Evaluation of Augmented Reality Technology in the English Language Field

Manoela Silva[1], Rafael Roberto[1], Veronica Teichrieb[1]

[1] Voxar Labs – Centro de Informática – Universidade Federal de Pernambuco
Recife, PE – Brasil

{mmos, rar3, vt}@cin.ufpe.br

Abstract. *Augmented Reality (AR) is heralded to be a promising technology for education. Therefore, it is important to properly evaluate it so practitioners feel more confident in its use. The aims of this study are threefold: (a) to evaluate the use of an AR tool for young children, in the English teaching field concerning linguistic concepts and competencies; (b) to discuss possibilities of use for AR tools in the classroom environment as well as reflect about the potential and difficulties involved in their introduction, and (c) to provide guidelines in order to assist researchers when conducting similar evaluations. The evaluation considered multiple metrics and involved the teacher as an instructional designer. Results supported the hypothesis that AR may help students in the learning process.*

1. Introduction

AR is a new technology that consists in adding virtual elements coherently in a real scene [Azuma et al. 1997]. Although all fields of knowledge can potentially take advantage from AR, [Tori et al. 2006] argue that education will be particularly modified by its introduction and changes in interaction promoted by the mix of virtual and real information. Coexistence of virtual and real information allows learners to visualize complex spatial relationships and abstract concepts. In [Silva et al. 2014], an AR application uses real puzzle pieces to display augmented content and, thus, foster geographic skills.

While AR offers new learning opportunities, it also creates challenges for education, such as technological, learning and pedagogical issues. Technological issues are related to cumbersome and expensive equipment, discomfort and poor depth perception. Learning issues include cognitive overload, requirement to apply and synthesize multiple complex skills in spatial navigation and collaboration [Wu et al. 2013]. Pedagogical issues refer to instructional design, flexibility of the content, constraints from schools and resistances among teachers [Kerawalla et al. 2006]. Therefore, it is important to mind the gap between teaching methods in use in most classrooms and the student-centered and exploratory nature of learning proposed by AR systems. Proper evaluation of AR tools is also required to understand their impact in the learning setting.

To evaluate AR's potential in the English Teaching field for young learners, a quasi-experimental study was carried out. The specific aims of this study are: (a) to evaluate the use of an AR tool in the English teaching field concerning linguistic concepts and competencies; (b) to discuss possibilities of use for AR tools in the

DOI: 10.5753/cbie.sbie.2015.577

classroom environment as well as reflect about the potential and difficulties involved in its introduction, and (c) to provide guidelines in order to assist researchers when conducting similar evaluations.

This work is organized as follows: Section 2 shows related works regarding the evaluation of AR in education and characteristics of young children. Section 3 describes the evaluation method in detail along with the tool and the participants. In Section 4, the evaluation results are presented and discussed. Finally, Section 5 draws some conclusions and suggests future directions for the research.

2. Related Works

So far the amount of AR systems formally evaluated is rather small [Dünser et al. 2007]. For example, literature surveys of user evaluation in AR have found that only around 8% of published AR papers include formal evaluations [Dünser et al. 2008]. One reason for this small percentage may be the lack of suitable methods for evaluating AR interfaces [Dünser and Billinghurst 2011].

Researchers in emerging interface fields such as Virtual Reality (VR) or AR cannot rely solely on design guidelines for traditional user interfaces since new interfaces afford new forms of interactions [Dünser and Billinghurst 2011]. It is also relevant to consider the point of view of both teachers and learners since they might differ. For instance, [Balog and Pribeanu 2010] had shown the perceived usefulness and the perceived enjoyment as relevant factors for student's acceptance of an AR application, while the perceived ease of use was not a significant factor for student's acceptance. Additionally, it is important to employ data collection and analysis techniques associated with both quantitative and qualitative data as a way of compensating the weakness of each method [Easterbrook 2008].

Many papers evaluate educational aspects by using mainly the pre-post test design [Arvanitis et al. 2009]. There are works that use multiple metrics in the evaluation, although many use them to evaluate not only educational aspects, but, other issues, such as: usability and satisfaction. It is also noticeable that few studies involve the teacher as an instructional designer. However, some studies use their expertise to craft their tools [Echeverría et al. 2012].

A recent survey reviewed applications intended to complement traditional curriculum materials for K-12 [Santos et al. 2014]. This study discovered that aside from the performance in pre and post-tests, other aspects such as motivation and satisfaction were usually observed in the literature.

In order to properly evaluate tools with young children it is important to consider their specific characteristics and needs. The main contrasts between children and adult learners are: (a) children are more likely to play with language; (b) they become more engaged through stories and games; (c) younger children are less likely to notice or correct errors [Peck 2001]. Children are holistic learners who need to use language for authentic purposes. Thus, it is important to teach them holistically and to provide varied material and child-centered activities [Reily and Ward 1997].

3. Evaluation

The evaluation of the ARBlocks was performed through a semi-experiment in an English language school located at Recife, Pernambuco in 2014 for approximately three months.

3.1. Experiment Design

Before the experiment, researchers got together with teachers to explain the objectives of the research and present the tool to be evaluated. Teachers manipulated it and, then, participated in a semi-structured interview about its potential. Teachers decided how to use the tool according to their classroom needs and were able to generate ideas for the activities. Researchers were responsible to make them accordingly. The teachers, coordinators and researchers decided which groups would be the case and control groups.

Different metrics both quantitative and qualitative were used. Firstly, it was considered student's academic achievement. All groups have 2 evaluation moments, the middle and final term, which is also a good indicator of students' previous knowledge. They are evaluated holistically regarding: continuous learning (i.e: students' overall development concerning language aspects), behavior, homework – frequency, homework – performance, participation and speaking. Grades are divided among four concepts: ED (excellent performance), BD (good performance), DS (borderline performance) and DM (unsatisfactory performance).

Students started to use the tool after the middle term evaluation and worked with it until the final term evaluation. At the end of the sessions, students from all groups answered an activity in order to evaluate their learning of the content seen throughout the semester. Performance in the content studied with the tool was compared to the performance in contents not worked with it.

Statistical analysis was performed for all groups concerning student's academic achievement and the final test applied by researchers. The Kolmorov-Smirnov test [Dancey et al. 2012] was used to check data distribution with significance level of 95%. For this analysis, the scores of the tests were converted into grades using a scale provided by the teachers: DM = less than 7; DS = 7 to 7.9; BD = 8 to 8.9; ED = 9 to 10.

As qualitative metrics, a research diary was used since it provides information about day to day activities that can be explored in a subsequent interview [Ortlipp 2008]. Questions in the diary covered aspects such as the kind of activity, skills practiced, type of interaction and teachers impressions. At the end of the semester, teachers participated in a semi-structured interview to evaluate the introduction of AR. All interviews were audio-recorded. The questions covered aspects, such as the interaction with the tool, its impact in the class and planning issues. A diagram with the entire process can be seen in Figure 1.

3.2. ARBlocks and the System Setup

The system used was ARBlocks [Roberto et al. 2013], which is an AR tool developed to scaffold education. It combines the principles of projective AR within tangible interaction. It allows for a variety of contents to be projected in its faces. The tool also provides visual and auditory feedback. Additionally, different activities can be done with the same set of blocks. Although preliminary, recent evaluation supported the hypothesis

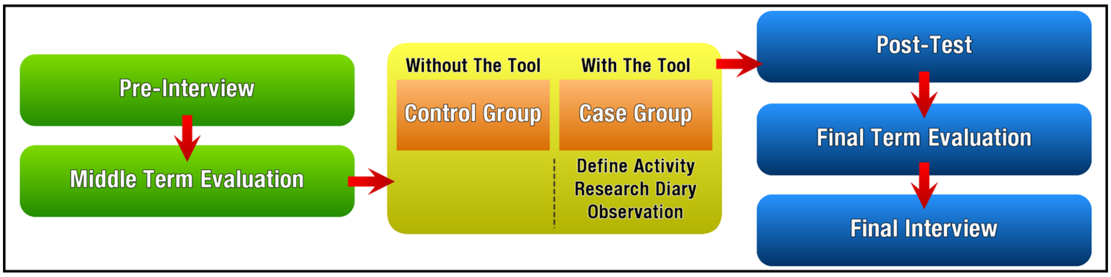

Figure 1. Diagram with the methodology design of the study. The middle and the final term evaluation correspond to students' academic achievement.

that the ARBlocks can help to motivate students and foster the literacy development [Silva et al. 2013].

For our work, the experiments were conducted in a regular classroom provided by the school. The ARBlocks run in an ordinary laptop having an Intel Core-i7 processor with 8 GB of RAM, a dedicated graphics card, a built-in speaker for the sonorous feedback and Windows 7. The computer was connected to an Epson projector EB-X10, similar to those found in several schools. It was used a Microsoft webcam LifeCam Cinema, that is also a standard model. The projector was attached to an adapted projector tripod and pointed down to the floor. The webcam was taped on the top of the projector in order to see the entire projection area.

3.3. The School

Language course curricula are usually designed for a semester, which provides a better overview of the process. In addition, the school chosen is known for its interest in applying technology tools.

3.4. The Participants

Teacher A is undergraduated in language teaching and has been lecturing for 7 years. She teaches the Pre-Kinder 2 (PK2) groups. PK2 groups have lessons in the morning shift twice a week. Classes last one hour and fifteen minutes. Case group has 12 students while the control group has 10 students.

Teacher B is undergraduated in computer science and has been teaching for 23 years. She has language certificates and teaches the Kids 1 (K1) groups. K1 classes last one hour and fifteen minutes twice a week. K1 case group has lessons in the morning shift and has 9 students. K1 control group has lessons in the afternoon shift and has 8 students.

3.5. The Activities

Both PK2 and K1 teachers elaborated activities based on what they were studying at the moment. ARBlocks sessions lasted 15-25 minutes each. PK2 group had sessions twice a week, while, due to time constraints, K1 group had sessions once a week.

The activities were mostly related to four types: (a) gap-filling, i.e: students had to fill in some sentences with the correct word or grammar structure; (b) matching, in which they had to match content, e.g: numbers with their respective names; (c) sorting, in which they had to categorize something, e.g: sort out vocabulary according to the correct article (a or an); and (d) forming words, e.g: they had to write names related to a certain content using the words projected in the blocks. Some of them can be seen in Figure 2.

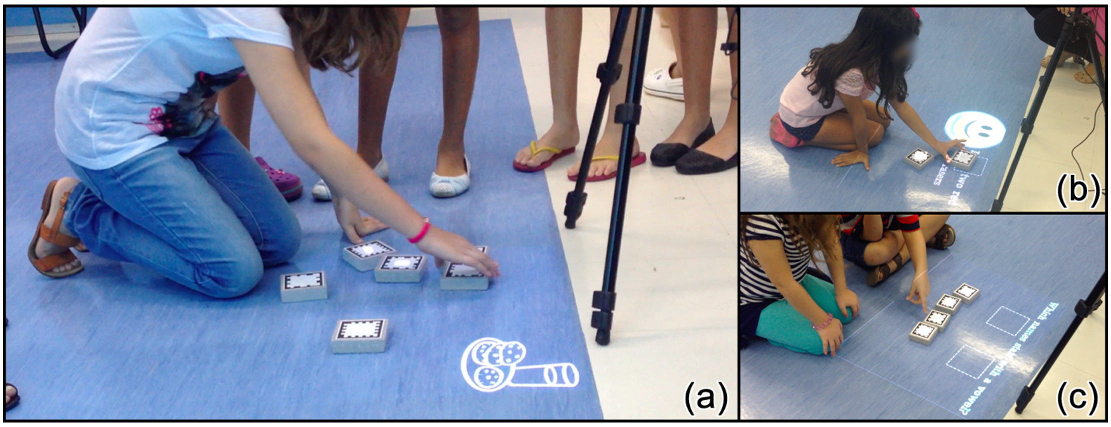

Figure 2. Examples of activities performed with the ARBlocks in both K1 and PK2 groups. (a) shows a forming word activity, (b) an example of gap-filling and (c) an example of sorting.

4. Results

In this section, the results of the experiment will be properly described and discussed.

4.1. PK2 Results

In the first interview, teacher A classified the tool as excellent and coherent with student's reality. She highlighted that it was a different way to review content and practice the language skills, particularly, reading.

After every session, she filled in the research diary. She reported 11 sessions. To begin with, she recorded basic information, i.e: level, time, date and the activity done. She mentioned reading, listening and speaking as skills practiced with the tool. She reported that students interacted well with the tool. She mentioned that they loved the exercise and were anxious and enthusiastic to take part of the activity. She reported that they showed a good comprehension of the rules as well as confidence and pleasure in moving the blocks.

She observed students learning and practicing different abilities, such as: categorization, listening, review vocabulary and pronunciation, spelling as well as consolidation of previous knowledge. To conclude, in the section for free comments, she usually wrote positive aspects. For instance, she mentioned that students loved one of the activities because it was different from what they were used to.

Another aspect evaluated was academic achievement. Through the middle term evaluation of the case group, it is possible to observe that overall students presented a good development and command of the language. In the final term evaluation, students progressed in all aspects of learning. The control group also presented a good command of the language. Overall, in the final term evaluation, control students also progressed. Concerning continuous learning, though, control group students did not show progress. The results from both case and control groups can be seen in Figure 3.

Statistic comparisons between the middle and final term evaluation showed that most of the categories were not significantly different. For the case group, only

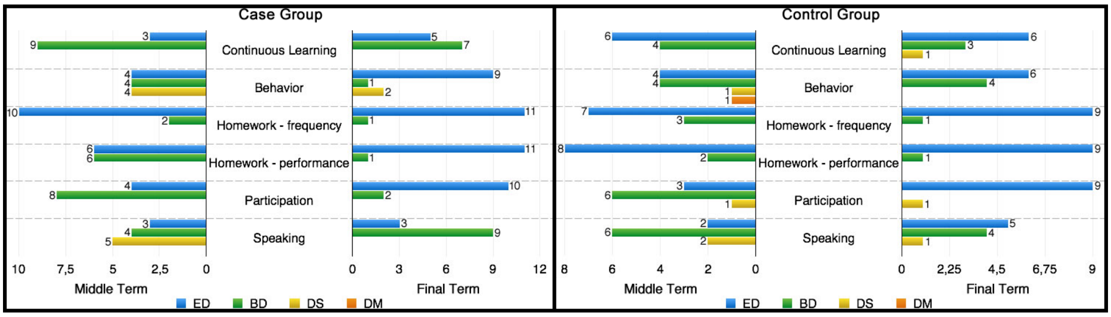

Figure 3. On the left side, PK2 case student's performance in both middle and final term evaluations. On the right side, PK2 control student's performance in both middle and final term evaluation.

the homework-performance presented statistical difference (p=0.025). This result may suggest that the use of the tool had a positive impact on student's performance regarding the topics studied. The other research instruments used reinforced this positive impact. For the control group, the results were also non-statistic significant, except for the behavior (p=0.049) and speaking (p=0.015). This result was coherent with what was observed in the classes and exposed by the teacher. This group presented behavior problems that were dealt with throughout the semester.

Teacher A worked with contents from 7 out of 9 units of the book. After the use of the tool, the post test was applied with all PK2 groups. It consisted of 19 items, divided in 16 questions about contents worked with the tool and 3 questions related to content not worked with it. The case group grade is divided in two parts (questions worked with the ARBlocks and questions not worked with it). The case group achieved a greater score in the questions using AR. Case group average grade in the AR part was 8.649 while their average grade in the non-AR part was 7.5. The score of the AR part of the test was slightly higher than the control group (8.538). Statistical comparisons showed that the results were non-statistic significant. In this test, the control group presented slight better scores than the case group. Although the observations and the other instruments suggested the ARBlocks had a positive impact on students learning, this impact was not statistically reflected in this test.

In the final interview, teacher A classified the experience as positive and constructive. In her words, the tool was simple and easy to use. In addition, she enjoyed the possibility of combining the writing to the listening and that she did not need to interfere so much during the activities. She considered the planning process smooth. She did all the planning basically in 2 moments. First, she thought about adapting the content, however, she took into account the best way to display the information in the blocks. She felt great connection among what the ARBlocks offered and what she wanted students to analyze and to what they were used to work.

She claimed that children seemed to be anxious to participate. She mentioned one of her students who demonstrated resistance to work with writing in a paper, but felt really motivated to manipulate the ARBlocks. When the sessions were over, she remembered that some students were still in the classroom playing with the tool. She also highlighted that students did not have problems even when the projection took longer to appear. The teacher mentioned that the dynamic of activities was varied. She explained the criteria used to decide the interaction. First, she thought about how to help students in their early literacy process. Then, she considered the best way to organize students for the activity. She mentioned that students were anxious to finish the class with the blocks because to them this was like bringing it to a perfect end.

She reported that she noticed students learning and that they got used to do the activities and waited for the feedback, which suggests that students were able to work independently. To conclude, she mentioned that the experience was positive and she resented not having used the tool with her other group. She also pointed out that students were able to interfere in the projection through their actions.

4.2. K1 Results

In the first interview, teacher B explained that the tool is strongly related to the school's work since technology is already part of their lives. She also believed that it might raise students' attention and, therefore, motivate them. Regarding the work with language skills, she envisioned different possibilities on activities involving matching, sorting out and gap filling. However, she mentioned some difficulties in realizing how to work with reading and speaking.

In the research diary, K1 teacher mentioned reading, listening, recognition, vocabulary and spelling as the skills practiced with the tool. In most of the sessions, students interacted with the tool individually. In the first one, they interacted in small groups as a competition. In the first 4 sessions, she reported that the students liked the activities. For instance, she mentioned that they were excited and eager to participate and to touch the blocks and that "they liked and helped each other". However, in the last 2 sessions, she reported that students were "a little bored" and that she felt they needed more interaction.

When asked if she observed they learning something using the application, her answers were mostly related to reviewing and reinforcement. She mentioned that most of the time students did not have difficulties using the tool, except in the clothing activity in which she reported that some pieces of clothing were not so clear. The space for free comments was left empty for most sessions. However, in the last one, she mentioned that "maybe the kind of interaction was not adequate for them".

As concerns the academic achievement, the middle term evaluation showed that K1 case group overall presented a good command of the language. In the final term evaluation, students progressed in most of the topics. However, for behavior, they remained with the same grades.

In the middle term evaluation, K1 control group presented a good command of the language. Although their scores were, in general, lower than the case group. In the final term evaluation, students overall progressed in almost all categories. All scores from K1 both case and control group are presented in Figure 4.

Statistical comparisons of academic achievement of the case group showed that the results were non-statistic significant, except for speaking (p=0.036). For the control group, results were also non-statistic significant, except for behavior (p=0.020) and speaking (p=0.08). The results concerning behavior were coherent with the other results obtained. This group presented behavior issues, which were dealt with throughout the

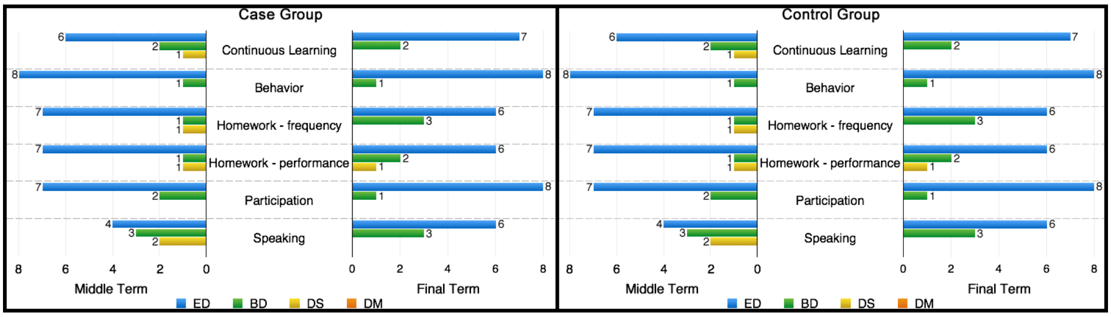

Figure 4. On the left side, K1 case student's performance in both middle and final term evaluation. On the right side, K1 control student's performance in both middle and final term evaluation.

entire semester. As for the speaking, students from both K1 groups and the PK2 control group presented statistic significant improvements. This may be explained by the fact that speaking is an ability that is generally more difficult for students to practice outside the school. Hence, teachers constantly reinforce this ability at school. For instance, in the K1, it was observed that the teacher had a chart in the class in which she marked every time students used Portuguese expressions.

After the work with the ARBlocks, a post test was applied. This evaluation consisted of 23 items divided in 15 items worked with the tool and 8 items related to content not worked with it. The teacher worked with contents from 4 out of 9 units of the book. For the post test scores, case group results are divided between the questions worked with and without AR support. In this test, case group had the same scores in the entire test (8.75). This grade was higher than the control group (8.152).

Statistic comparisons revealed that these results were statistically significant. The case group presented better scores (p=0.026). The observations and the other instruments suggested that the ARBlocks had a positive impact on the reinforcement of the content, although the teacher had some reservation regarding the interaction aspects of the tool.

In the final interview, teacher B explained that the experience was positive but there is a need to take into account some interaction aspects. She believes the interaction was repetitive since the activities ended up being just about vocabulary recognition. She stated that the technology is interesting but it needs to be expanded to explore more reading or speaking. She reported that it is a valid technology, however, it should be seen in a different way since students are used to competition. She affirmed that the problem was also related to group work in which students would interact with the tool at the same time. She mentioned that students who were not working were bored even though she tried to ask them to help. Regarding its impact, she claimed that it was an reinforcement of content.

The teacher claimed she did not know that she needed to plan the lessons and include the ARBlocks. She thought she needed to explain her lesson plan to programmers and they would adapt the activities. Therefore, she thought about her lesson plan and how some activities could be done with the blocks. The teacher mentioned that the tool had its contribution, for instance, she pointed out the reinforcement of language structure and the combination of the visual aid and the ability to touch and test hypotheses. She argued that it came to complement, as a novelty to help in the classroom. However, she listed some points of concern, such as the size of the projection on the floor (she believed it was small) and the limitation of the blocks. To conclude, she mentioned that in order to bring this tool to her school, it would be necessary to rethink some of the issues mentioned.

5. Conclusion

We designed a quasi-experiment in order to examine the impact of the ARBlocks in the English language teaching field. This evaluation encompasses multiple metrics, both qualitative and quantitative as well as the involvement of the teachers in elaborating suitable activities for their classroom needs.

In the PK2 groups, the results supported the hypothesis that the ARBlocks can help to motivate students and foster the development of their language skills. In the evaluation applied after using the tool, control group achieved a statistically higher score. However, the results from other metrics showed that AR use was beneficial for students. The results from K1 groups were also motivating. In the evaluation applied after using the tool, case group achieved a statistically significant higher score. However, K1 teacher reported that the tool supported limited interaction.

As for the evaluation process, we argue for the benefits of involving teachers in crafting activities directly related to their student's needs. This is a challenging type of evaluation since teachers are not merely subjects but actual partners in the process. Although defying, we believe this involvement may bring important insight for both researchers and developers. The use of multiple metrics both qualitative and quantitative is also favored since it allowed researchers to have a better overview of the big picture and counteract for weakness of individual methods as well as better understand problems that might have occurred.

As for limitations, it was not possible to have a larger sample of students from the same groups, which would have reinforced our conclusions and assured statistic significance of the results. With the ARBlocks, it was also possible to work not only with content but with behavior aspects as well, such as waiting for a turn. Future works might investigate how AR tools could influence student's competencies and social abilities.

6. Acknowledgments

Authors would like to thank CNPq for partially funding this research (process 134114/2013-2), professor Patrícia Smith from EDUMATEC - UFPE for her valuable suggestions, as well as the school, teachers and students who participated in this study.

References

Arvanitis, T., Petrou, A., Knight, J., Savas, S., Sotiriou, S., Gargalakos, M., and Gialouri, E. (2009). Human factors and qualitative pedagogical evaluation of a mobile augmented reality system for science education used by learners with physical disabilities. *Personal and Ubiquitous Computing*, 13(3):243–250.

Azuma, R. T. et al. (1997). A survey of augmented reality. *Presence*, 6(4):355–385.

Balog, A. and Pribeanu, C. (2010). The Role of Perceived Enjoyment in the Students' Acceptance of an Augmented Reality Teaching Platform: a Structural Equation Modelling Approach. *Studies in Informatics and Control*, 19(3):319–330.

Dancey, C. P., Reidy, J. G., and Rowe, R. (2012). *Statistics for the Health Sciences: A Non-Mathematical Introduction*. Sage Publications.

Dünser, A. and Billinghurst, M. (2011). Evaluating augmented reality systems. In Furht, B., editor, *Handbook of Augmented Reality*, pages 289–307. Springer New York.

Dünser, A., Grasset, R., and Billinghurst, M. (2008). A Survey of Evaluation Techniques Used in Augmented Reality Studies. Technical report.

Dünser, A., Grasset, R., Seichter, H., and Billinghurst, M. (2007). Applying HCI Principles in AR Systems Design. In *2nd International Workshop on Mixed Reality User Interfaces: Specification, Authoring, Adaptation (MRUI '07)*.

Easterbrook, S. e. a. (2008). Selecting empirical methods for software engineering research. *IEICE Transactions on Information and Systems*, page 285–311.

Echeverría, A., Améstica, M., Gil, F., Nussbaum, M., Barrios, E., and Leclerc, S. (2012). Exploring different technological platforms for supporting co-located collaborative games in the classroom. *Computers in Human Behavior*, 28(4):1170 – 1177.

Kerawalla, L., Luckin, R., Seljeflot, S., and Woolard, A. (2006). "making it real": exploring the potential of augmented reality for teaching primary school science. *Virtual Reality*, 10(3-4):163–174.

Ortlipp, M. (2008). Keeping and using reflective journals in the qualitative research process. *The Qualitative Report*, 13(4):695 – 705.

Peck, S. (2001). Developing children's listening and speaking in esl. In Celce-Murcia, M., editor, *Teaching English as Second or Foreign Language*, pages 139–149. Heinle Cengage Learning.

Reily, V. and Ward, S. (1997). *Very Young Learners: Resource Book for Teachers*. Oxford University Press.

Roberto, R., Freitas, D., Simoes, F., and Teichrieb, V. (2013). A dynamic blocks platform based on projective augmented reality and tangible interfaces for educational activities. *Journal on Interactive Systems, SBC*, 4(2):8–18.

Santos, M., Chen, A., Taketomi, T., Yamamoto, G., Miyazaki, J., and Kato, H. (2014). Augmented reality learning experiences: Survey of prototype design and evaluation. *Learning Technologies, IEEE Transactions on*, 7(1):38–56.

Silva, M., Roberto, R., and Teichrieb, V. (2013). Evaluating educational system based on projective augmented reality. In *Anais do SBIE 2013*, pages 10 pp.–.

Silva, M., Vilar, E., Reis, G., Lima, J. P., and Teichrieb, V. (2014). Ar jigsaw puzzle: Potencialidades de uso da realidade aumentada no ensino de geografia. In *Anais do SBIE 2014*, pages 8–17.

Tori, R., Kirner, C., and Siscoutto, R. (2006). *Fundamentos e tecnologia de realidade virtual e aumentada*. Editora SBC.

Wu, H.-K., Lee, S. W.-Y., Chang, H.-Y., and Liang, J.-C. (2013). Current status, opportunities and challenges of augmented reality in education. *Computers & Education*, 62(0):41 – 49.