Augmented Reality in Informal Learning Environments: Investigating Short-term and Long-term Effects

Peter Sommerauer
IT University of Copenhagen
psom@itu.dk

Oliver Müller
IT University of Copenhagen
olmy@itu.dk

Abstract

While many researchers have qualitatively examined the affordances and constraints of AR in educational settings, only few studies exist that tried to quantify the effect of AR on learning performance. To contribute to filling this research gap, we conducted a pretest-posttest-posttest crossover field experiment with 24 participants at a mathematics exhibition to measure the effect of AR on acquiring and retaining mathematical knowledge in an informal learning environment, both short-term (i.e., directly after visiting the exhibition) and long-term (i.e., two months after the museum visit). Our empirical results show that museum visitors performed significantly better on knowledge acquisition and retention tests related to augmented exhibits than to non-augmented exhibits directly after visiting the exhibition (i.e., short-term), but this positive effect of AR vanished in the long run.

1. Introduction

Augmented reality (AR) dynamically blends real world environments and context-based digital information [1]. Recent advancements in mobile computing made AR systems affordable for the broad public. Such mobile AR applications use cameras, GPS sensors, and Internet access of mobile devices to overlay real-world environments with dynamic, context-based, and interactive digital content.

It has been argued that education is one of the most promising application areas for AR [19]. The NMC Horizon Report 2016 identified AR as a technology to bring new opportunities for learning and to offer compelling applications for higher education; AR is especially expected to empower students in STEM (Science, Technology, Engineering, Mathematics) disciplines to engage in deep learning and prepare them for the future workplace [13]. Nonetheless, in their literature review on AR for teaching and learning Dunleavy and Dede [4] stated that "[d]ue to the nascent

and exploratory nature of AR, it is in many ways a solution looking for a problem" (p. 26) and that "relatively few research and development teams are actively exploring how mobile, context-aware AR could be used to enhance K-20 teaching and learning" (p. 8). Up to date, most empirical research on AR for teaching and learning is of a qualitative nature and focuses on exploring the affordances and constraints of AR. So far, relatively few quantitative studies exist that tried to measure the effect of AR on learning outcomes (exceptions include, e.g., [16, 5, 7]).

In order to address the current gap in the body of knowledge on AR for education, we conducted a field experiment in a mathematics exhibition to test the effect of AR on learning performance, both short-term and long-term. Our study was driven by the hypothesis that museum visitors learn better from museum exhibits enriched through AR than from exhibits that are accompanied by traditional physical information displays only (e.g., info boards, posters). The theoretical foundation for this hypothesis is based upon the cognitive theory of multimedia learning (CTML). More specifically, we posit that AR implements a subset of the design principles formulated in CTML, namely, the multimedia principle, the spatial and temporal contiguity principles, the modality principle, and the signaling principle.

The results of our experiment provide evidence for the short-term effectiveness of AR as a tool for supporting learning. Directly after the museum visit, participants were able to retain significantly more knowledge about augmented exhibits than about non-augmented exhibits. However, the advantage of AR over traditional learning materials disappeared when re-testing participants two months after the museum visit, pointing to the need for more research on the design of AR learning materials for supporting sustainable and deep learning experiences.

This paper is structured as follows: We first provide the theoretical background on the cognitive theory of multimedia learning and on the differences of short-term and long-term memory. We then outline our experimental design and present the analysis of our

HⱡCSS

empirical results. After discussing our results, we conclude with a brief summary and directions for future research.

2. Theoretical background

2.1. Cognitive Theory of Multimedia Learning

The cognitive theory of multimedia learning (CTML) provides potential explanations why AR may improve learning. In broad terms, CTML posits that people learn better from words and pictures than from words alone [10, 11]. CTML is based on three assumptions. First, humans possess two channels for processing information, an auditory/verbal channel and a visual/pictorial channel [14]. Second, each channel can process only a limited amount of information at one time [17]. Third, learning is an active process consisting of selecting relevant incoming information, organizing selected information into coherent mental representations, and integrating mental representations with existing knowledge [18]. Based upon these theoretical assumptions, CTML postulates principles for the design of effective multimedia instructions [11]. We argue that AR, designed and applied in the right way, inherently incorporates a subset of these design principles, namely, the (1) multimedia principle, (2) the spatial contiguity principle, (3) the temporal contiguity principle, (4) the modality principle, and (5) the signaling principle.

The multimedia principle states that people learn better from words and pictures than words alone. AR can implement this principle by overlaying printed texts with virtual pictorial content (e.g., integrating videos into a textbook) or, vice versa, by augmenting physical objects with virtual texts (e.g., displaying labels and measures when focusing on a technical object). The spatial and temporal contiguity principles state that learning is enhanced when the space and/or time between disparate but related elements of information is minimized. AR can implement the contiguity principles by superimposing virtual content onto physical objects in real-time and thereby spatially and temporally aligning related physical and virtual information. The modality principle states that learning can be enhanced by presenting textual information in an auditory format, rather than a visual format, when accompanying related visual content. AR can implement the modality principle by playing spoken text, instead of displaying printed text, when recognizing a trigger event. Finally, the signaling principle states that people learn better when cues highlight the organization of essential information in a learning environment. AR can implement signaling by directing and guiding people through learning environments using geographic location information and visual triggers.

2.2. Short-term and long-term memory

It is commonly accepted that three different types of human memory can be distinguished, namely, sensory memory, short-term memory, and long-term memory (for an overview of models of human memory see, e.g., [2]). External stimuli enter the human memory system through the sensory stores, which are characterized by their preattentive, modality-specific, and transient nature. If a subject pays attention to the information entering the sensory storage, it can be transferred to the short-term storage (STS), also known as working memory. Compared to the sensory storage, the STS has a much more restricted capacity but also a slower rate of forgetting. Through repeated rehearsal information can be transferred from the STS to the long-term storage (LTS), which has no known capacity limits. Compared to the STS, in which verbal information is coded phonemically, it is assumed that information in the LTS is stored largely semantically and maintained through repetition, organization, and integration with prior knowledge. CTML is largely based on the multi-store model of human memory (Figure 1). By representing information in efficient formats multimedia technologies bear the potential to overcome the capacity limitations of our working memory and thereby enable more effective short- and long-term learning.
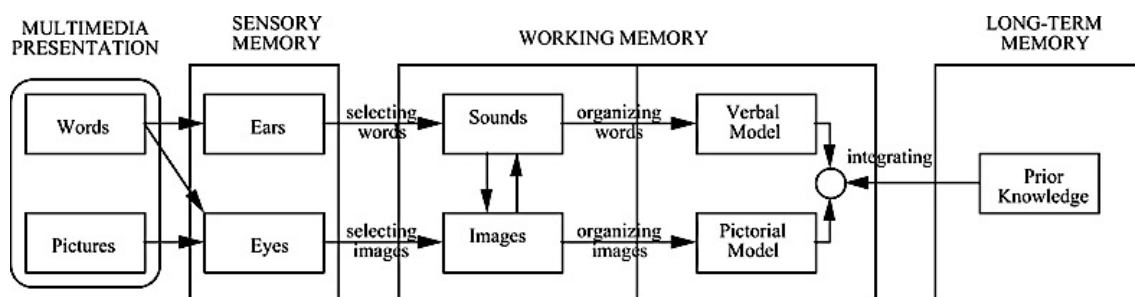

Figure 1. Overview of CTML [10]

3. Experimental design

3.1. Setup and participants

The objective of our study was to investigate whether AR is an effective educational technology in informal learning environments, both in the short run and in the long run. Consequently, the hypothesis underlying our study was that museum visitors learn better from augmented exhibits than from non-augmented exhibits.

We chose to conduct a framed field experiment [6], in which natural subjects (i.e., visitors) performed natural tasks (i.e., engaging with exhibits) in a natural place (i.e., museum). The only artificial component in the experimental setup was the fact that participants were aware that they are taking part in an experiment and that their behavior is recorded and analyzed. The field experiment was designed as a crossover study [8, 12], that is, participants received a series of different treatments over time (i.e., augmented and non-augmented exhibits) so that each participant could serve as its own control, thereby eliminating potential bias caused by between-subject variability. To rule out carryover and order effects, we designed experimental tasks that were logically and temporally independent of each other and let participants roam through the exhibition and complete tasks at their own order and pace.

Figure 2 graphically summarizes the design of the experiment. Participants were randomly assigned to one of two groups and had 15 minutes to take a pretest with 16 questions regarding the mathematical exhibits they were later to see. Participants in both groups were then given 90 minutes to visit the mathematics exhibition individually and at their own pace. Before entering the exhibition, all participants received a short hands-on training how to use the mobile AR app on their own devices in order to discover and activate hidden virtual contents within the exhibition. In addition, ten pre-configured iPads, two smartphones and various headsets were offered to those participants that had problems to get the application running on their own device. The same test, plus additional questions on demographics, was administered to all participants as a posttest directly after visiting the exhibition and in addition two months after the museum visit (participants were not told that the same questionnaire is used for the posttests).

The exhibition consisted of four separate rooms covering eight mathematical topics with a total of 275 exhibits. All objects of the exhibition were accompanied by traditional physical information displays (i.e., boards, posters, leaflets, quizzes, books, screens). For twelve exhibits, we created additional virtual augmentations, six accessible for participants in Group 1 and six accessible for participants in Group 2. All twelve augmented exhibits were tagged with markers.

We recruited a class of 26 pupils (K-20) and their mathematics teacher to take part in the experiment. The group consisted of two female and 24 male students and one male teacher. The students were between 15 and 18 years old and the teacher was 62 years old. While all 26 pupils participated in the first part of the experiment, only 23 could attend the long-term post-test session. Hence and due to the crossover design of the experiment [8], in which all participants receive to all treatments, both experimental groups contained 23 participants.

Random Assignment		Measurement	Treatment												Measurement	
			E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8	E_9	E_{10}	E_{11}	E_{12}		
	Group 1	Pretest Score Augm. Exhibits	AR	AR			AR			AR	AR		AR		Short-term Posttest Score Augm. Exhibits	Long-term Posttest Score Augm. Exhibits
	Group 2	Pretest Score Non-Augm. EXhibits			AR	AR		AR	AR			AR		AR	Short-term Posttest Score Non-Augm. Exhibits	Long-term Posttest Score Non-Augm. Exhibits

E_n: Exhibit n

Figure 2. Overview of Experimental Design

3.2. Treatments

We used Aurasma Studio (Version 2.0) to design augmentations for twelve selected exhibits. Nine objects were augmented with videos (incl. audio) in which the curator explained and demonstrated the mathematical exhibits, three objects were augmented with animations of the mathematical phenomenon described in the exhibit (Figure 2). The length of the augmentations varied between 60 and 252 seconds. Visitors used the Aurasma mobile app running on their own mobile devices like smartphones or on iPads to discover und unlock augmentations by pointing the tablet's camera at exhibits and trigger images. All devices were equipped with headphones to allow listening to sound without disturbing other visitors. We manipulated the treatments by assigning each augmentation to only one of the two experimental groups. Thereby we ensured that for each exhibit half of the participants were able to access the augmented virtual content and the other half had to rely on the physical information displays only. We used the channel concept of Aurasma to implement the grouping of participants and treatments.

As outlined in the background section, we argue that AR enables the efficient and effective implementation of a subset of the design principles stated in the cognitive theory of multimedia learning. For example, we incorporated the multimedia principle into the AR materials by explaining the mathematical concepts of an exhibit through rich motion pictures, that is, animations and videos, instead of static graphics and texts. For instance, while the physical information display for Exhibit 9 (Linear and exponential growth) illustrated exponential growth through a number series (2, 4, 8, 16, 32, 64, 128, …), the corresponding AR experience showed an animation of the wheat and chessboard problem using time-lapse and zooming features. The spatial contiguity principle was implemented by superimposing digital information onto physical exhibits. This removes the need to visually search the environment of an exhibit for explanatory information.

For example, in the AR experience of Exhibit 7 (The various nets of a cube's surface) the animation unfolded directly on top of the trigger image, while participants in the non-AR group for this exhibit had to spent cognitive resources to constantly switch their visual focus between a model of a cube and surrounding models of its eleven possible nets, and had to integrate these disparate information sources. In a similar vein, we used spoken narration by the curator to provide information about an exhibit at the same time at which the visitor is focusing on the exhibit, thereby implementing the temporal contiguity principle. Visitors in the control group, in contrast, had to decide whether to first take a look at the exhibit and then read

through the accompanying information, or vice versa, and then needed to integrate both types of information into one congruent mental model. This simultaneous visual and auditory information provisioning is also in line with the modality principle of CTML, which states that people learn better from animations with spoken narration than from animations with on-screen text. Finally, we implemented the signaling principle within and across AR experiences. Within individual AR 14 experiences, we inserted headings for subsections in order to give structure to videos and animations. Across the whole exhibition, we chose to augment only selected exhibits with AR in order to organize the overall museum visit and highlight the most important objects of each part of the exhibition. Table 1 provides an overview of the exhibits and AR experiences.

Table 1. Exhibits and AR experiences

Exhibit	Group	Exhibit and topic	AR Experience
1	1	Interactive model of a cycloid constructed of a three-lane marble track	*Video* in which the curator explains and illustrates that a cycloid has the properties of a tautochrone curve
2	1	Interactive model of a cycloid constructed of a three-lane marble track	*Video* in which the curator explains and illustrates that a cycloid has the properties of a brachistochrone curve
3	2	Interactive model of a hyperboloid constructed of strings	*Video* in which the curator explains why the cooling towers of nuclear power plants are constructed in the form of hyperboloids
4	2	Interactive model of a hyperboloid that is used for plugs in aircrafts; real aircraft plugs	*Video* in which the curator explains why a hyperboloid form guarantees full galvanic isolation of plugs
5	1	Interactive model of a double cone on a diverging monorail	*Video* in which the curator shows that a double cone on a diverging monorail seemingly rolls upwards

6	2	Explanation of the approximation of Pi in an annexed book and on exercise sheets	*Video* in which the curator explains how to approximate Pi by tying a rope around the earth's equator
7	2	Physical models of a cube and the various nets of its surface	*Animation* showing the unfolding of all different nets of a cube's surface
8	1	Interactive installation illustrating the attributes of a plain mirror; additional descriptions on exercise sheets	*Video* in which the curator illustrates the correlation between distance and height of the objects in the mirror
9	1	Illustration of linear/exponential growth through an interactive paper folding experiment and a representation of a exponentially growing number series on the steps of the entrance hall's stairs	*Animation* illustrating the exponential growth through the wheat and chessboard problem
10	2	The Monty Hall problem explained in book in the exhibition's reader's corner	*Animation* explaining the Monty Hall paradox
11	1	Fully functional exemplar of the Arithmometré mechanical calculator from Thomas de Colmar in a glass cabinet	*Video* in which the curator explains and demonstrates the functionalities of the Arithmometré calculator
12	2	Fully functional exemplar of the Heureka mechanical calculator in a glass cabinet	*Video* in which the curator explains and demonstrates the functionalities of the Heureka calculator

A key challenge when designing AR materials for experimental treatments is the issue of informational equivalence. According to Larkin and Simon [9], two representations are informationally equivalent if all the information from one representation can also be inferred from the other representation, and vice versa. On the one hand, informational equivalence is clearly a desirable feature for controlled laboratory experiments on educational technologies as it ensures that differences in effects stem from the mode of representation and not from the content of a representation. On the other hand, we argue that when designing realistic AR experiences it is difficult to achieve full informational equivalence without undermining the affordances of AR. For example, transcribing all spoken information of a two minutes AR experience would lead to long texts that no museum visitor would read, and, vice versa, transforming all information contained in the physical displays accompanying an exhibit in a science museum into AR would lead to overloaded AR experiences. Therefore, we designed AR materials that overlapped, rather than were equivalent, with physical information displays. Following the guidelines regarding informational equivalence in experimental studies given by Parsons and Cole [15], our questionnaire was then designed in a way that it was "possible to answer [all] questions correctly with any of the representational forms used as treatments in [the] experimental study" (p. 330). This way, we ensured that both learning experiences were "educationally equivalent", that is that they support the same learning objectives. Figure 3 illustrates this approach graphically.
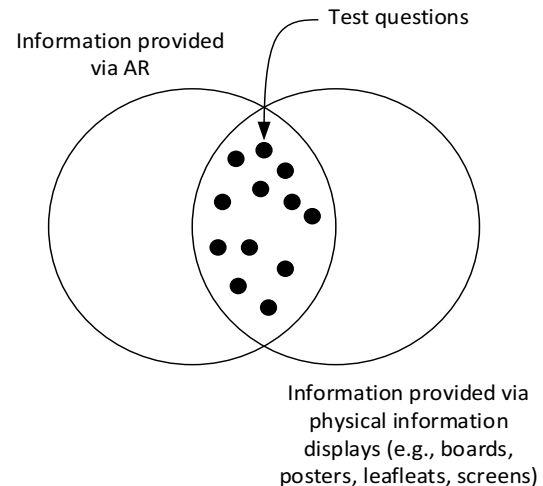
Figure 3. Alignment of information provided via AR, information provided via physical information displays and test questions

3.3. Measures

Following related experimental studies on the use of AR in education, we focused on knowledge retention as a measure of learning performance using a pretest-posttest measurement approach. All test questions were

single-choice questions. In the selection and design of the test questions we paid special attention that all question could be answered through both the virtual augmentations of the exhibits and the physical information displays accompanying the exhibits. We created one test question for each of the twelve exhibits being part of the experiment. We selected questions that were adaptations of well-known mathematical problems, for example: "What is the fastest descent between two points that are not above each other? A) Slope B) S-Curve C) Circular arc D) Cycloid" or "How tall a mirror do you need to see yourself? A) Half your height B) Two thirds of your height C) Equal to your height D) Twice your height". To establish content validity the curator of the exhibition, who was a retired mathematics high school teacher, reviewed all questions.

We aggregated the answers to the individual questions to six test scores (Figure 2). The pre-test score for augmented objects and the pretest score for non-augmented objects captured the level of previous knowledge regarding the mathematical exhibits. The short-term and long-term posttest scores for augmented objects and non-augmented objects captured the knowledge level after visiting the exhibition. The possible values of all scores ranged between 0 and 6.

4. Results

4.1. Descriptive statistics

Table 2 gives an overview of the test scores. The low scores on the pretest (on average 2 out of 6) suggest that participants had only little prior knowledge about the topics covered in the exhibition. Even after the visit, participants answered only about half of the test question correctly.

Table 2. Descriptive Statistics

| Test	Measure	Non-Augmented Exhibits	Augmented Exhibits
Partici-pants	N	23	23
Pretest Scores	Mean	2.05	2.00
	Median	2	2
	SD	1.27	1.03
	Min	1	0
	Max	5	4
Short-term Posttest Scores	Mean	2.59	3.42
	Median	3	4
	SD	1.34	0.65
	Min	1	0
	Max	5	5
Long-term Posttest Scores	Mean	3.04	2.88
	Median	3	4
	SD	1.11	1.68
	Min	1	0
	Max	5	5

4.2. Hypothesis tests

Usually, the statistical analysis of paired pretest-posttest data is done via paired t-tests or a repeated measures analysis of variance [3]. Yet, a Kolmogorov-Smirnov test indicated that the required assumption of normality for the dependent variables of the experiment was violated. Hence, we used the equivalent non-parametric Wilcoxon signed-rank test for statistical hypothesis testing. We conducted Wilcoxon signed-rank tests on various combinations of pretest scores, short-term posttest scores, and long-term posttest scores for augmented and non-augmented exhibits.

Figure 3 graphically summarizes the results of our statistical tests. First, to rule out that differences in test scores were caused by different levels of difficulty of question sets related to augmented and non-augmented exhibits we performed a Wilcoxon signed-rank test on the pretest scores of the two groups. The test showed no statistically significant differences in median scores between the two pretest-question sets ($p = 0.8649$).

Second, we compared medians of pretest scores and short-term posttest scores. For questions on both augmented and non-augmented exhibits participants showed significantly better performance at the short-term posttest than the pretest (AR: $p = 0.0005$, Non-AR: $p = 0.0516$). Hence, we can conclude that participants learned from both types of exhibits. However, the analysis also showed that participants performed significantly better for short-term posttest questions related to augmented exhibits (Mdn = 4) than for questions related to non-augmented exhibits (Mdn = 3), $p = 0.0482$. This result provides strong empirical evidence for the general effectiveness of AR as a tool for learning mathematical contents.

Finally, we analyzed the differences in test scores for the short-term (i.e., directly after the exhibition visit) and long-term (i.e., two months after the museum visit) posttests. Interestingly, we could not find significant differences in the performance of participants related to questions about augmented and non-augmented exhibits ($p = 0.6912$). This finding suggests that although participants remembered significantly more about AR exhibits than about non-AR exhibits directly after visiting the exhibition, AR seems to have no positive effect on long-term learning when comparing it with traditional, non-digital learning materials.
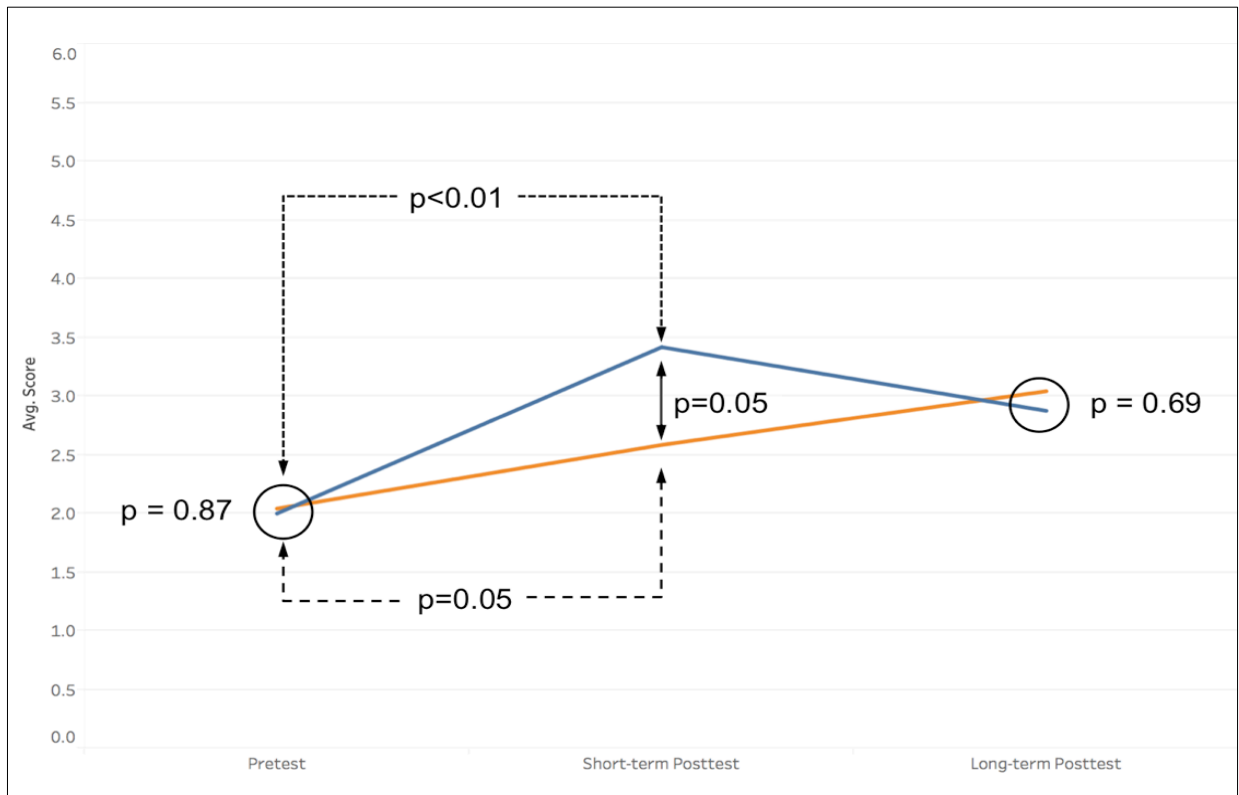
Figure 4. Results of Statistical Tests

5. Discussion and outlook

To the best of our knowledge, the here presented study is the first experiment on the effect of AR on learning mathematical contents that distinguished short-term and long-term retention effects.

The empirical evidence we gathered provides support for the proposition that AR has the potential to be an effective tool for acquisition and retention of formal contents in informal learning environments – at least for short-term learning. At the same time, our results suggest that AR is not necessarily more effective than traditional non-digital learning materials when it comes to long-term learning. One potential explanation for this surprising finding can be derived from the multi-store model of human memory. As outlined before, information first enters STS through increased and continued attention to stimuli from the preattentive sensory stores and is then stored in STS in phonetic form. By embodying the design principles of CTML, AR seems to effectively support both processes. The transfer of information from STS to LTS, in contrast, largely depends on a subject's ability to rehearse, semantically organize, and integrate the newly acquired information with prior knowledge. It seems that – at

least in our setting (i.e., informal learning of abstract mathematical content) – the AR materials designed by us did not effectively support these processes.

Even if we could add external validity in our field experiment through its realistic setting of the present experiment, many field experiments have to consider threats to internal validity. For example, we did not control the participants or their activities within their 90 minutes museum visit, which can be interpreted as potential confounding factor having influence on our results. Especially in self-directed learning settings as we used in our experiment, participants interactions with the exhibits differ in time and no of trials and we observed that some visitors paid more attention to either augmented or non-augmented exhibits issued in the pretest and posttest. However, this could be understood more as a positive effect derived from the use of technology rather than a threat.

Furthermore, another confounding factor results from the fact that a complete information equivalence of AR and non-AR materials could not be ensured. Thus, the AR materials were another representation of the exhibit's information. Finally, we asked the participants whether they visited the exhibition again between the two posttest activities which was denied by them, but

we could not proof this. Hence our results could be effected by multiple exhibition visits.

Since the participants are from the same class and the museum visit and the posttests were part of their lessons, further influence could result from the effect of learning from test situations [3, 8].

In conclusion, the present study contributes new insights on short-term and long-term learning with AR applications. In particular, it adds to the small but growing number of studies exploring the effective design for AR in teaching and learning. In order to confirm our findings, further studies on long-term retention should be carried out, particularly in other informal learning environments, for example, at workplaces.

6. References

[1] Azuma, R. 1997. "A survey of augmented reality", Presence: Teleoperators and Virtual Environments (6:4), pp. 355–385.

[2] Craik, F. I., & Lockhart, R. S. 1972. "Levels of processing: A framework for memory research", Journal of verbal learning and verbal behavior (11:6), pp. 671-684.

[3] Dimitrov, D. M., & Rumrill Jr, P. D. 2003. "Pretest-posttest designs and measurement of change", Work (20:2), pp.159-165.

[4] Dunleavy, M., & Dede, C. 2014. "Augmented reality teaching and learning", in The Handbook of Research for Educational Communications and Technology, J. Spector, M. Merrill, J. Elen, & M. Bishop (eds.), New York: Springer, pp. 735–745.

[5] Fonseca, D., Martí, N., Redondo, E., Navarro, I., & Sánchez, A. 2014. "Relationship between student profile, tool use, participation, and academic performance with the use of Augmented Reality technology for visualized architecture models", Computers in Human Behavior (31), pp. 434–445.

[6] Harrison, G., & List, J. 2004. "Field Experiments", Journal of Economic Literature (42:4), pp. 1009–1055.

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

[8] Johnson, D. 2010. "Crossover experiments", Wiley Interdisciplinary Reviews: Computational Statistics (2:5), pp. 620–625.

[9] Larkin, J., & Simon, H. 1987. "Why a diagram is (sometimes) worth ten thousand words", Cognitive Science, 11(1), 65–100.

[10] Mayer, R. E. 1997. "Multimedia learning: Are we asking the right questions?", Educational Psychologist (32:1), pp. 1–19.

[11] Mayer, R. E. 2009. Multimedia Learning (2nd ed.), Cambridge: Cambridge University Press.

[12] Mills, E., Chan, A., Wu, P., Vail, A., Guyatt, G., & Altman, D. 2009. "Design, analysis, and presentation of crossover trials", Trials (10:27), pp. 1–6.

[13] NMC. 2016. NMC Horizon Report > 2016 Higher Education Edition. Retrieved from http://cdn.nmc.org/media/2016-nmc-horizon-report-he-EN.pdf

[14] Paivio, A. 1990. "Mental representations: A dual coding approach", Oxford: Oxford University Press.

[15] Parsons, J., & Cole, L. 2005. "What do the pictures mean? Guidelines for experimental evaluation of representation fidelity in diagrammatical conceptual modeling techniques.", Data & Knowledge Engineering, 55(3), 327–342

[16] Sommerauer, P., & Müller, O. 2014. "Augmented reality in informal learning environments: A field experiment in a mathematics exhibition", Computers & Education (79), pp. 59-68.

[17] Sweller, J., Ayres, P., & Kalyuga, S. 2011. "Cognitive load theory", New York: Springer.

[18] Wittrock, M. 1992. "Generative learning processes of the brain," Educational Psychologist (27:4), pp. 531–541.

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