The Effect of Marker-less Augmented Reality on Task and Learning Performance

Peter Sommerauer, Oliver Müller, Leonard Maxim, and Niels Østman

IT University of Copenhagen, Department Business IT, Copenhagen, Denmark
`{psom,olmy,leom,nies}@itu.dk`

Abstract. Augmented Reality (AR) technologies have evolved rapidly over the last years, particularly with regard to user interfaces, input devices, and cameras used in mobile devices for object and gesture recognition. While early AR systems relied on pre-defined trigger images or QR code markers, modern AR applications leverage machine learning techniques to identify objects in their physical environments. So far, only few empirical studies have investigated AR's potential for supporting learning and task assistance using such marker-less AR. In order to address this research gap, we implemented an AR application (app) with the aim to analyze the effectiveness of marker-less AR applied in a mundane setting which can be used for on-the-job training and more formal educational settings. The results of our laboratory experiment show that while participants working with AR needed significantly more time to fulfill the given task, the participants who were supported by AR learned significantly more.

Keywords: Augmented Reality, Learning, Mobile Application

1 Introduction

Augmented Reality (AR) is known as a technology which augments the real environment with relevant digital information [3]. Such information can be superimposed on recognized objects using smartphones, tablets or AR goggles as user interfaces between the real and the virtual world. Additionally, AR allows a full 3D view of virtual objects and enables users to interact with them.

AR's potential has been shown in many use cases and in various settings, such as informal and formal learning environments, workplaces, museums and natural environments [2, 4, 5, 6, 10, 12, 21, 22, 24, 25]. In most settings which have been studied so far, trigger images or QR codes have been used for identifying objects in order to superimpose digital information on them [4]. Only few applications exist that use so-called marker-less AR [4]. Marker-less AR works in a way that the real environment itself and real objects therein are recognized by the app, which then augments digital information and adds functionality to the digitally enriched objects and environments [25], without any pre-defined trigger images or QR codes.

In this study we investigate the application and effectiveness of marker-less AR to support both the execution of a specific task in a mundane setting and the learning about

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

the underlying domain by executing the task (i.e., learning-by-doing). In particular, we intend to answer the following research questions:

- RQ1: How can marker-less AR be implemented in a real-world environment?
- RQ2: How does marker-less AR affect task and learning performance?

In the pursuit of answering our research questions, we developed a marker-less AR app, which enables the user to learn the names of objects from the real environment. We created a fictional learning situation with a given task and compared the results from two groups, one using an AR-based tool, the other using a traditional paper-based tool (Note that a direct comparison between marker-less and marker-based AR is not the aim of this study). Hence, our laboratory experiment uses a static group design with an experimental group and a control group. With this design we intent to investigate the differences in task and learning performance of the two groups by measuring task performance (i.e., time required for completing the task) and learning performance (i.e., answering a post-test questionnaire with questions about the task).

The remainder of this paper is structured as follows: To prepare the background, we first present related work and provide theoretical background on marker-less AR and its implementation. As our study was motivated by investigating task performance and learning performance, we also present associated performance metrics that are derived from learning theories. We then outline the app development process along with the embedding of a number of theory-ingrained design principles, followed by an introduction of the used dataset for image recognition and the setup and execution of our experiment. Next, we provide detailed insights into our data analysis, which prepares for the discussion of our results. Finally, we conclude with a brief summary and directions for future research.

2 Background

Our research background focuses on synthesizing the findings of published systematic literature reviews on AR learning and empirical studies about marker-less AR from the last decade. In order to identify relevant related work, we analyzed the most cited literature reviews on AR for education.

Most extant studies do not focus on using AR in real-life environments, but investigate its use for supporting a narrow and well-defined task in a controlled setting. Hence, it is not surprising that virtually all existing studies focus on the application of marker-based AR, which is easy to implement in a controlled laboratory setting, and that only few studies have investigated the use of marker-less AR so far [4, 5, 21, 26]. Moreover, marker-less AR is one key aspect discussed for implementing hybrid tracking for ubiquitous AR [5, 21, 25, 26].

What most studies have also in common is that they emphasize the need of further research on the features, use, advantages, and limitations of AR in educational settings [2, 4, 6, 10]. Reported advantages of AR in educational settings include learning gains, higher motivation, facilitated interaction, better collaboration, lower cost, better user experiences, just-in-time information, enabling of situated learning and student-centered approaches, increase of students' attention, enjoyment, exploration, increased capacity for innovation, creation of positive attitudes, more awareness, anticipation, and authenticity [2, 3, 4, 5, 6, 10, 11, 12, 21, 22, 24, 25, 26]. In contrast, repeatedly reported limitations of AR in education include the observation that AR apps are mostly designed for only one specific knowledge field [4], that teachers cannot create new learning content [2, 4, 10, 22], that there are difficulties maintaining superimposed information, that learners pay too much attention to the virtual information, that evaluation focused on short-term instead of long-term learning [4], and that AR can be perceived as an intrusive technology [4, 21, 26]. Still, most studies found positive evidence for the effectiveness of AR in education, for example, in the form of enhanced learning performance, higher learning motivation, improved perceived enjoyment, decreased cost, as well as adding creating positive attitudes towards education and fostering students' commitment [2, 3, 4, 5, 6, 10, 11, 12, 21, 22, 24, 25, 26].

In Bacca et al.'s review of AR for education, the authors report about 19 studies that use marker-based AR, 4 studies with marker-less AR, and 7 studies covering location-based AR [4]. They discuss challenges around the improvement of recognition algorithms (e.g., for human forms) in the process of achieving more immersive and not intrusive AR learning experiences. Furthermore, they recommend vocational educational training (VET) classes as target groups for future studies.

In their literature survey of AR, Billinghurst et al. [5] additionally focus on technology for user activity tracking considering input and interaction. They provided first design guidelines and interface patterns for AR development tools, starting with considering physical objects, virtual content and interaction metaphors and their connection. Additionally, they suggest future research directions as user tracking, user interaction, AR displays, and social acceptance of AR.

Dunleavy & Dede provide insights in AR teaching and learning, focusing on AR utilizing mobile, context-aware technologies (e.g. smartphones, tablets), thus enabling AR users interacting with digital information which is embedded within physical environments and in both, formal and informal learning environments [12]. They additionally investigate affordances and limitations for AR related to teaching, learning and instructional design and see AR as primarily aligned with situated and constructivist learning theory, stating, that AR positions learners within a real-world physical and social context while guiding, scaffolding and facilitating participatory and metacognitive learning processes (e.g. authentic inquiry, active observation, peer coaching, reciprocal teaching). Since AR legitimate users in peripheral participation with multiple modes of representation, they distinguish between location-aware and vision-based AR. In this context, AR has some limitations regarding student cognitive overload and managing level of complexity, which is a key instructional issue. Therefore, they recommend to decrease cognitive load by creating a simplified experience structure initially and increasing complexity as the experience progresses, thus scaffolding each experience explicitly at every step to achieve the desired experience or learning.

When Radu states that the educational community remains unclear regarding the educational usefulness of AR and regarding contexts in which this technology is more effective than other educational mediums, he refers to 26 publications comparing student learning with AR vs. non-AR apps [22]. Radu observed some negative consequences, such as attention tunneling, usability difficulties, ineffective classroom integration, and learner differences. His table of factors influencing learning in AR covers content representation, multiple representations that appear at appropriate time and space, learners are physically enacting educational concepts, attention is directed to relevant content, learners are interacting with 3D simulations, interaction and collaboration are natural.

Still, the benefits of AR in educational environments and the value of AR apps applied in educational environments has not yet been investigated in its entirety [8]. The different directions of AR apps differ regarding their potential benefits. In their systematic literature review to synthesize a set of 25 publications, Diegmann et al. [10] identified 14 different benefits clustered in six different groups. They considered dimensions like state of mind (e.g. increased motivation, increased attention, increased concentration, increased satisfaction), teaching concepts (e.g. student-centered learning, collaborative learning), presentation (e.g. increased details, information accessibility, interactivity), learning type (e.g. improved learning curve, increased creativity), content understanding (e.g. improved development of spatial abilities, memory), and reduction of costs [10]. They then mapped the benefits to five directions of AR in educational environments (discovery-based learning, objects modeling, AR books, skills training, AR gaming) and indicated that specific directions of AR apps are more likely to lead to certain benefits, such as increased motivation. Especially, they emphasize that future research is needed to investigate the causality between benefits and directions of AR.

In their review of AR in education from 2011 to 2016, Chen et al. focused on research which includes the uses, advantages, features, and effectiveness of AR in educational settings [6]. They recommended to undertake more studies considering the difference of cognitive process and psychological immersion between AR and reality settings, individual interaction, sense of identity, adaptive application in AR, AR classroom design and evaluation research, teacher's role model in AR educational setting, design and implementation of AR learning resources in K-12.

The literature review by Akçayır & G. Akçayır focuses on current advantages and challenges of? AR education. Although AR promotes enhanced learning achievement, they experienced a discrepancy for AR in terms of cognitive load and/or cognitive overload, and AR ease of use vs. challenges for AR app usability [2]. Since research studies report both, they advise AR developers to develop and consequently implement empirically proven design principles, focusing on AR use and educational outcomes, and AR apps designed for diverse populations (e.g. kids, students, lifelong learners). They emphasize the need to investigate students' satisfaction, motivation, interaction, and commitment, and provide insights from research and development comprising explanations of development processes and factors being considered in design.

Dunleavy, Dede, and Mitchell document in their review covering AR simulations for teaching and learning, how teachers and students describe and comprehend ways of participation in AR simulation, to aid or hinder teaching and learning [11]. By means of qualitative case studies across two middle schools they demonstrate that AR supports multi user environments and immersive collaborative simulation.

For professional education and training, Palmarini et al. focused on the state of the art of AR apps applied in maintenance [21]. Based on 30 primary studies between 1997-2017, they unveil most relevant technical limitations for AR and propose results indicating a high fragmentation among hardware, software and AR solutions which lead to a high complexity for selecting and developing AR systems, thus identifying areas where AR technology still lacks maturity (e.g. marker-less AR).

Further limitations for AR which are still present today were depicted by Zhou et al. for tracking techniques, interaction techniques, user interfaces, and AR displays, especially for head mounted displays (HMD) [26]. Although the development of AR hardware became more sophisticated in the past decade, the major technical issues are not sufficiently dissolved and need to be overcome, like low sensitivity trigger to recognition [2].

3 Methodology

In our study we followed the advice from Dunleavy & Dede (i.e., decreasing cognitive load by creating a simplified experience structure) [12], Diegmann et al. (i.e., causality between benefits of AR) [10], Chen et al. (i.e., AR classroom design and evaluation research, design and implementation of AR learning resources) [6], Palmarini (i.e., use of marker-less AR) [21]. In order to develop an AR app for both school and professional education (VET), we applied design principles from Billinghurst et al. (i.e., real physical objects/virtual elements to be displayed, linking interaction metaphor) [5] and Sommerauer & Müller (i.e., design elements derived from learning theories) [24].

For the evaluation of the effect of marker-less AR applied in a learning scenario we chose to design a controlled laboratory experiment to compare the support of AR with traditional, paper-based material inside a classroom. In this, we aimed to ensure that no or hardly any differences in information equivalence [16] could affect the results of our study. Finally, our research design aimed to support and control exactly those research design elements which were the key subject of investigation.

With the experiment we examined the usability of AR, its effectiveness and the potential for teaching and learning. The evaluation covered measures for perceived usefulness, perceived learning and students' motivation as well as objective performance in terms of time to completion for the task and number of mistakes made in a recall and retention test administered as a post-test. In addition, we employed the Systems Usability Scale (SUS) [23] to evaluate the usability of the applied AR system.

In our app development, we considered design elements from Billinghurst et al., who proposed to focus on physical objects, virtual content, the interaction metaphor, and their connections [5]. Additionally, we applied the conceptual framework by Sommerauer & Müller [24], which is inspired by Anderson's work on how learning can be enhanced using emerging technologies and applying learning theories [1]. At the heart of this framework are one or more learning sequences, each consisting of one or more connected learning activities. At the center of a single learning activity stands the learning content. This content should be designed according to different learning theories, indicated by the different concentric layers surrounding the learning content.

At the first layer, it is proposed to apply the 12 design principles of the cognitive theory of multimedia learning (CTML) [18]. In the second layer, design elements from mobile learning (e.g., Herrington et al. [13]) shall be considered for application design. Finally, it is proposed to implement design elements from game-based learning (e.g. leaderboard, mission) [14], simulations (e.g., storytelling, drama), experiential learning theory (e.g., diverging, assimilating) [15], and situated learning [19]. Additionally, collaborative learning elements can be introduced at the learning stage, where multiple learning activities are combined into a learning sequence [24].

We instantiated the above described conceptual framework by developing an AR learning app prototype. It supports the task of learning names related to physical objects used in a particular professional domain – in our case, the florist industry. More specifically, the app combines machine learning techniques for image recognition and machine translation to identify objects that are in the focus of the mobile phone camera in real-time and superimpose information such as the object's name in different languages onto the object. As a training application, the app can be used in any workplace environment and the trainee can select between exploration mode or quiz mode. In both, the user needs to focus the particular object using the device's camera (e.g. smartphone, tablet, any head-mounted device). Once the object is recognized, the app provides a selection of labels, comprising the three most likely names of the object using a percentage scale and colors. In quiz-mode, the app shows the most likely label and two randomly selected labels and the trainee has to pick the correct one. Figure 1 shows screenshots of the application and show the explore and quiz modes.

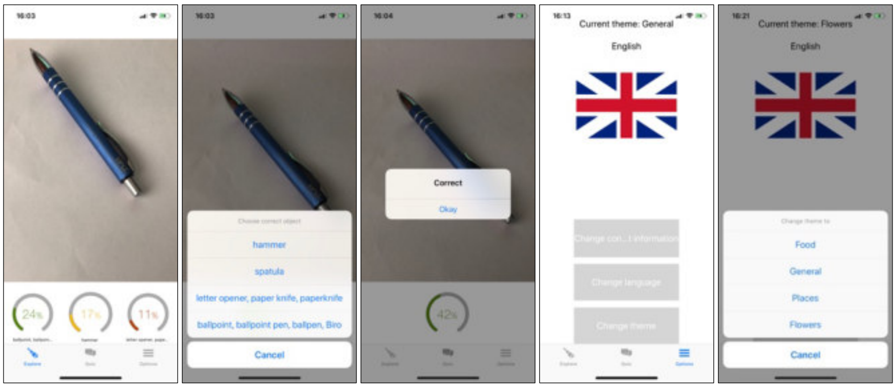


Figure 1. App in explore mode, quiz mode and selection of language and theme

The app design integrates design elements from CTML (i.e., the multimedia principle, the spatial contiguity principle, the temporal contiguity principle, and the signaling principle) with elements from the theory of mobile learning (i.e., users can use the app across space and time) and game-based elements). From a technical perspective, the app is based on Apple's ARKit framework[1] for implementing mobile

---

[1] https://developer.apple.com/arkit/

AR experiences, Google's MobileNets model[2], a convolutional neural network for efficient image recognition on mobile phones, and the Google Translate API[3] for automated translation of texts into multiple languages.

As a foundation for our flower identification app we used the flowers dataset by Nilsback & Zisserman [20], implemented as a selectable theme in our app. The flowers dataset consists of 8,189 images of flowers commonly occurring in the United Kingdom. The images are divided into 103 classes and each class consists of between 40 and 250 images. The images are scaled so that the smallest dimension is 500 pixels. The flowers are identified by different features describing different properties, e.g., color (HSV values of pixels), histogram of gradient orientations (HOG) [8], and distinctive image features (SIFT) [17] on foreground region and foreground boundary. In prior studies the recognition accuracy was measures at 72.8 percent.

The instructional design for the learning situation applied in the experiment contained elements from cognitive and constructivist learning theories. While the learning content was prepared based on CTML principles, elements of constructivist theory were implemented in the learning activity, such as, task orientation, mobile learning, and situated learning, by sending learners on missions including storytelling.

The laboratory experiment was based on a sequential quantitative method research [7] applying a static group design. The aim of the experiment was to identify differences in the application of AR vs. traditional learning. While the experiment group was supplied with mobile devices (iPhone 8+ and X) running the AR app, the control group received a traditional, paper-based tool (catalogue) to fulfil their task. Both groups received the same instructions and were required to fulfil the same task. At the end of the experiment both groups received a post-test questionnaire covering the same topics and questions. The questionnaire contained three sections. The first covered aspects for perceived usefulness, perceived learning, and students' motivation. The second part was a multiple-choice test asking for the names of five flowers shown as pictures. For each, participants could choose between three given names. The number of correctly identified flowers was used as an objective measure for learning performance. The third section of the questionnaire contained ten questions from the System Usability Scale (SUS), which was only available for the group using AR in the experiment.

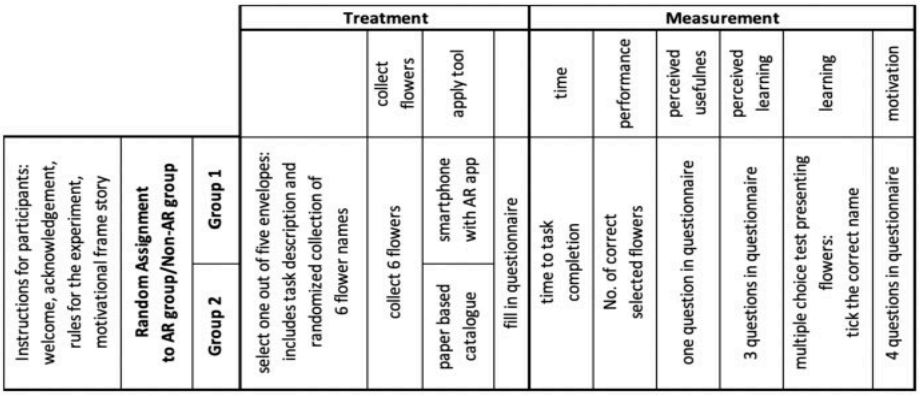
4 Experimental Setup

The laboratory experiment followed a static group design comprising an experimental group and a control group. With this design we intended to investigate the differences in task and learning performance of two groups: one supported by an AR tool and one using traditional tools (i.e. a catalogue). Following similar studies [2, 3, 4, 5, 6, 11, 12, 22, 24] and in line with our research questions, we used the item "time for task completion" as a measure for task performance and "No. of correctly identified flowers" from the questionnaire after the treatment as an indicator for learning performance. Figure 2 gives an overview of the randomized field experiment.

---

[2] https://github.com/tensorflow/models/tree/master/research/slim/nets/mobilenet

[3] https://cloud.google.com/translate/docs/apis



			Treatment				Measurement					
				collect flowers	apply tool		time	performance	perceived usefulnes	perceived learning	learning	motivation
Instructions for participants: welcome, acknowledgement, rules for the experiment, motivational frame story	**Random Assignment to AR group/Non-AR group**	**Group 1**	select one out of five envelopes: includes task description and randomized collection of 6 flower names	collect 6 flowers	smartphone with AR app	fill in questionnaire	time to task completion	No. of correct selected flowers	one question in questionnaire	3 questions in questionnaire	multiple choice test presenting flowers: tick the correct name	4 questions in questionnaire
		Group 2			paper based catalogue							

Figure 2. Overview of the randomized field experiment

We prepared two flower meadow, each consisting of 100 fake flowers composed of four different flower pictures per flower species and covering a selection of 25 different flower species from the flower dataset. The pictures were printed on paper and mounted on skewers. On the back side, the fake flowers were numbered according to an internal reference list to allow internal identification without the need for labels.

As a traditional tool for supporting participants in the experiment, we prepared a flowers catalogue covering exactly the 25 different flower species from the flower meadows. The flower pictures in the catalogue were different from those in the flower meadow and the catalogue was ordered alphabetically.

The questionnaire in the first section used a Likert scale containing five values from strongly disagree (1) to strongly agree (5) and covering eight questions:

- Perceived Usefulness:
  A. The AR app / catalogue was helpful to fulfil the task.
- Perceived Learning:
  B. With this activity I have learned something.
  C. I have learned about flowers.
  D. I can put together a bouquet on my own.
- Motivation: What do you think about the experiment and its setup?
  E. The introductory story was motivating.
  F. The task was simple and understandable.
  G. It was exciting to fulfill the task.
  H. The activity was entertaining.

Both rooms for the experiment where prepared in the same way. We set up the flower meadow with the fake flowers sticking in carton boxes and grouped by flower type. The carton boxes were placed on three tables in the center of the room. There was enough space to walk around the tables and to reach the flowers easily.

The main task for the participants was to collect six flowers from the meadow, which were named in form of a word-cloud on the instruction sheet in an envelope. We prepared five envelopes and the selection of the flower names for the word-cloud was done by a randomization process. Such, we used a webtool (www.randomizer.org) to collect 5 sets of 6 unique numbers per set within the range from 1 to 25. To arouse student attention and motivation, we narrated a story to send them on a mission, thus following design principles from game-based learning and simulation: “You fell in love with another person and have learned that you can break the ice between you and your crush with a smoothly arranged bouquet of flowers. Since you are absolutely unfamiliar with how to create a convincing flower bouquet, you ran a data analysis on your partner’s Facebook account and received a list of preferred flowers presented in the word-cloud below”. The mission to accomplish was formulated in the way, that “You know that love is like a little bird which flies away after some time and since you have just this one chance to score, give your best and collect the flowers as listed in the word cloud from the “self-service shop” as accurately and as fast as you can!”.

While the AR group could use a prepared iPhone (we used four iPhone 8+ and one iPhone X) to complete their mission, the control group (non-AR group) was provided with the aforementioned flowers catalogue. As noted earlier, we used different pictures for the catalogue and the production of the fake flowers.

The experimental process was designed in a way that after listening to the initial instruction participants were assigned an envelope with further instructions, the story, the mission, and either an iPhone or a flowers catalogue. Then the researcher started a timer and the students needed to collect the flowers as fast as possible. Afterwards, they came back to the researcher who recorded the collected flower numbers and asked students to complete the questionnaire. Since the students received a participant number, this number was noted on the questionnaire for later analysis. Once the participants completed all tasks, the fake flowers were put back to the flower meadows and the room was prepared for the next group.

5 Implementation

We invited 71 students from a Masters course in Information Technology at a technical university in northern Europe to participate in the experiment, but only 44 attended. The students were already divided into working groups from their course and we assigned them to sessions with a maximum of ten students per session and a duration of approximately 15 minutes. Participating students received a voucher from the university’s coffee shop as a reward right after the experiment.

The experiment started with a short introduction to welcome and thank the students for their participation. The participants were given some motivational instructions and were told to not chat with each other during the experiment or tell others about the experiment afterwards to not influence other students attending later. To split the group into the AR group (participants interacting with AR app during the experiment) and non-AR group (control group working with catalogue instead of AR app), students were told to choose between one of the two rooms by having equal numbered groups.

Participants could choose one of the five envelopes and when they started reading the instructions, a timer was set. After collecting the flowers, the students had to move to the research assistant and hand over their flower bouquet and all provided materials. To document the selected flowers and the required time to completion for the task, participants received a number to record their results for analysis. They then received the questionnaire to be answered on their own, marked with their participants number. After the students completed the questionnaire, they could leave the experiment.

Both experiment groups were treated in the same way, except of having different tools (AR app and paper catalogue) to fulfill the main task. There were no a priori time restrictions given, but students in the AR group were asked to terminate the collecting of flowers after 15 minutes.

6 Data Analysis

A participants' data record contained participant ID, group (AR, non-AR), gender (female, male), envelope number, IDs of the collected flowers, time to task completion, and the answers to the questions of the post-test questionnaire. In a first analysis, we assessed the number of correct flowers collected and the answers from the questionnaire. Overall, 18 female and 27 male students took part in the experiment, where 20 were assigned to the AR group and 25 the non-AR group.

In the AR group, 6 female and 14 male participants required from 510 to 1200 seconds to complete the given task (median 858.5 seconds, mean 864 seconds). They collected between 2 and 6 correct flowers from the given bouquet (median 4, mean 4.45). In terms of learning performance, the number of correct named flowers in their post-test questionnaire reached from 0 to 5 (median 2, mean 2.55).

Table 1. Correlation matrix

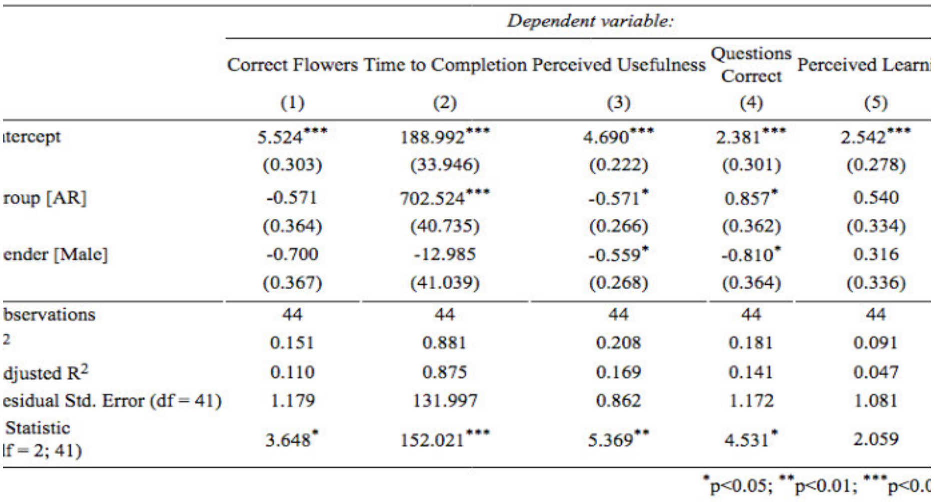
		Group	Gender	Envelope	t2compl	NoCorrFl	QuizRes	QA	QB	QC	QD	QE	QF	QG	QH
Group	Pearson Correlation	1	-,183	,113	-,927**	,287	-,229	,382**	-,241	-,352*	-,079	-,026	,099	-,014	,117
	Sig. (2-tailed)		,230	,458	,000	,056	,130	,010	,111	,018	,607	,866	,519	,930	,444
Gender	Pearson Correlation	-,183	1	-,220	,139	-,325*	-,283	-,363*	,154	,149	,150	,000	-,011	-,115	-,158
	Sig. (2-tailed)	,230		,146	,364	,029	,059	,014	,312	,329	,325	1,000	,942	,452	,300
Envelope	Pearson Correlation	,113	-,220	1	-,053	,122	,099	-,001	,065	-,098	,058	,231	-,091	,241	,235
	Sig. (2-tailed)	,458	,146		,732	,425	,516	,994	,670	,522	,703	,126	,552	,110	,121
t2compl	Pearson Correlation	-,927**	,139	-,053	1	-,306*	,295*	-,326*	,242	,324*	,120	,052	-,034	-,009	-,172
	Sig. (2-tailed)	,000	,364	,732		,041	,049	,029	,109	,030	,433	,733	,824	,951	,259
NoCorrFl	Pearson Correlation	,287	-,325*	,122	-,306*	1	,107	,176	-,227	-,281	-,053	-,037	-,013	,232	,201
	Sig. (2-tailed)	,056	,029	,425	,041		,485	,249	,133	,061	,730	,812	,933	,126	,186
QuizRes	Pearson Correlation	-,229	-,283	,099	,295*	,107	1	,166	-,248	-,119	,021	,230	,196	,183	,199
	Sig. (2-tailed)	,130	,059	,516	,049	,485		,276	,101	,435	,890	,128	,197	,228	,189
QA	Pearson Correlation	,382**	-,363*	-,001	-,326*	,176	,166	1	-,188	-,069	-,290	,257	,354*	,306*	,328*
	Sig. (2-tailed)	,010	,014	,994	,029	,249	,276		,216	,652	,053	,089	,017	,041	,028
QB	Pearson Correlation	-,241	,154	,065	,242	-,227	-,248	-,188	1	,709**	,428**	,109	,034	,241	,148
	Sig. (2-tailed)	,111	,312	,670	,109	,133	,101	,216		,000	,003	,476	,824	,111	,331
QC	Pearson Correlation	-,352*	,149	-,098	,324*	-,281	-,119	-,069	,709**	1	,322*	,148	-,031	,274	,238
	Sig. (2-tailed)	,018	,329	,522	,030	,061	,435	,652	,000		,031	,331	,842	,068	,116
QD	Pearson Correlation	-,079	,150	,058	,120	-,053	,021	-,290	,428**	,322*	1	,018	-,173	,109	,060
	Sig. (2-tailed)	,607	,325	,703	,433	,730	,890	,053	,003	,031		,908	,256	,475	,696
QE	Pearson Correlation	-,026	,000	,231	,052	-,037	,230	,257	,109	,148	,018	1	,175	,506**	,545**
	Sig. (2-tailed)	,866	1,000	,126	,733	,812	,128	,089	,476	,331	,908		,250	,000	,000
QF	Pearson Correlation	,099	-,011	-,091	-,034	-,013	,196	,354*	,034	-,031	-,173	,175	1	,221	,127
	Sig. (2-tailed)	,519	,942	,552	,824	,933	,197	,017	,824	,842	,256	,250		,144	,407
QG	Pearson Correlation	-,014	-,115	,241	-,009	,232	,183	,306*	,241	,274	,109	,506**	,221	1	,674**
	Sig. (2-tailed)	,930	,452	,110	,951	,126	,228	,041	,111	,068	,475	,000	,144		,000
QH	Pearson Correlation	,117	-,158	,235	-,172	,201	,199	,328*	,148	,238	,060	,545**	,127	,674**	1
	Sig. (2-tailed)	,444	,300	,121	,259	,186	,189	,028	,331	,116	,696	,000	,407	,000	

In the non-AR group, 12 female and 13 male participants needed between 68 and 330 seconds to complete the task (median 171 seconds, mean 182.24 seconds). They collected 0 to 6 correct flowers from the given bouquet (median 5, mean 5.16) and the number of correctly named flowers in the post-test questionnaire reached from 0 to 5 (median 2, mean 1.96). Between the two groups there was no difference in the distribution of envelopes, which was tested by performing a Kolmogorov Smirnov test.

Next, we ran an exploratory correlation analysis between all relevant pairs of variables in our dataset (Table 1). We found statistically significant correlations between group assignment and time to completion (mean of AR/non-AR: 864sec/182sec), perceived usefulness (QA) (mean of AR/non-AR: 3.65/4.40), and one of the questions related to perceived learning (QC) (mean of AR/non-AR: 3.4/2.8). Interestingly, we also found a significant correlation between gender and the number of correctly collected flowers (mean of female/male: 5.33/4.52, $p<0.01$), and perceived usefulness (mean of female/male: 4.50/3.77, $p<0.01$).

As our pseudo random assignment of students to groups did not produce an even distribution of males and females between the AR and non-AR group and because the correlation analysis indicated that gender is correlated with some of our dependent variables of interest, we decided to use regression models to test the main hypotheses of our experiment, namely that AR has a positive impact on (perceived) task performance and (perceived) learning performance. The advantage of a regression model over t-tests or ANOVA is in the ability to model the influence of multiple independent variables (in our case group and gender) on one dependent variable. Table 2 summarizes the results of this analysis.

Table 2. Regression results



	Dependent variable:				
	Correct Flowers	Time to Completion	Perceived Usefulness	Questions Correct	Perceived Learni
	(1)	(2)	(3)	(4)	(5)
itercept	5.524^{***}	188.992^{***}	4.690^{***}	2.381^{***}	2.542^{***}
	(0.303)	(33.946)	(0.222)	(0.301)	(0.278)
roup [AR]	-0.571	702.524^{***}	-0.571^{*}	0.857^{*}	0.540
	(0.364)	(40.735)	(0.266)	(0.362)	(0.334)
ender [Male]	-0.700	-12.985	-0.559^{*}	-0.810^{*}	0.316
	(0.367)	(41.039)	(0.268)	(0.364)	(0.336)
bservations	44	44	44	44	44
2	0.151	0.881	0.208	0.181	0.091
djusted R^2	0.110	0.875	0.169	0.141	0.047
esidual Std. Error (df = 41)	1.179	131.997	0.862	1.172	1.081
Statistic lf = 2; 41)	3.648^{*}	152.021^{***}	5.369^{**}	4.531^{*}	2.059

*p<0.05; **p<0.01; ***p<0.0

According to the regression results, participants in the AR group did not perform significantly better in terms of correctly identifying flowers than participants in the paper catalogue group. With regard to time needed to complete the task, participants in AR group even performed significantly worse than participants in the paper catalogue group. Hence, we did not find any empirical evidence that the AR app increased participants' objective task performance in terms of task accuracy and task time. Consistent with this finding, participants in the AR group evaluated the perceived usefulness of their tool (i.e. the AR app) significantly worse than participants in the non-AR group working with the paper catalogue.

However, when looking at objective learning performance, measured by the number of questions answered correctly in the post-test questionnaire, we found that participants in the AR group performed significantly better. This finding provides empirical support for the effectiveness of AR as a tool to enhance students' objective learning performance. With regard to perceived learning (measured by the average scores of questions B-D), we did not find a significant difference between the groups.

7 Discussion

In our experiment students achieved an observably better learning performance when using the AR flower identification app instead of a comparable paper catalogue, a result that is similar to prior research results comparing AR-based training to traditional paper-based training methods [4, 12, 22]. Therefore, and to answer RQ2, we conclude that AR can support students' learning performance. However, it may also be that the learning performance for the AR group was influenced by their longer task completion times, thus students were more engaged with the learning content and more motivated [2]. This can either be seen as a potential confounding factor which has to be controlled for in future studies (e.g. by predefining the available time for conducting a task), or as a positive side effect of using AR for teaching and learning [4, 5, 6, 12, 24, 25]. One could argue that when using AR students voluntarily spend more time with the learning materials, as compared to using traditional paper-based tools.

Considering participants' behavior during the experimental task, we noted that students in the AR group acted differently than those in the non-AR group. While participants in the AR group needed to investigate the flowers sequentially (because the app can only identify one object at a time) and thus examined nearly all flowers from the meadow, participants in the non-AR group selected a flower's name from the task description, searched for the name in the catalogue, and then located the flower by scanning the flower meadow with their eyes and matching the picture from the catalogue with the pictures on the meadow. On the one hand, this resulted in much shorter task times, as the human eye can focus on multiple objects at the same time (or at least can change focus much more quickly than AR technology), in comparison to the participants in the AR group who additionally had to perform the task of hand-eye coordination when using the app. On the other hand, when filling out the post-test questionnaire students realized that they had not inspected all flowers from the meadow and catalogue in sufficient detail in order to answer the questions correctly (the flowers students had to name in the post-test were different from those they had to collect).

A further observation related to the above point was that as participants in the AR group were forced by the app's functionality to look at each flower and since the app showed the three most likely names for identifying a flower and the elated confidence levels, students required more attempts to select the correct flower. We are convinced that this was a main driver behind the longer time needed to complete the task. Additionally, students from the AR group confirmed that it is more fun to look at the flowers with the app instead of just learning from a book.

It is remarkable that while the perceived learning of the AR group is not significantly higher compared to the non-AR group, their objective learning performance was significantly higher. The better objective learning performance may be explained by the different ways participants approached the task in the two groups. While students in the non-AR group focused on finding the flower picture for the given flower name and selecting a similar flower from the meadow, students in the AR group pointed their smartphone upon every single flower in the meadow to see its name. A single flower was represented multiple times in the flower meadow and students from the AR group visualized a particular flower more often. This finding corresponds to results from other studies, where AR is more effective than using traditional media [2, 4, 6, 10, 22, 24].

Since the paper catalogue prepared for the experiment was ordered alphabetically and only contained few pages covering the presented 25 flowers, students in the non-AR group had an advantage when matching flower names between the task description and catalogue. This could be a major limitation in our study in regard to the results for participants task performance times compared with participants from the AR group. Using a flower identification book with hundreds of pages ordered by species instead of alphabetically would have been more realistic for our comparison and would probably have led to different results, at least in terms of task completion times. However, this observation indicates that the prepared catalogue was designed to support task completion.

Our app is technically able to identify up to 60 pictures per second, comparing it with several thousands of pictures from the database. Thus, the setup of the experiment with only a handful of flowers did not challenge the full potential of the app, which is a further limitation in terms of system performance in comparison of traditional tools with AR based tools. Nonetheless, with our study we could contribute to the discussion about improvement of AR recognition and marker-less AR [4]. For future research and practical application, the AR app can be utilized in any other learning environment just by exchanging the underlying image recognition machine learning model. This represents a cost-efficient alternative to integrate AR into classroom trainings [22].

8 Conclusion

With the app development and its application in the experiment we could answer our RQ 1 and demonstrate how marker-less AR can be implemented for education in a real-world environment. Thus, we followed recommendations for further research in the directions of implementing AR in real-life settings [6] and applying image-based tracking [5] and marker-less AR [4, 21, 26] for ubiquitous learning [2]. Moreover, with our study we investigated how marker-less AR affects task and learning performance in a mundane setting, for example in our simulation of a florist’s job. Our results showed that from a learning aspect, students using the AR app performed better when it comes to recalling the learning content, similar to prior studies [2, 6, 10, 24]. Although students in both groups achieved the same level of accuracy in fulfilling the given task, those students in the AR group needed more time. Since the experimental setup unintentionally supported the control group in faster task completion time, which points towards the finding that tasks processed with AR need to be designed differently.

Relying on the predefined dataset and machine learning model from Nilsback & Zisserman [20] was an efficient decision and guaranteed a consistent recognition rate for each object in the experiment. However, participants had some troubles with finding the correct focus for the fake flowers because of reflections, shadows and different illumination caused by the changing daylight which is also mentioned in prior studies and therefore a limitation which should be investigated in future research [2, 11, 21].

Students from the non-AR group benefited from the reduced catalogue to accomplish their mission. Since the AR app is able to recognize up to 60 pictures in a second from a dataset containing 8,189 pictures, the comparison of both tools in the experiment and for the given task, to search and identify a flower by its given name, was not really fair. However, with our study we demonstrate a content application of AR in association with its benefits and directions, particularly its scalability in a mundane situation.

Conducting an experiment just with students is not always satisfying. However, in our larger research program this was only a first test to demonstrate the use of the marker-less AR app and to collect and analyze first empirical data to investigate its effectiveness. In fact, we are beyond this now and are currently testing the app with a target group of low-threshold skilled employees.

From the aspect of using marker-less AR in educational settings we have ascertained that the recognition sometimes lacks due to optical influences, which is still a common issue for AR applications [2, 4, 5, 12, 21, 24, 25, 26]. Hence, future technological development should focus on recognition algorithms and the preparation of large and validated datasets in order to support the implementation of marker-less AR in education and in various real-life situations. Furthermore, the application of object detection instead of image recognition inside AR applications provides potential for new findings about how full 3D support for such AR apps assists learning and a better understanding. First results from our continuing research already confirm that object detection facilitates the recognition of a series of objects in one single viewpoint.

References

1. Anderson, T.: Theories for learning with emerging technologies. Emerging technologies in distance education (2016)
2. Akçayır, M., Akçayır, G.: Advantages and challenges associated with augmented reality for education: A systematic review of the literature. In: Educational Research Review, vol. 20, pp. 1-11 (2017)
3. Azuma, R. T.: A survey of augmented reality. Teleoperators & Virtual Environments, vol. 6(4), pp. 355-385 (1997)
4. Bacca, J., Baldiris, S., Fabregat, R., Graf, S.: Augmented reality trends in education: a systematic review of research and applications. In: Journal of Educational Technology & Society, vol. 17(4), pp. 133 (2014)

5. Billinghurst, M., Clark, A., Lee, G.: A survey of augmented reality. In: Foundations and Trends® in Human–Computer Interaction, vol. 8(2-3), pp. 73-272 (2015)
6. Chen, P., Liu, X., Cheng, W., Huang, R.: A review of using Augmented Reality in Education from 2011 to 2016. In: Innovations in Smart Learning, pp. 13-18. Springer, Singapore, (2017)
7. Creswell, J. W.: Research design: Qualitative & quantitative approaches. Sage Publications, Inc. (1994)
8. Dalal, N., Trigg, B.: Histogram of oriented gradients for human detection. In: Computer Vision and Pattern Recognition, vol. 1, pp. 886-893. IEEE Computer Society (2005)
9. DeLone, W. H., McLean, E. R.: Information systems success: The quest for the dependent variable. In: Information systems research, vol. 3(1), pp. 60-95 (1992)
10. Diegmann, P., Schmidt-Kraepelin, M., Van den Eynden, S., Basten, D.: Benefits of Augmented Reality in Educational Environments - A Systematic Literature Review. In: Wirtschaftsinformatik, vol. 3(6), pp. 1542-1556 (2015)
11. Dunleavy, M., Dede, C., Mitchell, R.: Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning. In: Journal of science Education and Technology, vol. 18(1), 7-22 (2009)
12. Dunleavy, M., Dede, C.: Augmented reality teaching and learning. In: Handbook of research on educational communications and technology, pp. 735-745. Springer, New York (2014)
13. Herrington, A., Herrington, J., Mantei, J.: Design principles for mobile learning (2009)
14. Hirumi, A., Appelman, B., Rieber, L., Van Eck, R.: Preparing instructional designers for game-based learning: Part 2. In: TechTrends, vol. 54(4), pp. 19 (2010)
15. Kolb, D. A.: Experiential learning: Experience as the source of learning and development. FT press (2014)
16. Larkin, J., Simon, H.: Why a diagram is (sometimes) worth ten thousand words. In: Cognitive Science, vol. 11 (1), 65–100 (1987)
17. Lowe, D.: Distinctive image features from scale-invariant Keypoints. In: International Journal of Computer Vision, vol. 60(2), pp. 91–110 (2004)
18. Mayer, R. E.: Multimedia learning (2nd). Cambridge University Press (2009)
19. McLellan, H.: Evaluation in a situated learning environment. In: Situated Learning perspectives, Englewood Cliffs, NJ: Educational Technology Publications (1996)
20. Nilsback, M. E., Zisserman, A.: Automated flower classification over a large number of classes. In: Proceedings of the Indian Conference on Computer Vision, Graphics and Image Processing (2008)
21. Palmarini, R., Erkoyuncu, J. A., Roy, R., Torabmostaedi, H.: A systematic review of augmented reality applications in maintenance. In: Robotics and Computer-Integrated Manufacturing, vol. 49, pp. 215-228 (2018)
22. Radu, I.: Augmented reality in education: a meta-review and cross-media analysis. In: Personal and Ubiquitous Computing, vol. 18(6), pp. 1533-1543 (2014)
23. Sauro, J.: A practical guide to the system usability scale: Background, benchmarks & best practices. Denver, CO: Measuring Usability LLC (2011)
24. Sommerauer, P., Mueller, O.: Augmented Reality for Teaching and Learning – a Literature Review on Theoretical and Empirical Foundations. In: Proceedings of the 26th European Conference on Information Systems (ECIS), Portsmouth, UK, June 23-28 (2018)
25. Van Krevelen, D. W., Poelman, R.: A survey of augmented reality technologies, applications and limitations. In: International journal of virtual reality, vol. 9(2), pp. 1 (2010)
26. Zhou, F., Duh, H. B. L., Billinghurst, M.: Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR. pp. 193-202. IEEE Computer Society (2008)