Sommerauer, P., & Müller, O. (2014). Augmented Reality in Informal Learning Environments: A Field Experiment in a Mathematics Exhibition. ***Computers & Education***

http://dx.doi.org/10.1016/j.compedu.2014.07.013

Augmented reality in informal learning environments: A field experiment in a mathematics exhibition

Peter Sommerauer, Oliver Müller

PII: S0360-1315(14)00164-X
DOI: 10.1016/j.compedu.2014.07.013
Reference: CAE 2676

To appear in: *Computers & Education*

Received Date: 24 January 2014
Revised Date: 25 July 2014
Accepted Date: 27 July 2014

Please cite this article as: SommerauerP. & MüllerO., Augmented reality in informal learning environments: A field experiment in a mathematics exhibition, *Computers & Education* (2014), doi: 10.1016/j.compedu.2014.07.013.

Title: Augmented Reality in Informal Learning Environments: A Field Experiment in a Mathematics Exhibition

Author names and affiliations:

Sommerauer, Peter, University of Liechtenstein, Fürst-Franz-Josef-Strasse, 9490, Vaduz, Liechtenstein, peter.sommerauer@uni.li

Müller, Oliver, University of Liechtenstein, Fürst-Franz-Josef-Strasse, 9490, Vaduz, Liechtenstein, oliver.mueller@uni.li

Corresponding author: Müller, Oliver

Augmented Reality in Informal Learning Environments: A Field Experiment in a Mathematics Exhibition

Abstract

Recent advances in mobile technologies (esp., smartphones and tablets with built-in cameras, GPS and Internet access) made augmented reality (AR) applications available for the broad public. While many researchers have examined the affordances and constraints of AR for teaching and learning, quantitative evidence for its effectiveness is still scarce. To contribute to filling this research gap, we designed and conducted a pretest-posttest crossover field experiment with 101 participants at a mathematics exhibition to measure the effect of AR on acquiring and retaining mathematical knowledge in an informal learning environment. We hypothesized that visitors acquire more knowledge from augmented exhibits than from exhibits without AR. The theoretical rationale for our hypothesis is that AR allows for the efficient and effective implementation of a subset of the design principles defined in the cognitive theory of multimedia. The empirical results we obtained show that museum visitors performed better on knowledge acquisition and retention tests related to augmented exhibits than to non-augmented exhibits and that they perceived AR as a valuable and desirable add-on for museum exhibitions.

Keywords: Augmented Reality, Informal Learning, Mathematics, Field Experiment, Museum, Cognitive Theory of Multimedia Learning

1. Introduction

Augmented reality (AR) refers to technologies that dynamically blend real world environments and context-based digital information. More formally, AR has been defined as a system that fulfills three characteristics (Azuma, 1997): First, it combines the real and virtual world. Second, it allows real-time interaction. Third, it aligns real objects or places and digital information in 3D. In some professional contexts (e.g., military), AR technologies have been around for more than 50 years, but only the recent proliferation and consumerization of mobile technologies (e.g., smartphones, tablets) made affordable AR systems available for the broad public. Today's mobile AR applications leverage the built-in 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 asserted that education is one of the most promising application areas for AR (Wu, Lee, Chang, & Liang, 2013). The NMC Horizon Report 2012 identified AR as an emerging technology with high relevance for teaching, learning, and creative inquiry and predicted broad adoption by 2015 (NMC, 2012). Yet, in a recent literature review on AR teaching and learning Dunleavy and Dede (2014) 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). In fact, the majority of existing empirical research is of a qualitative nature (e.g., observations, interviews, focus groups) and concentrates on the elicitation of affordances and constraints of AR in education. Up to now, only few quantitative studies (e.g., experiments) exist that try to measure the effect of AR on learning outcomes.

In order to contribute to filing this research gap, we conducted a large-scale field experiment to test the effect of AR on learning performance. Due to its context-awareness and interactivity, many researchers see the biggest potentials in leveraging AR in informal learning environments (Dede, 2009; Greenfield, 2009), that is, voluntary and self-directed learning that takes place outside of the classroom (OECD, n.d.). We concur with this view and, therefore, conducted a field experiment at a mathematics exhibition, a typical example of an informal learning environment (Screven, 1993).

Our experiment was driven by the hypothesis that visitors learn better from augmented museum exhibits than from exhibits that are accompanied by traditional physical information displays only (e.g., boards, posters, leaflets, quizzes, books, screens). The theoretical foundation for this hypothesis is based upon the cognitive theory of multimedia learning (CTML). We argue that AR inherently implements a subset of the design principles formulated in the CTML, namely, the multimedia principle, the spatial contiguity principle, the temporal contiguity principle, the modality principle, and the signaling principle. The empirical results we obtained provide strong evidence for our hypothesis. Museum visitors learned significantly more from augmented exhibits than from non-augmented exhibits, perceived AR as a valuable add-on of the exhibition, and wish to see more AR technologies in museums in the future.

The remainder of this paper is structured as follows. We first present theoretical background on AR in education and related experimental studies that tried to quantify the effect of AR on learning outcomes. We then describe our experimental design in detail before we come to the statistical analysis of the results. In the discussion section we compare and contrast our findings with other studies and point out directions for future research. We conclude with a brief summary and outlook.

2. Theoretical Background

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 (Mayer, 1997, 2009). CTML is based on three assumptions. First, humans possess two channels for processing information, an auditory/verbal channel and a visual/pictorial channel (Paivio, 1990). Second, each channel can process only a limited amount of information at one time (Sweller, Ayres, & Kalyuga, 2011). 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 (Wittrock, 1992). Based upon these theoretical assumptions, CTML postulates principles for the design of effective multimedia instructions (Mayer, 2009). 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.

3. Related Work

Empirical studies have examined the use AR-based technologies for teaching and learning in natural science, medicine, engineering, languages, history, arts, and other subjects and in various learning environments, for example, kindergartens, schools, universities, laboratories, museums, parks, and zoos (Dunleavy & Dede, 2014; Wu et al., 2013). Given that mobile AR is still an emergent technology and field of study, it is not surprising that the majority of these studies is of a qualitative nature (using methods such as observations or interviews) and concentrates on the elicitation of affordances and constraints of AR for teaching and learning. Up to now, only few quantitative studies exist that try to rigorously measure the effect of AR on learning performance. In the following, we will briefly review extant experimental studies of AR for teaching and learning. As our field experiment focused on teaching general mathematical knowledge, we focused our review on studies that looked at teaching classical K-20 learning contents and excluded studies that looked at specialized professional trainings (e.g., maintenance, repair, medical training). We also excluded studies that lacked the rigorousness of true experimental designs (e.g., control groups, sufficient sample sizes, statistical hypothesis testing). Table 1 shows an overview of the studies we were able to identify.

Table 1: Overview of Experimental Studies on AR for Teaching and Learning

Study	Domain	Setting	Participants	AR Treatment	Control Group Treatment	Dependent Variables	Positive effect of AR
Dünser, Steinbügl, Kaufmann and Glück (2006)	Engineering	Classroom	215 high school students	AR via head-mounted displays	PC with CAD software	Spatial abilities	No
Liu, Tan and Chu (2009)	Ecology	Field trip	72 elementary school students	AR application on a PDA	Paper-based materials	Knowledge acquisition	Yes
Martín-Gutiérrez et al. (2010)	Engineering	Classroom	49 university students	AR book	Paper-based materials	Spatial abilities	Yes
Echeverría et al. (2012)	Physics	Classroom	45 secondary school students	AR game	Multiplayer computer game	Knowledge acquisition	No
Fonseca, Martí, Redondo, Navarro and Sánchez (2014)	Engineering	Classroom	57 university students	AR smartphone app	Paper-based materials	Academic performance (practical skills and spatial abilities)	Yes
Ibáñez, Di Serio, Villarán and Delgado Kloos (2014)	Physics	Classroom	64 high school students	AR smartphone app	Web-based learning application	Knowledge acquisition; Flow experience	Yes
Chang et al. (2014)	Arts	Museum	135 college students	AR smartphone app	Audio guide; No guide	Painting appreciation; Engagement with paintings; Flow experience	Yes

About half of the studies we found examined the effect of AR on learning spatial abilities; a finding that is not surprising as 3D is one of the key affordances of AR. In one of the first large-scale experiments Dünser et al. (2006) investigated the efficacy of AR for training spatial abilities using 215 high school students as participants. Applying a pretest-posttest control group design, the researchers compared an AR-based training application running on a head-mounted display with a CAD application running on a traditional computer with screen, keyboard, and mouse. A between groups comparison could not find clear evidence for the advantageousness of AR as a spatial ability learning tool. Martin-Gutierrez et al. (2010) also studied the effect of AR on learning spatial abilities using a textbook enhanced by a desktop AR system and found more promising results. In a pretest-posttest classroom experiment with 49 university students the AR group showed a significant gain in spatial abilities, whereas the control group using a traditional textbook did not show significant improvements. Finally, in a quasi-experimental study, Fonseca et al. (2014) used a mobile AR application as an educational tool in an architecture and building engineering course with 57 university students. Comparing students’ final grades related to practical skills and spatial abilities with the grades of students of the same course in the previous year (control group without AR), they found a significant statistical difference indicating that the application of AR technology in the course helped to improve students’ performance.

A second group of studies investigated the effect of AR on the acquisition of theoretical natural science knowledge. For example, Liu et al. (2009) conducted an experiment to measure the effect of a mobile AR application on the acquisition of ecological knowledge during a field trip to a nature park with 72 elementary school students. The researchers used a pretest-posttest design with a control group and found that the AR group significantly outperformed the control group in terms of learning improvement. Echeverria et al. (2012) compared an AR game running on tablet computers with touch screens and additional head-up displays with a multi-mice computer game running on standard PCs. In a pretest-posttest design they measured the acquisition of physics knowledge for both groups. The evaluation showed that both technologies had a significant effect on learning performance, but there was no statistical significant difference between groups. Finally, Ibanez (2014) conducted a classroom experiment with 64 high school students to test whether a mobile AR application for smartphones is more effective in supporting the acquisition of physics knowledge than a similar web-based application. The experiment indicated that students in the AR group perceived higher levels of flow experience during the lecture and also gained significantly more knowledge.

Finally, we found one experimental study that examined the use of AR in the context of arts education. Chang et al. (2014) designed a AR museum guide and tested its effectiveness against an audio guide and no guide at all. 135 college students participated in the experiment and the AR group showed significantly greater scores in a painting appreciation test than the two control groups. The researchers also investigated flow levels and amount of time spent focusing on paintings, but did not find clear differences between groups.

In sum, we can conclude that there is first promising quantitative evidence that AR has the potential to improve students' learning performance. Yet, the experimental results are not completely concordant. Two out of the seven reviewed studies did not find a significant difference between the AR group and the control group. Interestingly, both studies compared AR to other computer-based learning technologies, and not to paper-based learning materials. When looking at teaching and learning mathematics-related contents, which is in the focus of this paper, the picture is even more inconclusive. Three studies found positive evidence for the effectiveness of AR, while two studies did not. Finally, our brief review shows that the majority of studies (five out of seven) investigated the effect of AR on structured, organized, and intentional learning in the classroom (formal learning); only two studies were situated in informal learning environments.

4. Materials and Methods

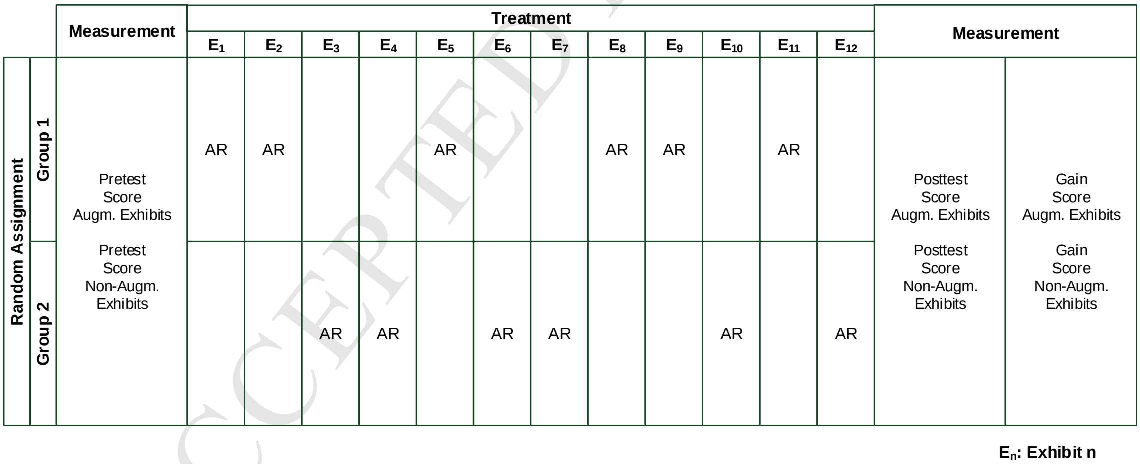
4.1. Experimental Design

The objective of our study was to investigate whether AR is an effective educational technology in informal learning environments. Consequently, the hypothesis underlying our study, which was conducted in the form of a field experiment during a mathematics exhibition at the ANONYMOUS national museum in spring 2013, was that museum visitors learn better from augmented exhibits than from non-augmented exhibits.

We chose to conduct a framed field experiment (Harrison & List, 2004), 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 (Johnson, 2010; Mills et al., 2009), 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 1 graphically summarizes the design of the experiment. Participants were randomly assigned to one of two groups. Participants in both groups were given 90 minutes to visit the mathematics exhibition individually and at their own pace. Before entering the exhibition, participants received a short hands-on training how to use the mobile AR app to discover and activate hidden virtual contents within the exhibition. In addition, all participants had 15 minutes to take a pretest with 16 questions regarding the mathematical exhibits they will later see. The same test, plus additional questions on demographics and user experience, was administered to all participants as a posttest after visiting the exhibition (participants were not told that the same questionnaire is used for the posttest).

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.



		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}**		
Random Assignment	**Group 1**	Pretest Score Augm. Exhibits	AR	AR			AR			AR	AR		AR		Posttest Score Augm. Exhibits	Gain Score Augm. Exhibits
	Group 2	Pretest Score Non-Augm. Exhibits			AR	AR		AR	AR			AR		AR	Posttest Score Non-Augm. Exhibits	Gain Score Non-Augm. Exhibits

E_n: Exhibit n

Figure 1: Overview of the randomized crossover field experiment

4.2. Participants

We recruited 101 participants to take part in the field experiment. The sample included heterogeneous genders, age groups, and educational levels (Table 2). Participants were recruited via mailing lists and local media as well as at the entrance of the museum itself. Participants received free entry into the exhibition as a compensation for taking part in the experiment.

Table 2: Participants of the field experiment

Gender		**Age**				**Education (highest degree achieved)**		
Male	**Female**	**14-20**	**21-40**	**41-60**	**61-79**	**Primary school**	**Secondary school**	**University**
62 (61%)	39 (39%)	35 (34%)	27 (27%)	26 (26%)	13 (13%)	40 (40%)	34 (33%)	27 (27%)

4.3. Treatments

We used Aurasma Studio (Version 2.0) to design augmentations for twelve selected exhibits (Table 3). 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 iPads (4th generation) to discover und unlock augmentations by pointing the tablet's camera at exhibits and trigger images. All tablets were equipped with headphones to allow listening to sound without disturbing other visitors. Manipulation of treatments was done 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.

Table 3: 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 (Figure 2)
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 and 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 (Figure 2)
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

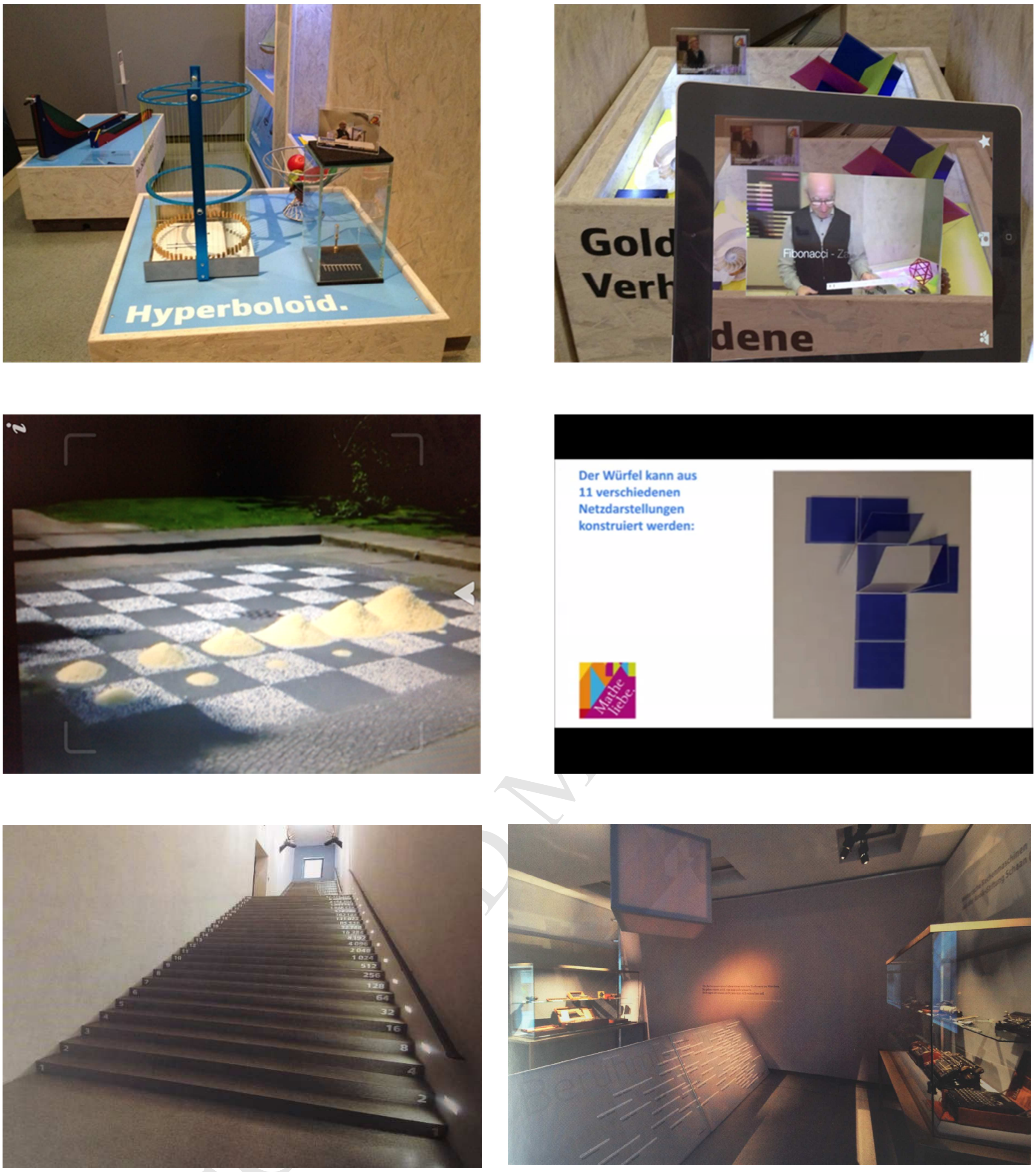


Figure 2: Interactive exhibits of hyperboloids (top left); AR experiences (top right: video in which the curator demonstrates an exhibit, middle: two animations illustrating mathematical problems); Illustration of exponential growth on the stairs of the entrance hall (bottom left); Historical calculators in glass cabinets (bottom right)

As outlined in the Theoretical 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. In the following, we explain how we incorporated these design principles into the experimental AR materials. 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 example, 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 (Figure 2). The *spatial contiguity principle* was implemented by superimposing virtual 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, Figure 2) 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 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.

A key challenge when designing AR materials for experimental treatments is the issue of informational equivalence. According to Larkin and Simon (1987), 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 (2005), 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.[1] Figure 3 illustrates this approach graphically.

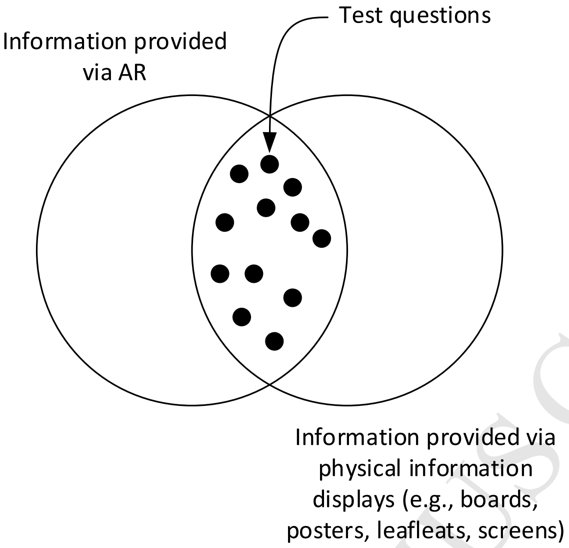


Figure 3: Alignment of information provided via AR, information provided via physical information displays and test questions

4.4. 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. This decision was driven by the guidelines outlined in Parsons and Cole (2005), who advocate the use of simple comprehension tests to compare different representations of information, as such tests focus on a representation's ability to effectively and efficiently convey information. Knowledge application or problem solving tests, in contrast, are intended to measure a deeper level of domain understanding in which information provided

---

[1] We gratefully thank one of the anonymous reviewers for providing us with the notion of "educational equivalence".

by a representation needs to be integrated with existing knowledge schema (e.g., a person's general mathematical understanding or mental arithmetic skills).

All pretest and posttest 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 all questions were reviewed by the curator of the exhibition, who was a retired mathematics high school teacher.

We aggregated the answers to the individual questions to six test scores (Figure 1). The pretest score for augmented objects and the pretest score for non-augmented objects captured the level of previous knowledge regarding the mathematical exhibits. The posttest score for augmented objects and the posttest score for non-augmented objects captured the knowledge level after visiting the exhibition. The possible values of pretest and posttest scores ranged between 0 and 6. Knowledge acquisition and retention was measured by computing gain scores as the difference between a participant's posttest and pretest scores. Analog to the pretest and posttest scores, we computed gain scores for augmented and non-augmented objects separately. Possible values of gain scores ranged between -6 and 6.

In addition to the above test questions we included four control questions into the pretest and posttest questionnaires to check for potential confounding factors. We added three control questions related to exhibits that were not augmented at all, neither for Group 1 nor for Group

2, and that were not tagged in any way. The answers to these questions were used to check whether visitors were biased towards tagged exhibits, even if they were not able to access the corresponding augmentation (as it was only accessible for participants in the other group). We also added one control question related to an additional exhibit which's augmentation was accessible for both groups. This question was used to check for unintended group differences (e.g., due to inappropriate randomization). The posttest questionnaire also contained a number of simple user experience questions and standard demographics questions.

5. Results

5.1. Descriptive Statistics

Table 4 and Figure 4 give an overview of the test scores. All results are in line with expectations. The low scores on the pretest 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 4: Descriptive statistics of test scores

	Pretest Scores					**Posttest Scores**					**Gain Scores**				
	M	**Mdn**	**SD**	**Min**	**Max**	**M**	**Mdn**	**SD**	**Min**	**Max**	**M**	**Mdn**	**SD**	**Min**	**Max**
Augmented Exhibits	1.75	2	1.11	0	5	3.64	4	1.31	0	6	1.89	2	1.50	-2	6
Non-Augmented Exhibits	1.81	2	1.16	0	5	2.59	3	1.28	0	6	0.78	1	1.46	-2	4

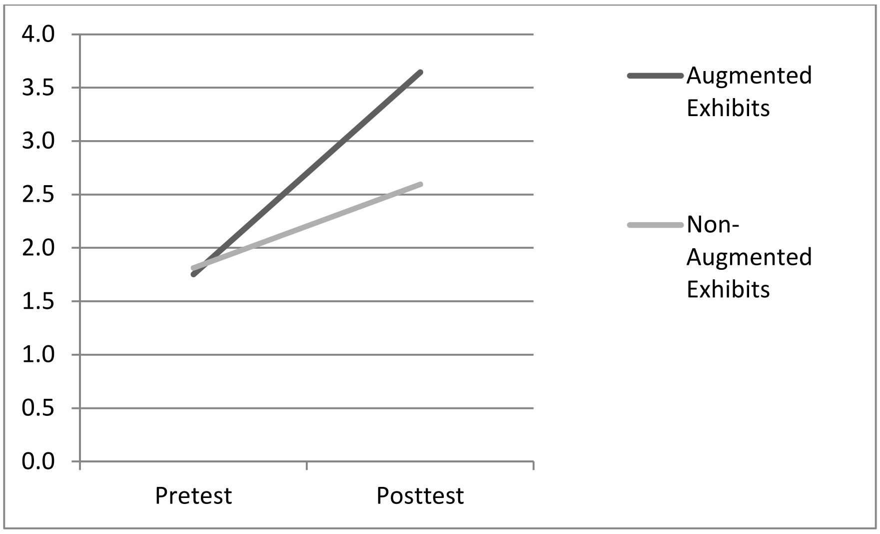


Figure 4: Comparison of pretest and posttest scores related to augmented and non-augmented exhibits

5.2. Hypothesis Testing

Usually, the statistical analysis of paired pretest-posttest data is done via paired t-tests or a repeated measures analysis of variance (Dimitrov & Rumrill, 2003). 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. Specifically, we conducted Wilcoxon signed-rank tests on pretest scores, posttest scores, and gain scores for augmented and non-augmented exhibits (Table 5) and for the additional control questions.

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 first performed a Wilcoxon signed-rank test on the pretest scores. The test showed no statistically significant differences in median scores between the two pretest question sets, z = -0.409, p = 0.682. From 101 participants, 37 participants performed better on questions related to augmented exhibits, 35 participants performed better on questions related to non-augmented exhibits, and 29 participants showed no difference in performance between questions related to augmented and non-augmented exhibits.

Next, we compared medians of posttest scores. Participants performed significantly better on posttest questions related to augmented exhibits (Mdn = 4) than on posttest questions related to non-augmented exhibits (Mdn = 3), $z = -5.069$, $p < 0.005$. From the 101 participants, 66 were better on questions related to augmented exhibits, whereas 17 were better on questions related to non-augmented exhibits; 18 participants showed no difference in performance.

To examine the magnitude of learning improvements, we proceeded with an analysis of gain scores. Participants learned significantly more from augmented exhibits (Mdn = 2) than from non-augmented exhibits (Mdn = 1), $z = -4.679$, $p < 0.005$. 62 participants gained more on questions related to augmented exhibits, 20 participant gained more on questions related to non-augmented exhibits, and 19 participants showed no difference.

We also computed the effect sizes for the Wilcoxon’s signed-rank tests using the formula given in Rosenthal (1991, p. 19). The effect size for the difference in posttest scores was $r = 0.36$ and the effect size for the difference in gain scores was $r = 0.33$, which can be considered medium effects (Cohen, 1992).

Table 5: Results of Wilcoxon signed-rank tests[2]

		N	**Mean Rank**	**Sum of Ranks**
Pretest Score Non-Augmented Exhibits – Pretest Score Augmented Exhibits	**Positive Ranks**	37[a]	33.58	1242.50
	Negative Ranks	35[b]	39.59	1385.50
	Ties	29[c]	-	-
a. PretestScoreNonAugmentedObjects < PretestScoreAugmentedObjects *b. PretestScoreNonAugmentedObjects > PretestScoreAugmentedObjects* *c. PretestScoreNonAugmentedObjects = PretestScoreAugmentedObjects*				
Posttest Score Non-Augmented Exhibits – Posttest Score Augmented Exhibits	**Positive Ranks**	66[a]	43.11	2845.50
	Negative Ranks	17[b]	37.68	640.50
	Ties	18[b]	-	-
a. PosttestScoreNonAugmentedObjects < PosttestScoreAugmentedObjects *b. PosttestScoreNonAugmentedObjects > PosttestScoreAugmentedObjects* *c. PosttestScoreNonAugmentedObjects = PosttestScoreAugmentedObjects*				
Gain Score Non-Augmented Exhibits – Gain Score Augmented Exhibits	**Positive Ranks**	62[a]	43.63	2705.00
	Negative Ranks	20[b]	34.90	698.00
	Ties	19[c]	-	-
a. GainScoreNonAugmentedObjects < GainScoreAugmentedObjects *b. GainScoreNonAugmentedObjects > GainScoreAugmentedObjects* *c. GainScoreNonAugmentedObjects = GainScoreAugmentedObjects*				

Finally, we analyzed the control questions to rule out further potential confounding factors. A Wilcoxon signed-rank test[3] showed no significant differences in median gain scores per question for control questions and for questions related to exhibits with inaccessible augmentations. We interpreted this as an indicator that visitors were not biased toward exhibits with inaccessible augmentations, as compared to totally “naked” exhibits, and vice versa. Regarding the control question related to the one exhibit which’s augmentation was accessible for

---

[2] A visual inspection of the shapes of the distributions of difference scores showed that the scores were approximately symmetrical and, hence, that all required assumptions of the Wilcoxon signed-rank test were met.

[3] A visual inspection of the shapes of the distributions of difference scores showed that the scores were approximately symmetrical and, hence, that all required assumptions of the Wilcoxon signed-rank test were met.

both groups, a Mann-Whitney U[4] test found no significant between-subjects difference in median gain scores. This gives indication that there were no differences in the use of AR between the two groups.

5.3. Post-hoc Analysis

In addition to the hypothesis tests, we carried out tests to check whether there were any differences in the effect of AR on learning performance between subgroups of our sample. A Mann-Whitney U test[5] with gender as a grouping variable showed neither for augmented nor for non-augmented exhibits a statistically significant difference in median gain scores between males and females. We performed two Kruskal-Wallis tests[6] to inspect whether the effect of AR on learning performance was different across educational and age groups. For the category education, the tests showed no significant differences. However, the scores were significantly different between the different age groups for augmented exhibits, $X^2(3) = 10.973$, $p = 0.012$. There were significant differences in gain scores for augmented exhibits between the age group 41-60 (Mdn = 3) and the age group 14-20 (Mdn = 2) ($p = .028$) and the age group 41-60 and the age group 61-79 (Mdn = 1) ($p = .035$), but not between any other combinations. For non-augmented exhibits, no statistically significant differences in gain scores across age groups were found.

---

[4] A visual inspection of the shape of the distribution of gain scores in each group showed that they were reasonably similar and, hence, that all required assumptions of the Mann-Whitney U test were met.

[5] A visual inspection of the shape of the distribution of gain scores in each group showed that they were reasonably similar and, hence, that all required assumptions of the Mann-Whitney U test were met.

[6] A visual inspection of the shape of the distribution of gain scores in each group showed that they were reasonably similar and, hence, that all required assumptions of the Kruskal-Wallis test were met.

5.4. Visitor Feedback

Besides measuring learning performance, we also asked participants whether they perceived the augmented exhibits as a positive experience. An overwhelming majority of participants reported that the mobile AR app was a valuable add-on for the exhibition, that the AR experience did not overload them, and that they wish to see more AR in museums in the future (Table 6). These results indicate that AR is not only an effective tool for learning in museums, but also a technology that museum visitors perceive as valuable and desirable.

Table 6: Visitor feedback on the AR experience

Do you think that AR is a valuable add-on for museum exhibitions?			
Yes, absolutely	***Yes, partly***	***Not really***	***Not at all***
72 (71.3%)	26 (25.7%)	2 (2.0%)	1 (1.0%)
Do you think that the enhancement of exhibitions through AR is "too much"?			
Yes, absolutely	***Yes, partly***	***Not really***	***Not at all***
4 (4.0%)	13 (12.9%)	32 (31.7%)	50 (49.5%)
Do you wish to see more AR in museums in the future?			
Yes, absolutely	***Yes, partly***	***Not really***	***Not at all***
58 (57.4%)	34 (33.7%)	6 (5.9%)	1 (1.0%)

6. Discussion

Our field experiment was driven by the hypothesis that museum visitors learn more from augmented exhibits than from non-augmented exhibits. We grounded this hypothesis in the cognitive theory of multimedia learning. The conducted field experiment produced empirical evidence that provides strong support for our hypothesis. Visitors performed significantly better on posttest questions related to augmented exhibits than on posttest questions related to non-augmented exhibits. Also, they showed significantly greater gain scores when comparing posttest and pretest question scores. The analysis of the effect size for both tests indicated that AR has a medium effect on learning performance.

This study contributes to the still emerging body of quantitative empirical evidence on the effect of AR on learning performance, especially learning mathematics-related contents in informal environments. Experimental results on the application of AR in this field are still inconclusive. For example, in contrast to the findings of Dünser et al. (2006) and Echeverria et al. (2012), who could not find a significant advantage of AR learning materials over other materials, we were able to obtain positive evidences for the efficacy of AR. However, it has to be noted that both studies compared AR to other computer-based treatments, and not to physical learning materials. Interestingly, Dünser et al. (2006) and Echeverria et al. (2012) discovered significant gender differences; in both studies male subjects profited from AR as compared to non-AR technologies and outperformed females using AR. We could not replicate these gender differences in our study. Our results are consistent with the results of other studies (Fonseca et al., 2014; Ibáñez et al., 2014; Martín-Gutiérrez et al., 2010), which found that AR can have a significant positive effect on knowledge acquisition performance. In particular, we could replicate and transfer the findings of a recent study of Ibáñez et al. (2014), who found that students using AR performed better on retention tests of physics knowledge than students using a web-base learning tool. The authors explained their results by arguing that AR technologies, as compared to traditional computer technologies, require a lower cogntive effort from users. This rationale is in line with our theoretical argument that AR allows for the efficient and effective implementation of CTML design principles, which, in turn, are partly based on cognitive load theory. When looking at the use of AR in informal learning environemts, our study extends the findings of Liu et al. (2009) and Chang et al. (2014). Both studies found empirical evidences for the efficacy of AR in field settings, but in non-matehmtical contexts. We demonstrated the value of AR for teaching formal contents (mathematics) in informal environments (museums). All extant AR studies in the mathematics context have been conducted in formal classroom situations. Our study, in contrast, investigated natural subjects (i.e., visitors) conducting natural tasks (i.e., engaging with exhibits) in a natural place (i.e., museum). Learning was not an organized and intentional process, but voluntary and self-directed. Taken together, the findings of our study and the above discussed studies suggest that AR has the potential to be an effective learning tool for mathematics-related and other contents in formal and informal learning environments.

The realistic field setting of our experiment added to its external validity. Yet, field experiments come with a number of threats to internal validity. For example, we were not able to control the actions of the experimental subjects during their 90 minutes museum visits. Hence, we cannot rule out that visitors paid more attention to augmented exhibits or to exhibits that were covered in the pretest. Especially the first case is a potential confounding factor that may have influenced our results. Yet, in self-directed learning settings, like the one used in this study, increasing voluntarily time spent on a task could also be understood as a positive side effect of a technology, and not as a threat. A second potential confounding factor stems from the fact that we were not able to ensure full informational equivalence of AR and non-AR materials, as the AR experiences we have designed were not artificial, but used in the museum on a daily basis.

Our study points out a number of possible directions for future research. First, although we have provided theoretical arguments for the proposition that the implementation of the principles of CTML makes AR an effective educational technology, our experimental design was not set up to “prove” that this theory really explains the causes for the observed effects. To do this, future studies should compare the effect of AR experiences that are designed in accordance to the principles of CTML with AR experiences that intentionally violate these principles.

Second, the post-hoc analysis of our experimental results showed that the effect of AR on learning performance differed significantly between age groups. In our experiment, the age group 41-60 profited the most from the use of AR. This is somewhat surprising, as one would usually expect that AR is especially effective with younger people. At the moment, we can only speculate about potential explanations. Our assumption, that builds upon the observations we made and the feedback we got during and after the experiments, is that this age group perceived the AR technology as something new and exciting and, at the same time, was not alienated by it. Yet, further research is needed to replicate, if possible, this result and find theoretically and empirically grounded explanations.

Third, we used a mobile AR app in combination with tablet computers and headphones for the experiment. This technology is omnipresent today, however, not without drawbacks. Some users complained that the tablets are heavy to carry around and hold when pointing at exhibits. As a result, users sometimes started shaking which, in turn, caused the camera to lose the focus and the app to stop the AR experience. Future research should investigate the consequences of such usability issues on the effect of AR and test different AR hardware (e.g., lightweight head-mounted displays).

Finally, our study is not without limitations. In particular, we solely focused on short-term knowledge acquisition and retention. First, it would be interesting to examine whether AR also has a positive effect on long-term knowledge retention. Second, we suggest that future studies should try to replicate our results for higher-order learning tasks, especially knowledge application (problem solving). The studies conducted by Martín-Gutiérrez et al.

(2010) regarding the effect of AR on spatial abilities and Fonseca et al. (2014) regarding the effect of AR on general academic performance have already provided first promising results in this respect.

7. Conclusion

Recent advances in mobile technologies – mobile cameras, GPS and Internet access – made AR available for everybody owning a smartphone. Consequently, many educators and developers started exploring the potential of AR for teaching and learning in various subjects and contexts. Yet, so far only few studies exist that tried to quantify the effect of AR on learning outcomes. To the best of our knowledge, the here presented study is the first *field* experiment on the effect of AR in learning mathematical contents. The empirical evidence we gathered provides strong support for the proposition that AR has the potential to be an effective tool for learning formal contents (mathematics) in informal learning environments (museums). Museum visitors learned significantly more from augmented exhibits than from non-augmented exhibits, perceived AR as a valuable add-on of the exhibition, and wish to see more AR experiences in museums in the future. Due to this combination of measurable utility and perceived user acceptance we think that AR bears the potential to replace traditional audio guides in museums in the near future; especially when considering the advent of next generation AR devices such as Google Glass.

References

Azuma, R. (1997). A survey of augmented reality. *Presence: Teleoperators and Virtual Environments*, *6*(4), 355–385.

Chang, K.-E., Chang, C.-T., Hou, H.-T., Sung, Y.-T., Chao, H.-L., & Lee, C.-M. (2014). Development and behavioral pattern analysis of a mobile guide system with augmented reality for painting appreciation instruction in an art museum. *Computers & Education*, *71*, 185–197.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159.

Dede, C. (2009). Immersive interfaces for engagement and learning. *Science*, *323*(5910), 66–69.

Dimitrov, D., & Rumrill, P. J. (2003). Pretest-posttest designs and measurement of change. *Work*, *20*, 159–165.

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

Dünser, A., Steinbügl, K., Kaufmann, H., & Glück, J. (2006). Virtual and augmented reality as spatial ability training tools. In *Proceedings of the ACM SIGCHI New Zealand Chapter's International Conference on Computer-Human Interaction* (pp. 125–132).

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

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*, 434–445.

Greenfield, P. M. (2009). Technology and informal education: What is taught, what is learned. *Science*, *323*(5910), 69–71.

Harrison, G., & List, J. (2004). Field Experiments. *Journal of Economic Literature*, *42*(4), 1009–1055.

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*, 1–13.

Johnson, D. (2010). Crossover experiments. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(5), 620–625.

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

Liu, T., Tan, T., & Chu, Y. (2009). Outdoor Natural Science Learning with an RFID-Supported Immersive Ubiquitous Learning Environment. *Educational Technology & Society*, *12*(4), 161–175.

Martín-Gutiérrez, J., Luís Saorín, J., Contero, M., Alcañiz, M., Pérez-López, D. C., & Ortega, M. (2010). Design and validation of an augmented book for spatial abilities development in engineering students. *Computers & Graphics*, *34*(1), 77–91.

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

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

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

NMC. (2012). *NMC Horizon Report: 2012 Higher Education Edition*. Retrieved from http://www.nmc.org/publications/horizon-report-2012-higher-ed-edition

OECD. (n.d.). Recognition of Non-formal and Informal Learning - Home. Retrieved January 12, 2014, from http://www.oecd.org/education/skills-beyond-school/recognitionofnon-formalandinformallearning-home.htm

Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford: Oxford University Press.

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.

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Sage Publications.

Screven, C. (1993). Museums and Informal Learning. *CMS Bulletin*, *1*(1).

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York: Springer.

Wittrock, M. (1992). Generative learning processes of the brain. *Educational Psychologist*, *27*(4), 531–541.

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*, 41–49.

Highlights:

- We conducted a cross-over framed field experiment to measure the effect of augmented reality (AR) on learning outcomes
- The field experiment was situated in a mathematics exhibition
- Participants learned significantly better from augmented exhibits than from non-augmented exhibits
- Participants perceived AR as a valuable add-on to the exhibition and wish to see more AR applications in museums in the future