ergonomics of specific AR devices, ease of use, usefulness for the learning process and students" motivation to learn.

The rest of this paper is organized as follows. In the next section we will briefly discuss some usability evaluation issues. In section 3 we will present the evaluation method and procedure. Next, we will present and discuss the evaluation results. The paper ends with conclusion in section 5.

2. Usability evaluation

ISO standard 9126:2001 defines usability as the capability of a software product to be understood, learned, used and attractive to the user, when used under specified conditions [14]. Over the last decade, usability evaluation concerns expanded in the area of user experience studies in order to better capture various aspects that are related to the hedonic aspects of interacting with computers [10].

Depending on its purpose, usability evaluation can be formative or summative [21]. Formative usability evaluation is performed in an iterative development cycle and aims at finding and fixing usability problems as early as possible [23]. The earlier these problems are identified, the less expensive is the development effort to fix them. Summative usability evaluation is carried on by testing with a relatively large number of representative users and aims at finding strengths and weaknesses as well as comparing alternative design solutions or similar systems.

There are several approaches to usability evaluation and, consequently many usability evaluation methods with various degree of effectiveness [7, 11]. Each method relies on a specific procedure to collect, process and analyze the usability data as well as on specific techniques to assess the reliability and validity of results [9, 13]. Despite the efforts done to improve the effectiveness of these methods, each one has specific weaknesses and the usability data provides with a limited view on the target system. The trend is to employ several methods and to take advantage of complementarities.

Triangulation was defined in the context of social sciences by Denzin as a combination of methodologies in the study of the same phenomenon [6]. Jick advocated that triangulation enables the researcher to get both a broader view on the unit under study and a deeper understanding of critical aspects [15].

Traditionally, approaches to evaluation are related to either to a quantitative or a qualitative paradigm. In this respect, mixed methods research is a new paradigm [5, 16] that recently gained considerable attention in various application domains such as information systems, social and behavioral sciences, and e-government. Researchers interested in mixed methods approached several aspects such as state-of-the art and challenges [8], integration of closed and open-ended items [2], or use of triangulation techniques [19].

3. Evaluation method and procedure

3.1. Participants and evaluation setting

ARTP is a "seated" AR environment: users are looking to a see-through screen where virtual images are superimposed over the perceived image of real objects placed on the table. The platform was registered by Fraunhofer IAIS under the trade mark Spinnstube® [24].

The real object is a flat torso of the human digestive system, as illustrated in Figure 1. A pointing device having a colored ball on the end of a stick and a remote controller Wii Nintendo as handler has been used as interaction tool that serves for three types of interaction: pointing on a real object, selection of a virtual object and selection of a menu item.

The user can select an organ with the pointing device. When the colored ball is onto the organ its augmentation is superimposed on the see-through screen. The user is confirming the selection by pressing the button B placed on the back of the controller. The button A of the controller was used to select a menu item.

Figure 1. Students testing the Biology scenario.

A total number of 139 students (13-14 years old), from which 65 boys and 74 girls tested the platform in 2008. All were 8th grade students enrolled in 3 general schools in Bucharest. None of them was familiar with the AR technology. The students came in groups of 7-8, accompanied by a teacher. The test was conducted on the platform of ICI Bucharest which is equipped with 4 Spinnstube® modules.

The participants have been assigned 4 tasks: a demo program explaining the absorption / decomposition process of food and three exercises: 1st exercise asking to indicate the organs of the digestive system and exercises 2 and 3, asking to indicate the nutrients absorbed / decomposed in each organ respectively the organs where a nutrient is absorbed / decomposed.

3.2. Method and procedure

A usability questionnaire was developed in this project that provides with both quantitative and qualitative measures of the educational and motivational values of a new learning scenario. The evaluation instrument has 28 closed items (quantitative measures) and 2 open questions, asking users to describe the most 3 positive and most 3 negative aspects (qualitative measures) regarding their interaction with the system.

The 28 closed items are targeting various dimensions such as ergonomics of the AR platform (5 items), perceived ease of use (10 items), perceived utility (4 items), perceived enjoyment (6 items) and intention to use (3 items).

Before testing, a brief introduction to the AR technology and ARiSE project had been done for all students. Each group of students tested ARTP in a session during 60 min.

During testing, effectiveness (binary task completion and number of errors) and efficiency (time on task) measures were collected in a log file. Measures were collected for all exercises performed. After testing, the students were asked to answer the usability questionnaire by rating the items on a 5-point Likert scale (1-strongly disagree, 2-disagree, 3-neutral, 4-agree, and 5-strongly agree). Reliability of the scale was 0.942 (Cronbach‟s alpha) which is acceptable.

4. Evaluation results

4.1. Quantitative data results

The measures of central tendency and variation for the items that are related to usability are shown in Table 1. Three items (15, 20, and 21) are general items that are measuring the overall perception of students as regarding the ease of use and enjoyment. The lowest mean values got two the items related with the clarity of visual perception (items 5 and 7) and difficulty to correct mistakes (item 13).

The items related to the ergonomics of ARTP were rated between 3.76 and 4.26. The mean value computed for this construct was 4.11. An analysis using Pearson‟s correlation indicated that there is a significant linear relationship ($R=0.543$, $p<0.001$) between observing the real object through the screen (item 5, $M=3.76$) and the overall ease of use (general item 15).

Table 1. Descriptive statistics for the Biology scenario (*N* = 139)

	Construct/Item	Mean	SD
	Ergonomics of the platform		
1.	Adjusting the "see-through" screen is easy	4.07	0.80
2.	Adjusting the stereo glasses is easy	4.26	0.79
3.	Adjusting the headphones is easy	4.24	0.91
4.	Working on the chair (work place) is comfortable	4.22	0.91
5.	Observing the real object through the screen is clear	3.76	1.07
	Perceived ease of use		
6.	Understanding how to operate with the AR application is easy	4.02	0.86
7.	The superposition between projection and the real object is clear	3.62	1.02
8.	Learning to operate with the AR application is easy	4.04	0.94
9.	Remembering how to operate with the AR application is easy	3.95	0.89
10.	Understanding the vocal explanations is easy	4.10	0.93
11.	Reading the information on the screen is easy	4.06	1.02
12.	Selecting a menu item is easy	3.95	1.10
13.	Correcting the mistakes is easy	3.79	1.04
14.	Collaborating with colleagues is easy	4.00	1.07
15.	Overall, I find the system easy to use	4.14	0.89
	Perceived enjoyment		
16.	The system makes learning more interesting	4.35	0.87
17.	Working in group with colleagues is stimulating	4.06	0.95
18.	I like interacting (move, touch, bring together) with real objects	3.91	1.05
19.	Performing the exercises is captivating	4.15	1.01
20.	Overall, I enjoy learning with the system	4.09	0.97
21.	Overall, I find the system exciting	4.13	0.89

The items related to the ease of use were rated between 3.62 and 4.14. The mean value computed for this construct was 3.97. An analysis using Pearson"s correlation indicated that there is a significant linear relationship (R=0.389, p<0.001) between item 7 (M=3.62) and the general item 15. Also, we found a significant linear relationship (R=0.358, p<0.001) between item 13 (M=3.79) and the general item 15. All these correlations highlight three usability problems that are influencing the overall ease of use.

The difference between the mean values of these two constructs (4.11 vs. 3.97) suggests that usability problems were mainly related to the ease of use (i.e. software) than to the ergonomics of the devices and accessories (i.e. hardware).

The highest mean value got the item 16, showing that the system makes learning more interesting. Except for item 18, the rest of the items related to the perceived enjoyment were scored over 4.00. This shows that students perceived the learning experience with ARTP as interesting, captivating and exciting. The mean value computed for this construct was 4.11.

An analysis using Pearson‟s correlation indicated that there is a significant linear relationship ($R=0.512$, $p<0.001$) between the general item 15 and the general item 20. This shows that the perceived enjoyment of learning with ARTP is related to its usability.

A comparison with the evaluation results of a previous version described in [4] and [20] revealed an improvement in usability of the Biology scenario.

In general, students testing ARTP in 2007 (N=42) scored lower than students participating at testing in 2008 (general mean of 3.96 vs. 4.06). An independent samples t-test revealed that differences were statistically significant ($\alpha=0.05$, $DF=179$) for the following items: 1 ($t=2.800$, $p=0.006$), 5 ($t=2.303$, $p=0.002$), 7 ($t=2.254$, $p=0.025$) and 15 ($t=2.152$, $p=0.033$).

4.2. Qualitative data results

The answers to the open questions were analyzed in order to extract key words (attributes). Some students only described one or two aspects while others mentioned several aspects in one sentence thus yielding a total number of 304 attributes related to positive aspects and 226 attributes related to negative aspects.

The attributes were then grouped into predefined categories having in mind the goal of our study: to integrate quantitative and qualitative data in order to get a broader view on the usability of ARTP and a higher reliability of the evaluation results.

Therefore, grouping was done in two stages by two independent experts. The first step was to group attributes that are related to the same aspect. Then we aggregated these categories in broader categories following the dimensions targeted by the closed items.

The main categories of positive aspects are presented in Table 2. The students liked the 3D visualization and perceived ARTP as enjoyable and useful for learning. Most positive aspects were related to the perceived enjoyment (33.88%).

Table 2. Main categories of positive aspects

Category	Frequency	%
ARTP capabilities	90	29.61
Perceived usefulness	82	26.97
Perceived enjoyment	103	33.88
Perceived ease of use and other	29	9.54
Total	304	100.00

The positive aspects that are related to the perceived enjoyment are further detailed in Table 3.

Table 3. Categories of positive aspects related to enjoyment

Category	Frequency	%
Interesting	43	41.75
Captivating	23	22.33
Novel, attractive	15	12.62
Funny, stimulating, exciting	13	12.62
Enjoyable, in general	11	10.68
Total	103	100.00

Students perceived ARTP as interesting („*The lesson is more interesting"; „It is very interesting since is another learning system"*; „*Exercises are very interesting"*), captivating („*I liked this system because is captivating"; „Using the system is captivating"*), novel („*Is something new"*), attractive („*Is amazing how beautiful can be*"), funny („*It was funny*"; „*A funny way to learn*"), stimulating („*Is a true stimulant for our mind*", „*Is an exciting way of learning*"), and, in general, enjoyable („*I liked everything*", „*Enjoyable to use*").

Overall, these excerpts from students" comments show that learning with ARTP was perceived as an enjoyable experience and highlight the motivating value of the AR technology for the educational process.

The main categories of negative aspects are presented in Table 4. Most of them are related to the ease of use (35.40%), ergonomics of the ARTP (33.19%), and manipulation of the real object (24.34%). In this respect, the answers at open-ended questions proved to be a valuable aid in understanding the usability problems.

Table 4. Main categories of negative aspects

Category	Frequency	%
Perceived ease of use	80	35.40
Ergonomics of the ARTP	75	33.19
Real object	55	24.34
Other	16	7.08
Total	226	100.00

These three categories of negative aspects related to usability problems are further detailed in Table 5.

Most of the negative aspects were related to the torso shared by two students ("*I didn't like to drag the torso in order to reach the oral cavity*", "*Moving the torso to perform the exercises*", "A torso is needed for each student", "*The fact that we had to work two with the same torso*").

Other usability problems were related to headphones ("*Headphones are not comfortable*"), sound ("*At some times the sound was interrupted*"), selection

("*It was very difficult to select objects*"), glasses ("*I experienced headaches after use*"), AR setting too narrow and temperature too high ("*The work place should be larger*"; "*Too little space between the screen and the table*", "*Too hot*"), superposition problems ("*Superposition of objects is not clear*"), and remote control problems ("*Button B was blocked several times*").

Table 5. Categories of negative aspects related to usability

Category	Frequency	%
Perceived ease of use, from which		
- selection problems	37	17.63
- superposition accuracy	15	7.14
- remote control problems	10	4.76
- other difficulties	18	8.57
Ergonomics of the ARTP, from which		
- head phones and sound	38	18.10
- glasses	25	11.90
- AR setting	12	5.71
Real Object		
- torso too big and difficult to move	55	26.19
Total	210	100.00

The qualitative data analysis helped to better understand why students scored higher or lower on some closed items and how specific usability problems are affecting the ease of use. The difference between the number of negative aspects in each category also shows that usability problems were mainly related to the ease of use (i.e. software) than to the ergonomics of the devices and accessories (i.e. hardware).

4.3. Integrating findings and discussion

On the positive side, the ARTP capabilities were highly appreciated by students. This shows a limitation of the evaluation instrument that did not target this dimension. Most of the positive aspects that were related to enjoyment are showing that ARTP makes learning more interesting (41.75%). This is consistent with the highest mean value got for item 16 (M=4.35). These findings are showing that the advantage of triangulation is twofold: (a) to overcome the limitations of a single method and (b) to increase the confidence in the evaluation results.

On the negative side, the usability problems accounted for 92.92% from all the negative aspects mentioned by students. This is consistent with the low mean value computed for all the items targeting the perceived ease of use.

It was surprising that students were not happy to share the torso with a colleague. This was an unexpected usability problem which was revealed by the qualitative data analysis. It was also useful to find out that selection and superposition issues together accounted for 24.77% from the total of usability problems. Triangulation of results confirmed that the clarity of visual perception seriously affects the usability of interaction techniques.

An interesting research question is how usability problems are affecting the user experience with ARTP. An analysis using Kendall‟s Tau indicated that there is a significant correlation ($\tau=-0.196$, $p<0.05$) between the number of negative aspects related to the usability of the ARTP and item 18 (I like interacting with real objects) that got the lowest mean value among the closed items related to enjoyment (M=3.91).

The negative correlation explains why this item got a surprisingly low rating that is sharply contrasting with the fact that students perceived ARTP as enjoyable while interacting with real objects is essential for AR technologies. On the other hand, the qualitative data analysis revealed the reason of this perception: the torso had to be shared between two students thus making the interaction a fight for the real object.

5. Conclusion

In this paper we took a mixed methods approach to usability evaluation of an AR-based learning scenario by integrating quantitative data based on closed items with qualitative data based on open-ended questions. The analysis was structured following the dimensions of closed items that are related to usability: ergonomics of the ARTP, perceived ease of use and perceived enjoyment. Not only the triangulation increased confidence in the evaluation results but also provided with a broader view on the strengths and weaknesses of the learning scenario.

Comparison of the quantitative and qualitative data made it clear that usability problems were mainly related to the software part of the ARTP. A critical usability problem for desktop AR systems is the accuracy of the visual perception which depends on both the see-through screen (hardware issue) and superposition of the augmentation with the real object (software issue).

A specific usability problem with this scenario was the size of the real object (a flat torso of the human body) and the fact that it had to be shared by two students. The triangulation revealed that this usability problem had a negative impact on the students‟ perception regarding the AR-based interaction.

Nevertheless, both categories of data show that ARTP makes learning more interesting and the Biology scenario was perceived by students as an enjoyable and exciting learning experience. Overall, the integration of quantitative and qualitative proved to be useful since it enabled a deeper understanding on how specific usability problems are affecting the user experience with ARTP.

Acknowledgments

We gratefully acknowledge the support of the European research project FP6 027039 and the national research project TEHSIN 503/2009.

REFERENCES

[1] Aart, J., Bartnek, C., Hu, J., Rauterberg, M., Salem, B. (2010) How to behave as Alice in Wonderland – about boredom and curiosity. *Entertainment Computing 1*, 125-137.

[2] Arnon, S., Reichel, N. (2009). Closed and open-ended question tools in a telephone survey about „The good teacher‟: An example of a mixed methods study. *Journal of Mixed Methods Research, 3 (2)*, 172-196.

[3] Lanzilotti R., Ardito, C., Costabile M.F., De Angeli, A. (2010) Do patterns help novice evaluators? A comparative study. *International Journal of Human-Computer Studies 69(1-2),* 52-69.

[4] Balog, A., Pribeanu, C. (2008) A comparative usability evaluation of two augmented reality learning scenarios. *Proceedings of SIGMAP 2008*, INSTICC, 370-375.

[5] Denscombe, M. (2008) Communities of practice. A research paradigm for the mixed methods approach. *Journal of Mixed Methods Research, 2 (3)*, 270-283.

[6] Denzin, N. (1978) *The Research Act*, New York: McGraw-Hill.

[7] Dunser, A., Grasset, R., Billinghurst, M. (2008) A survey of evaluation techniques used in augmented reality studies. *Proceedings of SIGGRAPH* Asia, ACM, 1-27.

[8] Fidel, R. (2008) Are we there yet? Mixed methods research in library and information science library. *Library and Information Science Research 30*, 265-272.

[9] Gabbard, J., Swann, E. (2008) Usability engineering for augmented reality: Employing user-based studies to inform design. *IEEE Transactions on Visualization and Computer Graphics 14(3)*, 513-525.

[10] Hassenzahl, M., Diefenbach, S., Goritz, A. (2010) Needs, affect, and interactive products – Facets of user experience, Interacting with Computers, 22, 353-362.

[11] Hornbaek, K., 2006. Current practice in measuring usability: Challenges to usability studies and research". *International Journal of Human Computer Studies 64*. 79-102.

[12] Huang, H.M., Rauch, U., Liaw, S.S. (2010) Investigating learners‟ attitude towards virtual reality learning environments: based on a constructivist approach, Computers & Education, 55, 1171-1182.

[13] Iordache, D.D., Pribeanu, C. (2009) "Comparison of Quantitative and Qualitative Data from a Formative Usability Evaluation of an Augmented Reality Learning Scenario" *Informatica Economică Journal, 13 (3)*, 67-74.

[14] ISO 9126-1:2001 Software Engineering - Software product quality. Part 1: Quality Model.

[15] Jick, T. (1979). Mixing Qualitative and Quantitative Methods: Triangulation in Action, *Administrative Science Quarterly 24(4)*. 602-611.

[16] Johnson, R.B., Onwuegbuzie, J., Turner, L. (2007) Towards a definition of mixed methods research, *Journal of Mixed Methods Research, 1 (2)*, 112-133.

[17] Krauss, M., Riege, K., Winter, M., Pemberton, L. (2009) Remote Hands-On Experience: Distributed Collaboration with Augmented Reality. *Proceedings of. EC-TEL 2009, LNCS 5794*, Springer, 226-239.

[18] Lamanauskas, V., Pribeanu, C.., Vilkonis, R., Balog, A., Iordache, D.D., Klangauskas, A. (2007) Evaluating the educational value and usability of an augmented reality platform for school environments: some preliminary results. *Proceedings of WSEAS/IASME - Mathematics and Computers in Science and Engineering,,* 86-91

[19] Modell, S. (2009) In defense of triangulation: A critical realist approach to mixed methods in management accounting, *Management Accounting Research 20*, 208-221.

[20] Pribeanu, C., Iordache, D.D., Balog, A. (2008) Evaluarea utilizabilităţii unui scenariu de învăţare a biologiei implementat pe o platformă de realitate îmbogăţită. *Revista Română de Interacţiune Om-Calculator 1(1*), 39-56.

[21] Scriven, M. (1991) *Evaluation thesaurus*. 4th ed. Newbury Park, CA: Sage Publications.

[22] Tang, A., Owen, C., Biocca, F., Mou, W. (2003) Comparative effectiveness of augmented reality in object assembly. *Proceedings of CHI 2003, 5(1)*, 73-80.

[23] Theofanos, M. & Quesenbery, W. (2005) Towards the Design of Effective Formative Test Reports. *Journal of Usability Studies 1(1)*, 27-45.

[24] Wind, J., Riege, K., Bogen M., 2007. Spinnstube®: A Seated Augmented Reality Display System, *Virtual Environments: Proceedings of IPT-EGVE – EG/ACM Symposium*, 17-23.