Published as: Tiede, J., Mangina, E., & Grafe, S. (2021). Evaluation Design Methodology for Piloting Two Educational Augmented Reality STEM Apps in European Elementary Schools. *EdMedia + Innovate Learning 2021*, Online, July 6–8, 2021, 185–190.

Evaluation Design Methodology for Piloting Two Educational Augmented Reality STEM Apps in European Elementary Schools

Jennifer Tiede
Institute of Pedagogy, University of Würzburg, Germany
jennifer.tiede@uni-wuerzburg.de

Eleni Mangina
School of Computer Science, University College Dublin, Ireland
eleni.mangina@ucd.ie

Silke Grafe
Institute of Pedagogy, University of Würzburg, Germany
silke.grafe@uni-wuerzburg.de

Abstract: According to research findings, Augmented Reality (AR) can have a positive impact on student STEM learning and knowledge acquisition. However, sound concepts for assessing this impact especially in an international sample are scarce, even though an international comparative perspective opens up important viewpoints and helps understand individual and national conditions of the usefulness within specific concepts. This paper will introduce the research methodology applied in the European H2020 ARETE project to systematically assess and analyze an intervention utilizing AR in STEM classes in elementary education across Europe, compared within the background work of this research area. This methodology allows for a grounded data collection approach in order to bring forward important research results on a European level.

Introduction: Augmented Reality and STEM Learning

STEM learning has become an important objective of today's school education. Starting in elementary school already, lessons for example in mathematics and geography support students in understanding the world we live in and in acquiring elementary 21st century skills (Johnson, Peters-Burton, & Moore, 2016). As Ibáñez and Delgado-Kloos (2018) analyse, Augmented Reality (AR) has proven to be a valuable tool for supporting and enhancing STEM learning, e.g. by fostering affective states of students and advancing knowledge acquisition. It also has the potential to decrease cognitive load and to support skills such as critical thinking and problem-solving. However, the authors state that there are also caveats related to the novelty effect or to usability problems.

The research project Augmented Reality Interactive Educational System (ARETE) builds on related findings and aims to systematically support fourth to fifth grade students' STEM learning by two interactive Augmented Reality apps. These two apps are used in a pilot study which includes approximately 160 classrooms across Europe to enhance teaching and learning processes in mathematics and geography lessons.

To evaluate the impact the apps have on student learning, on knowledge acquisition and retention and on teaching and learning processes in fourth and fifth grade STEM lessons, it is essential to design a sound and valid evaluation approach. In the project presented, this evaluation approach includes pretests, posttests and retention tests of different formats targeting both teachers and students. The development of this methodology will be presented in the following sections, embedded in the relevant state of the art concerning AR apps and based on related research findings. The approach aims to respond to the following research question: "Which evaluation design and metrics are suitable for systematically assessing students' STEM knowledge acquisition and retention under use of Augmented Reality? ".

State of the Art of AR-Apps in STEM-Learning in the Primary Sector

There is a growing body of evidence of the impacts of informal education on learners, e.g., valuing science and the natural environment, increasing self-efficacy, making scientifically influenced decisions and developing 21st century skills. Knowledge about informal learning experiences is published in the fields of science education and has become increasingly accessible to the STEM research community in repositories such as InformalScience.org. Within the efforts of enabling engagement within STEM subjects, interactive technologies embedding AR (Matcha & Rambli, 2013) have already been applied in primary schools, although not within a formal curriculum setting but rather as investigatory, experimental pilots (Fuchsová, Adamková, & Lapšanská, 2019; Ross, 2019). The main elements associated with current pilots of AR in STEM subjects focus on evaluating students' experiences through comparative studies on specific tasks, utilizing the AR equipment (Ibáñez & Delgado-Kloos, 2018). Most silo products engage the students with 3D content either through marker/markerless applications and/or tangible objects, such as the following AR tools for classrooms:

- 3DBeaR AR: Teachers can create and assign lessons through a web-based dashboard, and students use the app to create scenes. The subjects associated within this tool are: Arts, Science, Creativity (https://www.commonsense.org/education/app/3dbear-ar). Contents and scenes are imported from Thingiverse (https://www.thingiverse.com/).
- CoSpaces Edu: Students can advance their coding skills through the creation of virtual 3D worlds. Teachers create a class and post assignments. Images and 360-degree photos can be uploaded (https://cospaces.io/edu/).
- Froggipedia: Students can use this biology-based app to explore the life cycle and anatomy of a frog through a guided virtual dissection (https://apps.apple.com/us/app/froggipedia/id1348306157).
- JigSpace: Students can view 3D objects (3D heart, solar system etc.) through the content library and each "Jig" is a 3D presentation of how everyday things work, explained in simple steps (https://jig.space/).
- MERGE Cube: Experiences through a hologram. With a MERGE Cube, teachers can create lessons and activities to explore STEAM concepts, illustrate complex systems, and enable students to "experience" history or science (https://mergeedu.com/educators).
- Moatboat: The student provides simple commands through the creation engine app for augmented reality and virtual reality and the creations can then be "placed" on a table to share with others (http://www.moatboat.com/).
- CleverBooks: AR apps for STEM subjects, where the student utilises AR apps and tangible objects associated with each subject with limited educational content and no access to content library and no available collaborative functionality (https://www.cleverbooks.eu/).

Current studies apply similar design features based on simple interactions with the digital objects to achieve knowledge discovery (Ibáñez & Delgado-Kloos, 2018). However the evaluation is not provided based on a digital assistance; hence, most of the studies reviewed evaluated the effects of augmented reality technology in fostering students' conceptual understanding, followed by those investigating affective learning outcomes (Ibáñez & Delgado-Kloos, 2018). Although these silo products have been evaluated with impact analysis, primary studies focused on the evaluation metrics. Supported initiatives are required to address the teacher preparation for curriculum design to augment STEM subjects at primary education. There is no current work to facilitate the interception of educational pedagogy in the full spectrum of the learning outcomes of the course to evaluate the impact of augmented reality in the learning process for students in primary school.

The main research gap in the literature review focuses on the fact that current AR products underperform in terms of updates in the educational market since the evaluation is taking place within nonstandard performance attributes. There are a number of future directions for the application of AR in STEM subjects:

- Design features within the AR facilitating the acquisition of STEM competencies through collaborative learning,
- Inclusion of experimental support for inquiry-based learning activities supported from a cloud-based 3D content library,
- Re-usable and interoperable 3D learning objects reachable from teachers and students to enable cross-disciplinary data intensive science.

Current Evaluation of AR in STEM Education

The potential of Augmented Reality within formal and informal settings has accelerated over the last 15 years, with the highest number of case studies in the literature review focused on promoting interactive experiences within STEM fields (Science, Technology, Engineering), followed by applications in humanities, medicine and health (for the years 2007-2019; Da Silva, Teixeira, Cavalcante & Teichrieb, 2019). The majority of the studies evaluated knowledge retention and/or performance and usability (users' attitudes and satisfaction). Other parameters have been monitored and assessed in few studies only (cognitive skills, behavior/motivation, learning styles, creativity, teaching effects, opinions, flow experience; Ibáñez & Delgado-Kloos, 2018).

The evaluation of AR in STEM applications is closely related to the type of activities offered from the AR educational tool, and current literature shows that most applications offer exploration or simulation activities (Ibáñez & Delgado-Kloos, 2018). The design features associated with such applications are based on specific interactions with 3D digital learning objects (Mangina, 2017). In some studies assistance is offered through the application to enhance the concepts' understanding and achieve effective learning outcomes (Ibáñez & Delgado-Kloos, 2018).

The main outcomes investigated so far for the evaluation of AR in STEM focus on quasi-experimental studies (Bursztyn, Walker, Shelton, & Pederson, 2017; Estapa & Nadolny, 2015), experimental design (Chen, Chou, & Huang, 2016; Ibáñez, Di Serio, Villarán, & Kloos, 2014), and interpretative studies (Cuendet, Bonnard, Do-Lenh, & Dillenbourg, 2013; Akçayır, Akçayır, Pektaş, & Ocak, 2016) with emphasis on fostering effects on satisfaction, motivation, enjoyment, attitude and remembering of information. There is little evidence so far on how the AR-based interventions support standardised instructional strategies. The current AR based applications do not include the learning goals and the characteristics (inclusion/exclusion) of students, that might affect the instructional design approach.

Along with the technology advancement, more capabilities of AR applications need to be investigated, such as collaborative learning. With the pandemic enforcing online learning, academics and students have realised the importance of peer-learning within the learning process and advancements have been made towards AR and VR Collaborative Learning (Simmons School of Education & Human Development, n.d.). Introducing AR within specific tasks in the curriculum (Petrov & Atanasova, 2020) affects the pedagogical approach as well as the evaluation criteria as it provides the learner with control of the learning process while utilising the ability to repeat a learning process and encourage reflection on action.

Evaluation Design

The project ARETE addresses the desiderata identified in the literature review, e.g. with regards to a comprehensive evaluation approach taking into account outcomes of AR applications as well as pedagogical implications. The focus of ARETE pilot 2 is on advancing European primary students' skills and knowledge in STEM learning. Against the background of the findings described, an intervention has been designed to test and evaluate two AR-enhanced apps in approximately 160 classrooms from eleven European countries. The comprehensive intervention does not only provide access to suitable apps but also provides pedagogical guidance. The two apps in the study have been developed for application in Mathematics and Geography lessons in 4^{th} and 5^{th} grade to support and enhance teaching and learning processes. The mathematics app presents 2D and 3D geometric shapes in Augmented Reality and facilitates spatial visualisation. The geography app adds AR features to maps and fosters visualisation e.g. of weather, the earth, or plants. Against the background of the state of the art outlined above, the apps represent common elements such as exploration and inquiry and also encourage collaborative learning activities.

In the intervention, teachers include one of the two apps in their regular classes according to their own pedagogical implication scenarios. The research question for this intervention is: "Can an Augmented Reality-app help students improve their test scores in STEM by 33 % and their retention rate by 100 %? ".

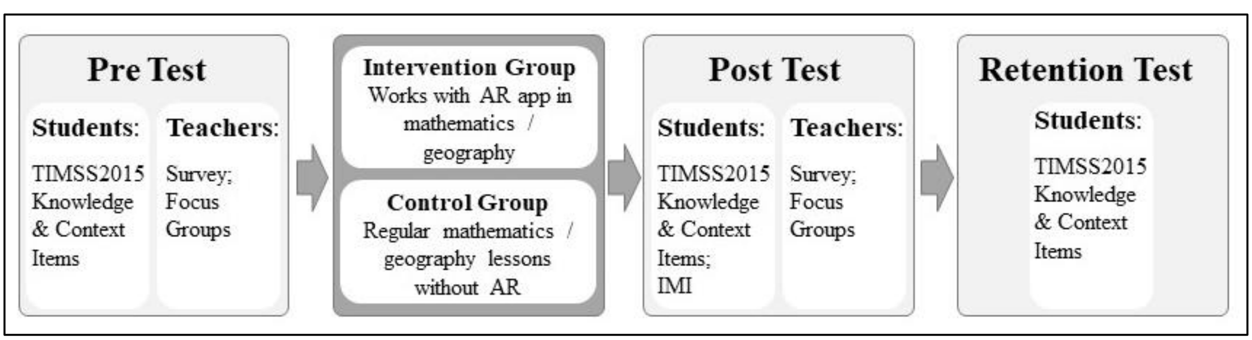

Figure 1: Evaluation design process for STEM pilot in the ARETE H2020 project

In the intervention evaluation, data are collected to assess these effects of the AR-apps. Additional qualitative data also clarify pedagogical and practical implications of AR use in STEM classes. In accordance with these research objectives, both students and teachers are addressed in the multimethod research design (Dünser, Grasset, & Billinghurst, 2008; Da Silva et al., 2019). For this purpose, a quasi-experimental approach with intervention and control group has been designed (cf. fig. 1).

Students

Metrics

In the following the different measures of the evaluation approach are explained in further detail.

1) Content knowledge test

To test for an increase in student knowledge soundly, it takes a standardized measurement instrument which is applied pre and post intervention both in an intervention and control group. In ARETE pilot 2, the respective items are taken from TIMSS 2015 (Trends in International Mathematics and Science Study). TIMSS 2015 is a well established international large scale study. It includes comprehensive scales for mathematics and science in fourth and eighth grade (Mullis & Martin, 2013). TIMSS 2015 was applied in 57 countries and translated into 43 languages, which makes it a powerful international comparative instrument. Reliabilities were found to be relatively high; for the fourth grade items, the median of the reliability coefficients for all countries was .83 for mathematics and .78 for science. Also, scoring reliability was confirmed to be high with 96 % or above within-country, and 86 % or higher internationally for fourth grade scales (Martin, Mullis, & Hooper, 2016).

A range of items was selected for use in the ARETE study. These items show a high congruency with the contents addressed by the two ARETE apps and have thus been defined as suitable for assessing the respective knowledge gain.

2) Background questionnaire

The background questionnaire enrichens the data collection by important correlating information on student context level. In addition to demographic questions, it presents a self-assessment scale including questions on media equipment and habits of media use, and items assessing beliefs and perceived self-efficacy concerning mathematics/mathematics lessons and science/science lessons (IEA, 2019).

3) Motivational scale

The student post-test survey also includes a motivational scale to identify important motivational implications of the AR-supported teaching and learning scenarios. Considering students' motivation when using AR is relevant because motivation is an important predictor for academic achievement (Steinmayr, Weidinger, Schwinger, & Spinath, 2019).

The items selected for the ARETE pilot 2 study are taken from the Intrinsic Motivation Inventory (IMI), which is a well-established measure providing a Likert scale for students to self-assess their motivational perceptions during an intervention across seven subscales (Decy & Ryan, 2003). It has been applied successfully in other AR-related educational studies before (Sotirious & Bogner, 2008; Buchner & Zumbach, 2018; Hanafi, Said, Wahab, & Samsuddin, 2017). In accordance with the items from TIMSS 2015 which are available in all test languages, this IMI item selection will be translated into all native languages for the pilot to ensure a high validity of responses.

Sample

The convenience sample of students will include approximately 3000 participants, half of which will work on mathematics while the other half will work on science lessons. Due to the intervention and control group design, four subsamples result with approximately 750 students each. The intervention groups will work with the mathematics or science app while the control groups will proceed with their regular classes. Inclusion into the sample depends on the teachers who apply with their classes. To be eligible, students must be in 4^{th} or 5^{th} grade and 7 to 12 years old. They need access to suitable devices and broadband internet.

Teachers

Metrics

Teachers have a significant impact on their students' achievements. Their attitudes and beliefs towards digital media, their educational practice and experience and their digital educational competencies frame the conditions for students' knowledge acquisition (Mishra & Koehler, 2006; Redecker, 2017). Hence, it is relevant to include the teachers' perspectives when assessing the effects of an intervention.

For this purpose, surveys have been developed in ARETE assessing intervention and control group settings and pedagogical implications of the Augmented Reality application. This includes questions e.g. on factors that hinder or facilitate a successful implementation, on group settings when using the app, and on teacher attitudes and beliefs towards Augmented Reality in teaching (Granić & Marangunić, 2019). Additional focus groups with teacher coordinators serve to deeper explore the aspects addressed in the surveys pre and post intervention.

Sample

The convenience sample of teachers consists of approximately 170 participants, with 85 teachers teaching mathematics and 85 teaching science. Teachers are recruited pairwise, i.e., two teachers per school apply together who both teach the same subject (mathematics or science). Together with their classes, one of these two is randomly allocated to the intervention group sample while the other teacher is assigned to the control group. This pairwise recruitment serves to ensure comparability of the school curricula of intervention and control groups.

To be eligible for inclusion into the sample, teachers must teach a 4^{th} or 5^{th} grade either in mathematics or science, be able to provide access to suitable devices for his or her students, and have a second teacher from his or her school to team up with for application. For the application all European countries were eligible. For the final sample, applications were accepted from Spain, Croatia, Romania, Portugal, Greece, Poland, Serbia, Italy, Ireland, Sweden and Turkey. Due to organisational conditions of the study, the sample is a self-selecting sample with teachers applying for the pilot study voluntarily.

Out of the group of teachers, eight teachers additionally work as teacher coordinators. They function as contact persons for the teachers in their country or language group and help disseminating information, organizing and troubleshooting to ensure a reliable and failure-free pilot administration.

Discussion

Based on the literature review, the research approach introduced within this paper has been designed to generate valid conclusions on the effects AR can have on student STEM knowledge acquisition and retention and on teaching and learning processes in general. For the interpretation of data collected with this approach, it will be important to consider certain limitations and caveats: for example, the nature of the self-selecting sample poses limitations to the generalizability and transferability of results because certain types of teachers might apply disproportionately often and generate biased results. Also, the spread of participants per country (socio economic background of participants) and per region in Europe is not equally distributed. This is also related to challenges in the recruitment of participants due to the impact of the COVID19 pandemic. Regarding the measurement, further correlating variables beyond knowledge, context and motivation would be interesting to consider but had to be excluded to keep the questionnaire appropriate for the age group targeted.

Yet, it can be assumed that the variety of international participants and the sample size will yield relevant results and advance this important field of research on Augmented Reality applications for STEM education. An interpretation of results will allow for conclusions e.g. on the learning achievements of students from a specific country after the intervention with the Augmented Reality app, on the correlation between certain pedagogical impact factors and success of the implementation, or on student motivation for STEM learning with an AR app.

It will be desirable also for future studies in this field to consider the multi-perspective approach applied and to include different viewpoints and methods for a thorough and comprehensive evaluation of Augmented Reality in education. The international comparative perspective contributes to this objective and enhances the applicability of conclusions to a wider European target group.

Acknowledgements

This research has been supported by the European Union's Horizon 2020 research and innovation program under grant agreement No. 856533, project ARETE.

References

Akçayır, M., Akçayır, G., Pektaş, H.M., & Ocak, M.A. (2016). Augmented reality in science laboratories: The effects of augmented reality on university students' laboratory skills and attitudes toward science laboratories. *Computers in Human Behavior, 57,* 334–342.

Buchner, J., & Zumbach, J. (2018). Promoting Intrinsic Motivation with a Mobile Augmented Reality Learning Environment. *14th International Conference Mobile Learning*, Lisbon, Portugal, 55–61.

Bursztyn, N., Walker, A., Shelton, B., & Pederson, J. (2017). Increasing undergraduate interest to learn geoscience with GPS-based augmented reality field trips on students' own smartphones. *Geological Society of America Today, 27*(5), 4–11.

Chen, C. H., Chou, Y. Y., & Huang, C. Y. (2016). An augmented-reality-based concept map to support mobile learning for science. *The Asia-Pacific Education Researcher, 25*(4), 567–578.

Cuendet, S., Bonnard, Q., Do-Lenh, S., & Dillenbourg, P. (2013). Designing augmented reality for the classroom. *Computers & Education, 68,* 557–569.

Deci, E. L., & Ryan, R. M. *Intrinsic Motivation Inventory*. [Website]. Retrieved from https://selfdeterminationtheory.org/intrinsic-motivation-inventory/#:~:text=Scale%20Description,target%20activity%20in%20laboratory%20experiments

Dünser, A., Grasset, R., & Billinghurst, M. (2008). A survey of evaluation techniques used in augmented reality studies. *ACM SIGGRAPH ASIA 2008 courses*, Singapore, 1–27.

Estapa, A., & Nadolny, L. (2015). The effect of an augmented reality enhanced mathematics lesson on student achievement and motivation. *Journal of STEM Education: Innovations and Research, 16*(3), 40–48.

Fuchsová, M., Adamková, M., & Lapšanská, M. P. (2019). Chapter 8: Uses of Augmented Reality in Biology Education. In T. Prodromou (Ed.), *Augmented Reality in Educational Settings* (pp. 168–194). Leiden, The Netherlands: Brill | Sense.

Granić, A. & Marangunić, N. (2019). Technology acceptance model in educational context: A systematic literature review. *British Journal of Educational Technology, 50*(5), 2572–2593.

Hanafi, H. F., Said, C. S., Wahab, M. H., & Samsuddin, K. (2017). Improving Students' Motivation in Learning ICT Course With the Use of a Mobile Augmented Reality Learning Environment. *IOP Conf. Series: Materials Science and Engineering, 226,* 012114.

Ibáñez, M. -B. & Delgado-Kloos, C. (2018). Augmented reality for STEM learning: A systematic review. *Computers & Education, 123,* 109–123.

Ibáñez, M. -B., Di Serio, Á., Villarán, D., & Kloos, C. D. (2014). Experimenting with electromagnetism using augmented reality: Impact on flow student experience and educational effectiveness. *Computers & Education, 71,* 1–13.

International Association for the Evaluation of Educational Achievement [IEA] (2019). *TIMSS 2015 Context Questionnaires*. [Website]. Boston, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College, and International Association for the Evaluation of Educational Achievement. Retrieved from https://timssandpirls.bc.edu/timss2015/questionnaires/index.html

Johnson, C. C., Peters-Burton, E. E., & Moore, T. J. (Eds.) (2016). S*TEM Road Map. A Framework for Integrated STEM Education*. New York, NY: Routledge.

Mangina, E. (2017). 3D learning objects for augmented/virtual reality educational ecosystems. *2017 23rd International Conference on Virtual System & Multimedia (VSMM),* Dublin, Ireland, 1–6. doi: 10.1109/VSMM.2017.8346266.

Matcha, W. & Rambli, D. R. A. (2013). Exploratory study on collaborative interaction through the use of Augmented Reality in science learning. *Procedia Computer Science, 25,* 144–153.

Martin, M. O., Mullis, I. V. S., & Hooper, M. (Eds.) (2016). *Methods and Procedures in TIMSS 2015*. Boston, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College, and International Association for the Evaluation of Educational Achievement (IEA). Retrieved from https://timssandpirls.bc.edu/publications/timss/2015-methods.html

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for integrating technology in teachers' knowledge. *Teachers College Record, 108*(6), 1017–1054.

Mullis, V. S., & Martin, O. (Eds.) (2013). *TIMSS 2015 Assessment Frameworks*. Boston, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College, and International Association for the Evaluation of Educational Achievement (IEA). Retrieved from https://timssandpirls.bc.edu/timss2015/frameworks.html

Petrov, P.D. & Atanasova, T. V. (2020). The Effect of Augmented Reality on Students' Learning Performance in Stem Education. *Information, 11*(4), 209. https://doi.org/10.3390/info11040209

Redecker, C. (2017). *European Framework for the Digital Competence of Educators: DigCompEdu*. Y. Punie (Ed.). Luxembourg, Luxembourg: Publications Office of the European Union.

Ross, T. (2019, February 25). Top 10 Augmented Reality Tools for the Classroom. *EBSCOpost*. [Website]. Retrieved from https://www.ebsco.com/blogs/ebscopost/top-10-augmented-reality-tools-classroom

Da Silva, M., Teixeira, J., Cavalcante, P., & Teichrieb, V. (2019). Perspectives on how to evaluate augmented reality technology tools for education: a systematic review. *Journal of the Brazilian Computer Society, 25*(3), 1–18.

Simmons School of Education & Human Development (n.d.). AR and VR can support experiential and collaborative learning. *Augmented Reality and Virtual Reality for STEM Collaborative Learning. The Future of Immersive Learning*. [Website]. Retrieved from https://www.smu.edu/Simmons/Research/Center-for-Virtual-Reality-Learning-Innovation/STEM-Collaborative-Learning

Sotirious, S., & Bogner, F. X. (2008). Visualizing the Invisible: Augmented Reality as an Innovative Science Education Scheme. *Advanced Science Letters, 1,* 114–122.

Steinmayr, R., Weidinger, A. F., Schwinger, M., & Spinath, B. (2019). The Importance of Students' Motivation for Their Academic Achievement – Replicating and Extending Previous Findings. *Front. Psychol., 10,* art. 1730.