How gamification motivates: an experimental study of the museum experience using handheld augmented reality

Mikkel Mogensen, Raminta Stanulionyte

***Abstract*—As of late, the renewed interest in Augmented Reality (AR) has seen a wave of gamified applications for handheld devices. Gamification is one approach that can help increase engagement of users by introducing traditional game elements and participatory elements. In this paper we thoroughly investigate how gamification can help museum visitors engage with exhibits while using our own handheld augmented reality application called 'AR Explorer'. Moreover, we were interested in finding out to what extent there could be an interest in using this type of application. The 'AR Explorer' application was designed for museums, leveraging the functionality of augmented reality on smartphones. The purpose of the application was to preserve the visitors' contact with the exhibit, while allowing them to explore features (i.e. symbols, wear marks, damaged details) of the exhibition in detail. Two versions of the applications were developed, a gamified and non-gamified, that were used for the independent variable in a between-group control study carried out on 30 participants. A replica of a museum object was used for the experiment, where 10 features had been implemented. For each feature the participants interacted with, a quiz was shown that they had the choice either to answer or skip. We measured task completion for each of the groups from these quizzes, and likened the quality of answers to their level of engagement. A questionnaire was given to participants who finished the experiment which measured general usability and the experience using our application. The results of our study indicated that gamification increased engagement of participants. Furthermore, respondents of the questionnaires reported that they would be interested in using this or a similar application in the future.**

I. Introduction

Given the broadened interest in augmenting the human experience, the implications of such augmentation and the extent of which we allow ourselves to infringe on our senses is a disputed matter. The example that we wish to highlight is Mixed Reality (MR) applications. These types of applications have grown in popularity in recent years, largely owing to the influence of technology companies such as *Google*, *Apple*, and *Microsoft*. This among other things has contributed to increased commercial interest in the area. As the pace of adoption of MR quickens, cultural heritage sites have begun focusing on the possibilities for implementing this innovative technology. This paradigm shift points to the knowledge that: *"the typical museum visitor is discontent, restless and on a quest for stimuli"* [1]. Relevant to the pursuit of bettering the museum experience, or any other for that matter, is the extensive corpus on gamification and its positive effects on the engagement of the user. Gamification in cultural heritage and in particular museums has been an area of rigorous scientific studies [2] [3] [4]. However, many of the solutions presented in these papers do not preserve the contact between visitors and museum object i.e. information may be detached from the museum exhibit and/or occludes the exhibit. In this paper, we try to address this concern whilst investigating the interaction between gamified elements and the user experience using a novel Handheld Augmented Reality (HAR) application.

The aim of our research is to report on the factors of gamification influencing the engagement of museum visitors using such systems, and to which extent there could be an interest in these or similar applications.

In section 2 we provide a background on HAR and gamification and their role in cultural heritage. Section 3 presents our research methods and limitations. Section 4 will focus on technicalities of our application. In section 5 we outline our procedure before presenting the results of our experiment in section 6. The results are discussed in section 7 and a conclusion is drawn in section 8.

II. Background of HAR and Gamification in Cultural Heritage

The notion of "Mixed Reality" (MR) can be traced back to an academic paper by Milgram and Kishino [5]. It describes a subclass of technologies that involve the combination of real and virtual elements, better known as VR and AR. The latter has seen a surge in popularity over the past few years, due to renewed interest and advancements in the technology. Augmented Reality (AR) describes the concept of overlaying a view of the real world with artificial images. A majority of research and development in this field is devoted to its mobile applications for smartphones and tablets. Handheld AR (HAR) is a popular term among scholars that encapsulates this; as the term implies, it deals with handheld devices with see-through capabilities. To this day, *HAR* is presumably one of the fastest growing research areas in the field of augmented reality [6]. A motivating factor for this could very well be attributed to the*"emergence and widespread uptake of smart-phones that provide powerful platforms for supporting augmented reality on a mobile platform"* [7].

Mikkel Mogensen is with Aalborg University Copenhagen
Raminta Stanulionyte is with Aalborg University Copenhagen

According to [8] one of the important aspects of AR is its ability to*"[. . .] enhance a user's perception of and interaction with the real world"*. AR promises limitless possibilities in a wide range of industries i.e. it may be used as a backbone for education as well as for navigation and entertainment. AR is still rudimentary given that it has yet to reach its full potential and be normalized [9]. That being said it is already seeing a widespread adoption in many industries.

One of the areas that have taken a keen interest in AR is cultural heritage and tourism sites. In museums, for example, AR can be used to actively engage visitors by various means such as overlaying information about the exhibit or serve as a virtual tour guide. Museums have proven to be proactive in seeking to expand their audience and their offerings; in this regard, AR is a good match as it could help cater to technology-savvy individuals [10]. This should be taken in the context of research suggesting that visitors want more information and orientation in museums; it has also been found that for the majority of visitors, social and recreational leisure activities are just as important [11]. This goes to show that there is more to engaging visitors than the technology alone. In fact, there seems to be a movement toward greater entertainment-oriented experiences, that employ gamification principles to encourage visitors to explore and engage socially [12].

According to [13] gamification is "*[. . .] the deliberate implementation of features of games in contexts not normally associated with games. Organizations incorporate gamification in order to encourage participants to feel a certain way, exhibit a certain behaviour and/or perform a certain action, which may not occur otherwise.*" The increasing used of games in non-entertainment is a transformation which has the ability to not only immerse, engage and motivate but also sustain participation *[13].* Such use of gamification in the cultural heritage and tourism sites could altogether cater to the needs of both the museumgoers as well as the museum stakeholders, delivering essential information regarding exhibits in more user-friendly, enjoyable and entertaining manner.

III. Research Methods

After examining recent trends and the applications of AR in cultural heritage sites we have identified gamification as an efficient means to enhance the visitors' overall experiences. We now explain the details of our chosen research methods, the tasks and our hypotheses in our user study. The following questions guided our research:

1) *How can gamification be used to engage museum visitors using HAR applications on mobile devices?*
2) *To what extend could there be an interest in a HAR application for museums?*

To answer our research questions we developed an HAR application that allows users to investigate various features of the exhibit. By features we mean the intrinsic characteristics of the artifact:

- Wear marks (i.e. abrasions and fractures)
- Biomarkers (i.e. lipids)
- Symbols or carvings
- Missing or damaged details

Since engagement is a fairly complex issue we opted to measure both quantitative and qualitative data. The application gathers quantitative data from quizzes during the session and sends it to a remote server. We are specifically interested in data regarding task completion (on the basis of these quizzes) and time spend in the application. We define the task as answering correctly to the quizzes for selected features. The quality of these answers are measured and grouped as follows:

- Complete (the user answered correctly to a quiz)
- Skipped (the user decided to skip the quiz)
- Incorrect (the user answered incorrectly to a quiz)
- Re-tries (the user repeated the same quiz)

We included additional measures of play duration, amount of discovered and undiscovered features in addition to the measurements of task completion mentioned above.

For the qualitative aspect we employed a questionnaire comprised of 15 questions; the scale we used was a 7-point Semantic Differential (SD). Respondents are asked to rate their overall experience with the application and usability of the application. As a base for the questionnaire, we use the Standard Usability Questionnaires for Handheld Augmented Reality [14]. The HAR Usability Scale (HARUS) was developed specifically to evaluate the usability in term of perception and ergonomic issues. These qualities are also referred to as comprehensible and manipulable. Comprehensibility refers to the "*[. . .] ease of understanding the information presented by the HAR system*" and manipulability is the " *[. . .] ease of handling the HAR device as the user performs the task*" [15]. Both factors can be used to calculate a HAR score, although, we note that the HARUS score is not computed for this specific scenario, as we are not interested in the ergonomic issues. The full questionnaire can be found in the appendix. The HARUS was chosen mainly because it has good internal consistency according to Cronbach's alpha; we later check the internal consistency of our own questionnaire.

A user study was conducted on two groups using a between-group protocol to avoid carryover effects. Both groups used our HAR application on a mobile device where the dependent variable was the gamification of the application (refer to section 4). Willing participants were found and were randomly assigned to either the gamified or non-gamified condition. The following hypotheses were chosen to answer our first research question on the engagement of visitors:

- *H_1: There is a positive relationship between elements of gamification and play duration*
- *H_2: There is a positive relationship between the gamified experience and task completion*

Results from the questionnaires will be used to answer the

second research question on the the general interest in our application or similar ones.

IV. Application Overview

In the following section, we take a look at the 'AR Explorer' that was the outcome of our research and development. The resulting HAR application was deployed on Android operating systems and has two variations: gamified and non-gamified. We elaborate on the differences later on in this section.

The purpose of the HAR application was to study the effects of gamification in a controlled experiment comparing a gamified approach to a non-gamified. We anticipated the application to work as a natural extension of the exhibit. That is, we do not wish to remove the emphasis away from the legacy of the museum artifact. We believe that museum objects have a material and cultural value, a view that is shared by many experts [16].

The resulting application allows users to select and explore features on the museum exhibit using a 3D cursor overlaid on the object. Information about the feature is then made visible through the interface. At the same time, a 3D representation of the museum object provides a detailed view of the selected feature. Furthermore, we wanted to give users the ability to freely explore; this is based on research which suggests that "*[. . .] diversion, curiosity and spontaneity are more characteristic of visitor intentions than structured learning*" [11]. Users can explore features in any order and revisit features later on. Since one key aspect of the application was to preserve the contact to the exhibit, it was important to keep the application's design elements simple and homogeneous. We accomplish this by using reliable design guidelines for our application interface. This and more will be explained in detail later on.

Two different views were implemented in the application: E*xploration view*, F*eature view*. What follows is a description of each of these views, their purpose, features, and interfaces.

A. Exploration view

The exploration view is visible once the application starts. In the exploration view the user may view and select features on the museum object. Features that can be explored are shown using small colored indicators overlaid on the surface of the object. Depending on their color, the indicators disclose different game states:

- White: undiscovered (indicates that the user has not yet interacted with the feature)
- Red: incorrect answer (indicates a feature that has been discovered, but the quiz was answered incorrectly)
- Yellow: skipped (indicates that the feature was discovered, but the quiz was skipped)
- Green: correct answer (indicates a feature that has been discovered and that the quiz was answered correctly)

Indicators are important as they comprise a mean rhetoric, that communicates the procedures[1] and explain what actions are available to the user [17]. They also help the user to easily navigate the application.

1) Navigation: Navigation between features is accomplished by pointing a 3D cursor on any of the indicators and tapping on the screen. Tapping triggers a smooth transition where a 3D representation of the museum object moves to frame the symbol so it is displayed in the center of the screen (refer to the 'Exploration View' and 'Feature View' in figures 2 and 3). At the same time the application provides an auditory cue of the selection.

B. Feature view

After selecting a feature users are transported to the feature view (refer to the middle of figures 2 and 3). The screen is divided into a *viewport* (top) framing the feature, and *interface* (bottom) containing various GUI elements (text, images and buttons). The user is unable to navigate to other features while in the feature view.

1) Guidelines: The interface follows Google's material metaphor guidelines [18]. Having guidelines for design was important to streamline the experience and reduce mental effort. Material design puts emphasis on movements, as a way to respect the continuity of the experience even as the UI transforms and reorganizes. The layout of the interface is split in two parts: a top navigation bar and a content canvas containing a card. Cards help organize content such as text and images and can be manipulated by sliding them vertically on the screen.

2) Navigation : In the feature view the users' movement is restricted, in the sense that they cannot move between features. They can, however, navigate the interface which is done by tapping or sliding elements on the screen. On the lower half of the interface, in the content canvas, the user can scroll the text. In the navigation bar the user can select different options such as voice-over by tapping the note icon or exit the current activity by pressing the 'X' button. Upon exiting the activity the a quiz screen is presented to the users. From here they can choose to answer; this is done by toggling one of the radio buttons and pressing 'OK'. To skip the quiz the user can press the 'SKIP' button.

C. Differences between the gamified and non-gamified applications

Since we created a gamified and non-gamified application, some key features differ. Notably the 3D cursors used for navigating to different features and the achievement system that rewards players for certain actions. We elaborate on these differences below.

[1]Procedures referring to game procedures

1) 3D cursors: We created two different variations of the 3D cursor for the gamified and non-gamified version respectively (refer to figure 1):

- Non-gamified: circular 2D sprite
- Gamified: magnifying glass

The magnifying glass is a thematic cue that helps contextualizing and give meaning to rules, information and game elements; usually through metaphors [17]. We anticipated that the metaphor of a magnifying glass could motivate users to act like real-world explorer thereby encouraging them to engage and spend time with exhibit for an extended duration.

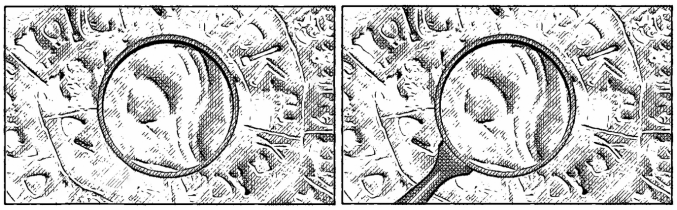

Figure 1. Non-gamified cursor (left) and gamified cursor (right)

2) Achievement system: An achievement system was implemented to the gamified application (refer to figure 3) that rewards the user for completing predefined tasks. The actions that triggers a reward are:

- Exploring an undiscovered feature (+5 pts)
- Answering correctly to a quiz (+10 pts)

The player score is displayed in the corner of the screen as a small icon with a counter that increments. When the score increases an animation is shown (see rightmost image in figure 3). This type of rhetoric can be characterized as valence feedback. Feedback rhetoric is necessary in games because it gives the user feedback about his actions while also increasing task completion [19]. We expected the motivational factor to be greater in the gamified application since users were rewarded with points. This might produce a measurable effect in our study in the task completion criterion (refer to chapter 5).

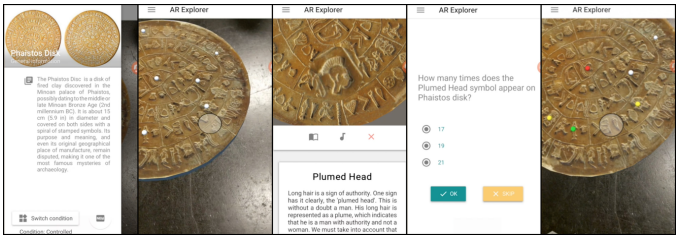

Figure 2. Non-gamified mobile application: Navigation drawer (1st on the left), exploration view (2nd from the left), feature view (middle), quiz screen (2nd from the right), exploration view with updated indicators (rightmost).

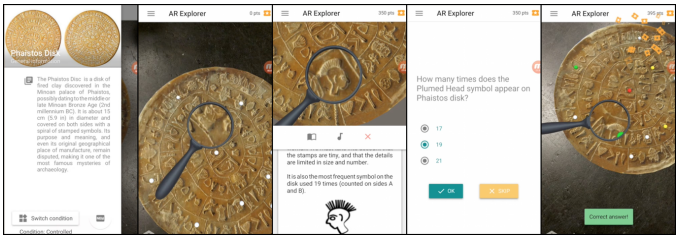

Figure 3. Gamified mobile application: Navigation drawer (1st on the left), exploration view (2nd from the left), feature view (middle), quiz screen (2nd from the right), exploration view with updated indicators (rightmost). Indications of the point system are visible on the top right corner of the app bar.

V. Procedure

In this section we present the experimental procedure that was carried out as a part of this study. We had two conditions, one control which had participants use the 'AR Explorer' application, without elements of gamification and experimental condition which had these elements (refer to table 1 below).

Table I. Dependent variables

Feature	Control	Experimental
3D cursor	Circular 2D sprite	Magnifying glass metaphor
Achievements	No	Yes

The compound effects of the dependent variables were studied in a between-group design where participants were randomly allocated. Each followed the same test procedure which we divide into three parts:

- Video introduction (different videos for each condition)
- Play session
- Questionnaire (different questionnaires for each condition)

Both tests were carried out on the university campus of Aalborg University (A. C. Meyers Vænge 15, 2450 København SV) during and after the lunch break. The video introduction was designed to minimize our influence on test participants and avoid bias during the learning phase. Each test condition had a video tailored to them because of the dependent variables. Both are similar in that they shortly described the purpose of the application in the museum context and introduced users to the museum object. The following elements were introduced in the video for the control condition: (1) *Cursor: circular 2D sprite*, (2) W*hite, green, yellow and red indicators*, (3) *Feature view*, (4) *Voice-over option*, (5) *Quiz screen*, (6) *Skip button*, (7) *End button.* In a similar fashion the video for experimental condition introduced the following: (1) *Cursor: magnifying glass*, (2) W*hite, green, yellow and red indicators*, (3) *Feature view*, (4) *Voice-over option*, (5) *Quiz screen*, (6) *Skip button*, (7) *Achievement system*, (8) *End button.* The lengths of both videos were approximately 2.5 minutes. Inevitably the video for the experimental conditions was a bit longer.

After the video presentation participants were asked to proceed with the session using the 'AR Explorer' application. In

each of the test conditions, participants had a chance to explore 10 features and answer 10 quizzes. In the video presentation, it was specified that participants did not have to explore all of the features and that they could stop playing at any time. The participants' performances were recorded automatically by the application. Once participants had explored the desired amount of features they were asked to fill out a questionnaire according to their test condition. Each questionnaire consisted of 20 questions regarding general information about testers (their age and gender), usability, engagement and the overall experience of the AR Explore application. Questionnaires for both conditions can be found in the appendix.

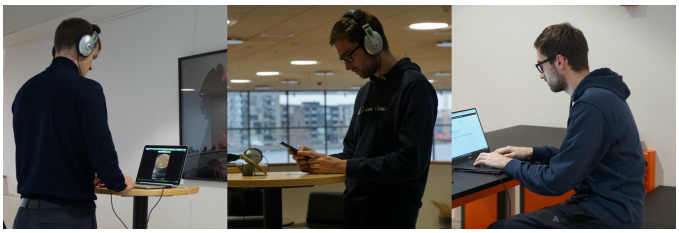

Figure 4. Visual representation of the test setup from left to right: Intro video (left), object exploration (middle), questionnaire (right)

VI. Results

We now report on the results of our user study with respect to task completion, engagement and overall experience of using our application. In total 30 participants in the age range 20 to 30 years, took part in our experiment. 86.7% of those were male and 13.3% female. The overall duration of the experiment varied from 12 to up to 20 minutes depending on the participant.

A. Task Completion

For each of the conditions, we obtained 15 recorded trials that were logged automatically through the application. After applying the K-S test for normality on the recorded data (refer to Table II) we see that the "Play duration" for the control condition $(D\,(15) = .173,\ p < .05)$ and the experimental condition $(D\,(15) = .126,\ p < .05)$ fulfill the normality criteria. The rest of the data pairs failed to meet the normality requirements $(p > .05)$. Levene's test (refer to Table IV) indicated equal variances $(F = 0.332,\ p = .569)$ revealing that the data for "Play duration" is parametric. We proceeded with an independent sample t-test (refer to Table IV). The results from the independent sample t-test indicated that play duration for the experimental condition $(M = 363.73,\ SD = 120.88)$ was not significantly different from the control condition $(M = 326.80,\ SD = 137.73),\ t(28) = 0.7806,\ p > .05$. The effect size or this analysis $(d = 0.284783)$ was found to be small when refering to Cohen's convention [20].

The remaining data was evaluated using non-parametric tests. From the rank table of Mann-Whitney (refer to table V) we see that mean of ranks as well as sums of ranks visibly differ for both conditions. However, looking at the significance value $(\alpha = .05)$ for Mann-Whitney U test (refer to table VI), it applies only for three data samples "Complete" $(U = 52.5,\ Z = -2.561,\ p = .010)$, "Discovered" $(U = 68.0,\ Z = -2.091,\ p = .037)$ and "Incorrect" $(U = 67.5,\ Z = -2.027,\ p = .043)$. The three remaining data pairs ("Skipped", "Undiscovered" and "Retries") showed no significance between experimental and control conditions.

Table II. Normality test

		Kolmogorov-Smirnova[a]		
	Condition	**Statistic**	**df**	**Sig.**
Play duration	Control	.173	15	.200*
	Experimental	.126	15	.200*
Complete	Control	.190	15	.149
	Experimental	.361	15	.000
Skipped	Control	.437	15	.000
	Experimental	.514	15	.000
Incorrect	Control	.345	15	.000
	Experimental	.438	15	.000
Discovered	Control	.242	15	.019
	Experimental	.485	15	.000
Undiscovered	Control	.247	15	.014
	Experimental	.485	15	.000
Retries	Control	.303	15	.001
	Experimental	.161	15	.200*

*. This is a lower bound of the true significance.
a. Lilliefors Significance Correction

Table III. Group Statistics for Play duration data pair

Condition	**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**
Control	15	326.8000	137.7291	35.5615
Experimental	15	363.7333	120.8757	31.2100

B. User experience

The questionnaires consisting of 13 items showed high internal consistency for the control condition $(\alpha = .733)$ and

Table IV. Independent Samples Tests on "Play duration"

Levene's Test for Equality of Variances		**t-test for Equality of Means**	
F	**Sig.**	**t**	**df**
0.3323	0.5689	0.7806	28
t-test for Equality of Means			
Sig. (right tailed)	**Mean Difference**	**Std. Error Difference**	
0.2208	36.9333	-16.8535	

Table V. Ranks for task completion

	Condition	**N**	**Mean Rank**	**Sum of Ranks**
Completed	Control	15	11.50	172.50
	Experimental	15	19.50	292.50
	Total	30		
Skipped	Control	15	16.70	250.50
	Experimental	15	14.30	214.50
	Total	30		
Incorrect	Control	15	18.50	277.50
	Experimental	15	12.50	187.50
	Total	30		
Discovered	Control	15	12.53	188.00
	Experimental	15	18.47	277.00
	Total	30		
Undiscovered	Control	15	18.27	274.00
	Experimental	15	12.73	191.00
	Total	30		
Retries	Control	15	12.67	190.00
	Experimental	15	18.33	275.00
	Total	30		

Table VI. TEST STATISTICS FOR TASK COMPLETION.[A]

	Complete	Skipped	Incorrect
Mann-Whitney U	52.500	94.500	67.500
Wilcoxon W	172.500	214.500	187.500
Z	-2.561	-1.070	-2.027
Asymp. Sig. (2-tailed)	.010	.285	.043
Asymp. Sig. (1-tailed)	.005	.1425	.0215
Exact Sig. [2*(1-tailed Sig.)]	.011[b]	.461[b]	.061[b]
	Discovered	**Undiscovered**	**Retries**
Mann-Whitney U	68.000	71.000	70.000
Wilcoxon W	188.000	191.000	190.000
Z	-2.091	-1.950	-1.809
Asymp. Sig. (2-tailed)	.037	.051	.071
Asymp. Sig. (1-tailed)	.0185	.0255	.0355
Exact Sig. [2*(1-tailed Sig.)]	.067[b]	.089[b]	.081[b]

a. Grouping Variable: Condition
b. Not corrected for ties.

Table VII. DESCRIPTIVE STATISTICS (CONTROL)

#	N	Min	Max	Mean	Median	Mode	Std. Deviation
Q1	15	4	6	4.80	5	5	0.561
Q2	15	4	6	5.00	5	5	0.378
Q3*	15	1	6	4.27	4	4[a]	1.280
Q4	15	3	6	4.73	5	5	0.799
Q5*	15	3	6	4.93	5	5[a]	0.961
Q6	15	4	6	5.20	5	5[a]	0.775
Q7*	15	3	6	5.00	5	5	0.845
Q8	15	4	6	5.07	5	5	0.704
Q9	15	2	6	4.87	5	6	1.356
Q10	15	2	6	5.07	5	5	1.163
Q11*	15	1	6	3.53	3	2[a]	1.922
Q12	15	0	6	3.87	4	3	1.807
Q13*	15	0	6	4.00	4	6	2.000
Q14	15	0	6	3.93	5	5	1.870
Q15	15	1	6	2.87	3	3	1.457

a. Multiple modes exist. The smallest value is shown
*. Reverse coded questions

for the experimental condition ($\alpha = .745$). Note that we reverse coded items that were negatively worded (Q3, Q5, Q7, Q11, Q13) to correlate responses in the same direction.

Table VII shows the descriptive results of responses in the control condition. On most items the mode is relatively high; we find a mode of 6 on both Q9 ("I thought that the information displayed on screen was consistent.") and Q13 ("The object itself was not interesting at all."). Note that since we reverse coded, the values should be interpreted positively. For Q13 we see that the median is just above neutral ($Mdn_{Q13} = 4$) and the minimum value is 0. The standard deviation for Q13 is also very high ($SD_{Q13} = 2.000$). Q5 and Q6 have multiple modes where the lower is shown in the table VII. As the lowest mode is 5 for these the highest mode is 6. The lowest mode is found for Q11 ("I did not want to learn about every single symbol on Phaistos Disk."). For the same question the median was lower than the mode ($Mdn_{Q11} = 3$). The highest mean is found for Q6 ($M_{Q6} = 5.20$) ("I felt that the information display was responding fast enough.") and the lowest for Q15 ($M_{Q15} = 2.87$) ("I enjoyed the voice over functionality while learning about the object.").

On figure 5 we show a summary of the responses gathered from the control condition is shown. Responses are generally positive except for a few (Q3, Q11, Q12, Q13, Q14, Q15). The figure reiterates the results from earlier. In particular, since respondents seem to be very positive we would expect the vast majority of modes to be above 3 (neutral). This seem to be the case as 9 of the 15 questions have a mode of 5. Moreover, it is observed that Q3-Q9 from the HARUS questionnaire are generally positive. Questions regarding user experience are mixed (Q10-Q15) except when asked about the overall experience and interest in the application (Q1, Q2).

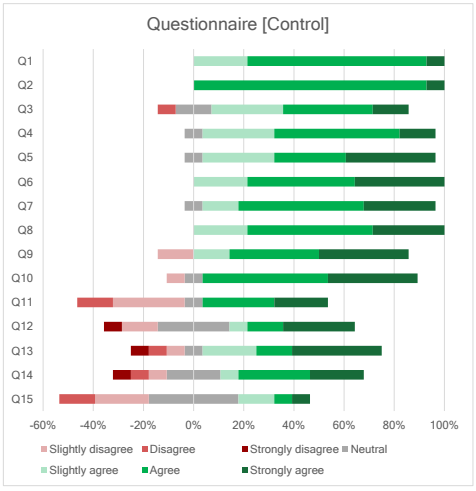

Figure 5. Summary of responses for the control condition

The sentiment in regards to several of the features were measured in the questionnaire. Participants rated the features either positively (+1) or negatively (-1). The sum of these scores gives us the results as seen in figure 6 . For the control condition, all of the scores were positive. The highest scoring features "were information about the object's features" (+12), "Zoom animation on object selection" (+11) and "Navigation (on the object)" (+11). The lowest scoring was the "Cursor (that appears on the object)" (+2).

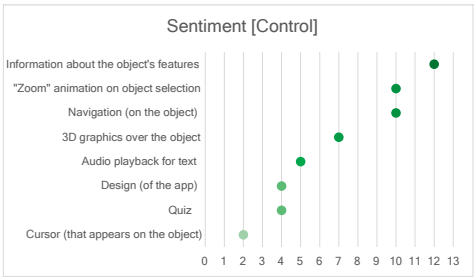

Figure 6. Sentiment rated on features for the control condition

Table VIII shows the descriptive results of responses in the experimental condition. Similar to control condition the mode among most of the respondents rated the experimental condition positively, however, the mode never reached 6. The highest mode is 5 and as is the case with in control condition the majority had this value (Q1, Q2, Q6, Q7, Q8, Q9, Q10, Q14). The lowest mode was 2 and can be found on responses

Table VIII. DESCRIPTIVE STATISTICS (EXPERIMENTAL)

#	N	Min	Max	Mean	Median	Mode	Std. Deviation
Q1	15	3	6	4.47	5	5	0.834
Q2	15	4	6	5.07	5	5	0.799
Q3*	15	1	6	3.13	3	4	1.506
Q4	15	2	6	4.07	4[a]	4	1.280
Q5*	15	2	6	4.40	4	4	1.056
Q6	15	2	6	4.60	5	5	1.298
Q7*	15	3	6	4.80	5	5	0.775
Q8	15	2	6	4.73	5	5	1.223
Q9	15	4	6	5.07	5	5	0.704
Q10	15	3	6	4.53	5	5	0.990
Q11*	15	1	6	3.27	3	3	1.710
Q12	15	1	6	3.47	3[a]	2	1.685
Q13*	15	2	6	3.93	4[a]	3	1.280
Q14	15	1	6	4.20	5	5	1.781
Q15	15	0	6	3.00	3	3	1.813

a. Multiple modes exist. The smallest value is shown
*. Reverse coded questions

for Q12 ("I wanted to learn about every single symbol on Phaistos Disk."), however, notice that there are two modes for this particular question. The highest mean value belonged to Q2 ($M_{Q2} = 5.07$) ("Would you like to use this or similar application at a museum?") and Q9 ($M_{Q9} = 5.07$) ("I thought that the information displayed on screen was consistent."). Both of the question have small standard deviations ($SD_{Q2} = 0,799 \mid SD_{Q9} = 0,704$) and small differences between Min and Max values ($Min_{Q2} = 4,\ Max_{Q2} = 6 \mid Min_{Q9} = 4,\ Max_{Q9} = 6$).

On figure 7 we show a summary of the responses gathered from the experimental condition is shown. Responses are generally positive, but a negative responses are prevalent in Q3, Q11, Q12, Q14 and Q15. It is observed that Q3-Q9 from the HARUS questionnaire are generally positive, but contain negative responses (especially the case in Q3). Questions regarding user experience are mixed (Q10-Q15) except when asked about the overall experience and interest in the application (Q1, Q2).

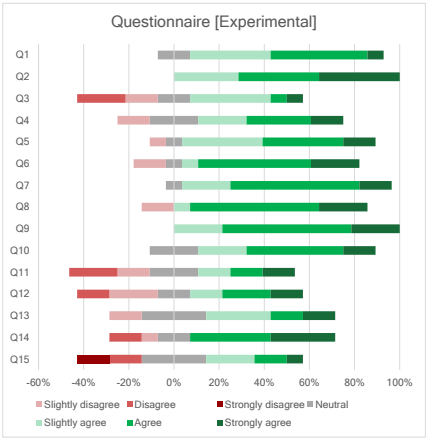

Figure 7. Summary of responses for the experimental condition

The sentiment analysis for the experimental condition was primarily positive except for one of the feature "Magnifying glass" which scored -2 (see figure 8). The highest scoring features were "Information about the object's features" and "Design (of the app)" which scored 11 and 10 respectively. "Zoom animation on object selection" again scored relatively high (+6). The same goes for "Animation when receiving points" and "Points (for exploration).

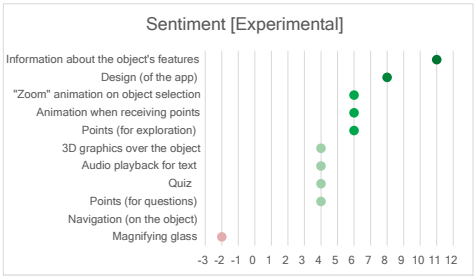

Figure 8. Sentiment rated on features for the experimental condition

A Mann-Whitney U test indicated that Q3 ("*I think that interacting with this application requires a lot of mental effort.*") required a higher mental effort in the experimental condition ($Mdn = 3$) than for the control condition ($Mdn = 4$), $U = 61.5$, $p = .030$. For the remaining questions the Mann-Whitney U test did not show any significant difference in the two conditions.

VII. DISCUSSION

In this chapter, we will briefly explain the findings following our experiment. After that follows a discussion about the obtained results, our interpretations, and limitations of the testing procedure. To summarize the results of the hypotheses, H_1 ("*There is a positive relationship between elements of gamification and play duration*") from the independent sample t-test showed no significant difference in 'play duration' between the two conditions. We therefore failed to reject the null hypothesis. We rejected the null hypothesis for H_2 (" *There is a positive relationship between the gamified experience and task completion* ") as task completion was significantly higher in the experimental condition according to the independent sample T-test. The following research questions we stated at the outset of the paper:

1) *How can gamification be used to engage museum visitors using HAR applications on mobile devices?*
2) *To what extent could there be an interest in a HAR application for museums?*

We found that gamification can be used to engage visitors, but in this case, it presents a compromise between task completion (as an indicator of engagement) and user experience. In general, it looked as though participants were more discontent with the interface and 3D cursor in the experimental condition. However, a non-parametric test revealed that only the perceived mental effort was higher for the experimental condition. Quantitative data gathered from the sessions illustrate that although there was a negative effect on the perceived mental effort from the gamified application, a measurable and significant effect for task completion was readily apparent.

Table IX. RANKS FOR QUESTIONNAIRES

	Condition	N	Mean Rank	Sum of Ranks
Q1	Controlled	15	17.17	257.50
	Experimental	15	13.83	207.50
	Total	30		
Q2	Controlled	15	15.03	225.50
	Experimental	15	15.97	239.50
	Total	30		
Q3	Controlled	15	18.90	283.50
	Experimental	15	12.10	181.50
	Total	30		
Q4	Controlled	15	17.83	267.50
	Experimental	15	13.17	197.50
	Total	30		
Q5	Controlled	15	17.60	264.00
	Experimental	15	13.40	201.00
	Total	30		
Q6	Controlled	15	17.30	259.50
	Experimental	15	13.70	205.50
	Total	30		
Q7	Controlled	15	16.63	249.50
	Experimental	15	14.37	215.50
	Total	30		
Q8	Controlled	15	16.10	241.50
	Experimental	15	14.90	223.50
	Total	30		
Q9	Controlled	15	15.73	236.00
	Experimental	15	15.27	229.00
	Total	30		
Q10	Controlled	15	15.50	232.50
	Experimental	15	15.50	232.50
	Total	30		
Q11	Controlled	15	16.40	246.00
	Experimental	15	14.60	219.00
	Total	30		
Q12	Controlled	15	15.67	235.00
	Experimental	15	15.33	230.00
	Total	30		
Q13	Controlled	15	16.20	243.00
	Experimental	15	14.80	222.00
	Total	30		
Q14	Controlled	15	15.70	235.50
	Experimental	15	15.30	229.50
	Total	30		
Q15	Controlled	15	15.27	229.00
	Experimental	15	15.73	236.00
	Total	30		

Table X. TEST STATISTICS FOR QUESTIONNAIRES

#	Mann-Whitney U	W[a]	Z	Asymp. Sig.[b]	Exact Sig.[c]
Q1	87.500	207.500	-1.166	.244	.305[d]
Q2	105.500	225.500	-.339	.735	.775[d]
Q3	61.500	181.500	-2.173	.030	.033[d]
Q4	77.500	197.500	-1.519	.129	.148[d]
Q5	81.000	201.000	-1.367	.172	.202[d]
Q6	85.500	205.500	-1.201	.230	.267[d]
Q7	95.500	215.500	-.785	.432	.486[d]
Q8	103.500	223.500	-.416	.677	.713[d]
Q9	109.000	229.000	-.155	.877	.902[d]
Q10	112.500	232.500	.000	1.000	1.000[d]
Q11	99.000	219.000	-.568	.570	.595[d]
Q12	110.000	230.000	-.105	.916	.935[d]
Q13	102.000	222.000	-.444	.657	.683[d]
Q14	109.500	229.500	-.129	.897	.902[d]
Q15	109.000	229.000	-.148	.882	.902[d]

a. Wilcoxon W
b. 2-tailed
c. [2*(1-tailed Sig.)]
d. Not corrected for ties

Participants completed more task in the experimental condition than the control condition while spending the same amount of time. Taken together with the fact that we could not measure a significant difference in skipped quizzes it goes to show that the quality of the answers in the gamified application was higher than for the non-gamified. To reiterate, we equate this quality of answers with higher engagement. To answer the second question, we looked at the data acquired from respondents of the questionnaires. Overall the results indicate that there is a huge interest in HAR applications for museums; this is true for responses gathered from both conditions.

A. Quantitative data

Despite the fact that the difference in 'play duration' for the control and experimental condition was insignificant and failed to reject the null hypothesis in the independent sample T-test, it has to be mentioned that we had a small sample size for both conditions (N = 15). This is apparent when looking at the calculated effect size $(d = 0.284783)$ which was small according to Cohen's convention with respect to play duration. This leads us to believe that we need a larger sample size as it would produce a more pronounced effect.

The rest of the data pairs turned out to be non-parametric due to the lack of normal distribution among the results. The violation of the normality condition of the sampling distributions could be attributed to several factors such as: irregularities during the experiment or insufficient sample size as previously mention. It could also be that some participants' were biased because they have recruited from the university, possibly during their break. It would have made sense to sign up people a few days in advance to ensure that time was not at the essence. We could also have asked participants if they were in a hurry and use it as a criterion to filter the data. Software issues could also explain some of the observations, but it is unlikely, as we did not encounter any when pilot testing the application. One thing that we did notice within the application was that participants forgot to end the experiment, causing the timer to run overtime. We did our best to manually check this after each session.

B. Qualitative data

In terms of users' feedback, we found that many of the responses did not differ significantly for the control and experimental condition. Participants did express that experimental condition required more mental effort than the control condition and we confirmed this using a non-parametric test. We can recapitulate many of the above points as for why there did not seem to be a huge effect from the experimental condition. For instance, a bigger sample size would certainly help to detect a meaningful difference. The sentiment analysis did show us that people were not very pleased with the magnifying glass compared to circular sprite for the 3D cursor, but again, it does not produce a significant difference in perceived usability. A few of the issues in the application leading to higher mental strain can probably be mitigated in

future iterations, by tweaking or changing the interface in accordance with gathered feedback.

VIII. Conclusion

In this research paper we presented a study concerning the use of HAR application in cultural heritage sites. Moreover, we examined the effect of gamification on the aforementioned application, to determine how it can be used to engage museum visitors and to what extent there could be an interest in this type of application. Results from evaluations of the "AR Explore" application was indicative of significantly greater number of tasks completed in the gamified application, compared to the non-gamified. Moreover, since there was no significant difference in numbers of skipped quizzes, this is equated with higher quality answers and a higher level of engagement. A small impact on the perceived mental effort was evident from the questionnaire data, but it could be attributed to differences in the interface. In regards to the interest in HAR applications, respondents of the questionnaire reported that there were very interested in using this or similar application in museums.

IX. Appendix

Questionnaire for control condition:

Qa: Gender:

1. Male
2. Female

Qb: Age:

1. 18 or Under
2. 19 - 22
3. 23 - 26
4. 27 - 30
5. Over 30

Q1*: How do you rate your overall experience using the AR Explore application?

Q1.1: Shortly explain your overall experience:

Q2*: Would you like to use this or similar application at a museum?

Q3*: I think that interacting with this application requires a lot of mental effort.

Q4*: I thought the amount of information displayed on screen was appropriate.

Q5*: I thought that the information displayed on screen was difficult to read.

Q6*: I felt that the information display was responding fast enough.

Q7*: I thought that the information displayed on screen was confusing.

Q8*: I thought the words and symbols on screen were easy to read.

Q9*: I thought that the information displayed on screen was consistent.

Q10*: I was interested in exploring the object in detail.

Q11*: I did not want to learn about every single symbol on Phaistos Disk.

Q12*: I wanted to learn about every single symbol on Phaistos Disk.

Q13*: The object itself was not interesting at all.

Q14*: I prefer reading the text rather than listening to the recording of the information text.

Q15*: I enjoyed the voice over functionality while learning about the object.

Q.A: Which application elements you found **attractive**? (Choose as many as you like)

1. Quiz
2. Cursor (that appears on the object)
3. Information about the object's features
4. Audio playback for text
5. Design (of the app)
6. Navigation (on the object)
7. 3D graphics over the object
8. "Zoom" animation on object selection

Q.B: Which application elements you found **NOT** attractive? (Choose as many as you like)

1. Quiz
2. Cursor (that appears on the object)
3. Information about the object's features
4. Audio playback for text
5. Design (of the app)
6. Navigation (on the object)
7. 3D graphics over the object
8. "Zoom" animation on object selection

**** 7-point Semantic Differential scale (1 - Strongly disagree, 4 - neutral, 7 - Strongly agree)***

Questionnaire for experimental condition:

Q.a: Gender:

3. Male
4. Female

Q.b: Age:

6. 18 or Under
7. 19 - 22
8. 23 - 26
9. 27 - 30
10. Over 30

Q1*: How do you rate your overall experience using the AR Explore application?

Q1.1: Shortly explain your overall experience:

Q2*: Would you like to use this or similar application at a museum?

Q3*: I think that interacting with this application requires a lot of mental effort.

Q4*: I thought the amount of information displayed on screen was appropriate.

Q5*: I thought that the information displayed on screen was difficult to read.

Q6*: I felt that the information display was responding fast enough.

Q7*: I thought that the information displayed on screen was confusing.

Q8*: I thought the words and symbols on screen were easy to read.

Q9*: I thought that the information displayed on screen was consistent.

Q10*: I was interested in exploring the object in detail.

Q11*: I did not want to learn about every single symbol on Phaistos Disk.

Q12*: I wanted to learn about every single symbol on Phaistos Disk.

Q13*: The object itself was not interesting at all.

Q14*: I prefer reading the text rather than listening to the recording of the information text.

Q15*: I enjoyed the voice over functionality while learning about the object.

QA: Which application elements you found **attractive**? (Choose as many as you like)

1. Quiz
2. Magnifying glass
3. Information about the object's features
4. Audio playback for text
5. Design (of the app)
6. Navigation (on the object)
7. 3D graphics over the object
8. "Zoom" animation on object selection
9. Points (for exploration)
10. Animation when receiving points

QB: Which application elements you found **NOT** attractive? (Choose as many as you like)

1. Quiz
2. Magnifying glass
3. Information about the object's features
4.Audio playback for text
5. Design (of the app)
6. Navigation (on the object)
7. 3D graphics over the object
8. "Zoom" animation on object selection
9. Points (for exploration)
10. Animation when receiving points

7-point Semantic Differential scale (1 - Strongly disagree, 4 - neutral, 7 - Strongly agree)

REFERENCES

[1] L. Neuburger and R. Egger, "Augmented Reality: Providing a Different Dimension for Museum Visitors," in *Augmented Reality and Virtual Reality*. Springer International Publishing, sep 2017, pp. 65–77. [Online]. Available: https://doi.org/10.1007%2F978-3-319-64027-3_5

[2] F. Liarokapis, P. Petridis, D. Andrews, and S. de Freitas, "Multimodal Serious Games Technologies for Cultural Heritage," in *Mixed Reality and Gamification for Cultural Heritage*. Springer International Publishing, 2017. [Online]. Available: https://doi.org/10.1007%2F978-3-319-49607-8_15

[3] R. Hammady, M. Ma, and N. Temple, "Augmented Reality and Gamification in Heritage Museums," in *Serious Games*. Springer International Publishing, 2016, pp. 181–187. [Online]. Available: https://doi.org/10.1007%2F978-3-319-45841-0_17

[4] S. de los Ríos, M. F. Cabrera-Umpiérrez, M. T. Arredondo, M. Páramo, B. Baranski, J. Meis, M. Gerhard, B. Prados, L. Pérez, and M. del Mar Villafranca, "Using Augmented Reality and Social Media in Mobile Applications to Engage People on Cultural Sites," in *Lecture Notes in Computer Science*. Springer International Publishing, 2014, pp. 662–672. [Online]. Available: https://doi.org/10.1007%2F978-3-319-07440-5_60

[5] P. Milgram and F. Kishino, "A Taxonomy of Mixed Reality Visual Displays," *IEICE Trans. Information Systems*, vol. vol. E77-D, no. 12, pp. 1321–1329, dec 1994.

[6] R. Azuma, M. Billinghurst, and G. Klinker, "Special Section on Mobile Augmented Reality," *Computers & Graphics*, vol. 35, no. 4, pp. vii–viii, aug 2011. [Online]. Available: https://doi.org/10.1016%2Fj.cag.2011.05.002

[7] A.-C. Haugstvedt and J. Krogstie, "Mobile augmented reality for cultural heritage: A technology acceptance study," in *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, nov 2012. [Online]. Available: https://doi.org/10.1109%2Fismar.2012.6402563

[8] R. T. Azuma, "A Survey of Augmented Reality," *Presence: Teleoperators and Virtual Environments*, vol. 6, no. 4, pp. 355–385, aug 1997. [Online]. Available: https://doi.org/10.1162%2Fpres.1997.6.4.355

[9] D.-I. Han, T. Jung, and A. Gibson, "Dublin AR: Implementing Augmented Reality in Tourism," in *Information and Communication Technologies in Tourism 2014*. Springer International Publishing, 2013, pp. 511–523. [Online]. Available: https://doi.org/10.1007%2F978-3-319-03973-2_37

[10] N. Kotler and P. Kotler, "Can Museums be All Things to All People?: Missions Goals, and Marketing's Role," *Museum Management and Curatorship*, vol. 18, no. 3, pp. 271–287, jan 2000. [Online]. Available: https://doi.org/10.1080%2F09647770000301803

[11] ——, "Can Museums be All Things to All People?: Missions Goals, and Marketing's Role," *Museum Management and Curatorship*, vol. 18, no. 3, pp. 271–287, jan 2000. [Online]. Available: https://doi.org/10.1080%2F09647770000301803

[12] S. de los Ríos, M. F. Cabrera-Umpiérrez, M. T. Arredondo, M. Páramo, B. Baranski, J. Meis, M. Gerhard, B. Prados, L. Pérez, and M. del Mar Villafranca, "Using Augmented Reality and Social Media in Mobile Applications to Engage People on Cultural Sites," in *UAHCI 2014: Universal Access in Human-Computer Interaction. Universal Access to Information and Knowledge*, C. Stephanidis and M. Antona, Eds. Cham: Springer International Publishing, 2014, pp. 662–672, dOI: 10.1007/978-3-319-07440-5_60. [Online]. Available: https://doi.org/10.1007/978-3-319-07440-5_60

[13] F. Liarokapis, P. Petridis, D. Andrews, and S. de Freitas, "Multimodal Serious Games Technologies for Cultural Heritage," in *Mixed Reality and Gamification for Cultural Heritage*. Springer International Publishing, 2017. [Online]. Available: https://doi.org/10.1007%2F978-3-319-49607-8_15

[14] M. E. C. Santos, J. Polvi, T. Taketomi, G. Yamamoto, C. Sandor, and H. Kato, "Toward Standard Usability Questionnaires for Handheld Augmented Reality," *IEEE Computer Graphics and Applications*, vol. 35, no. 5, pp. 66–75, sep 2015. [Online]. Available: https://doi.org/10.1109%2Fmcg.2015.94

[15] ——, "Toward Standard Usability Questionnaires for Handheld Augmented Reality," *IEEE Computer Graphics and Applications*, vol. 35, no. 5, pp. 66–75, sep 2015. [Online]. Available: https://doi.org/10.1109%2Fmcg.2015.94

[16] "We ask the experts: why do we put things into museums?" nov 2013. [Online]. Available: http://www.cam.ac.uk/research/discussion/we-ask-the-experts-why-do-we-put-things-into-museums

[17] C. Marache-Francisco and E. Brangier, "The Gamification Experience: UXD with a Gamification Background," in *Gamification: Concepts, Methodologies, Tools, and Applications: Concepts, Methodologies, Tools, and Applications*. IGI Global, mar 2015.

[18] Google, "Google Design." [Online]. Available: https://design.google/resources/

[19] C. Marache-Francisco and E. Brangier, "The Gamification Experience: UXD with a Gamification Background," in *Gamification: Concepts, Methodologies, Tools, and Applications: Concepts, Methodologies, Tools, and Applications*. IGI Global, mar 2015.

[20] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences (2nd Edition)*. Routledge, 1988. [Online]. Available: https://www.amazon.com/Statistical-Power-Analysis-Behavioral-Sciences/dp/0805802835?SubscriptionId=0JYN1NVW651KCA56C102&tag=techkie-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=0805802835

[21] N. Lercari, M. Forte, and L. Onsurez, "Multimodal reconstruction of landscape in serious games for heritage: An insight on the creation of Fort Ross Virtual Warehouse serious game," in *2013 Digital Heritage International Congress (DigitalHeritage)*. IEEE, oct 2013. [Online]. Available: https://doi.org/10.1109%2Fdigitalheritage.2013.6744759

[22] P. Milgram and H. Colquhoun, "A Taxonomy of Real and Virtual World Display Integration," in *Mixed Reality*. Springer Berlin Heidelberg, 1999, pp. 5–30. [Online]. Available: https://doi.org/10.1007%2F978-3-642-87512-0_1

[23] ——, "A Taxonomy of Real and Virtual World Display Integration," in *Mixed Reality*. Springer Berlin Heidelberg, 1999, pp. 5–30. [Online]. Available: https://doi.org/10.1007%2F978-3-642-87512-0_1

[24] P. Milgram and F. Kishino, "A Taxonomy of Mixed Reality Visual Displays," *IEICE Trans. Information Systems*, vol. vol. E77-D, no. 12, pp. 1321–1329, dec 1994.

[25] E. Woods, M. Billinghurst, J. Looser, G. Aldridge, D. Brown, B. Garrie, and C. Nelles, "Augmenting the science centre and museum experience," in *Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Austalasia and Southe East Asia - GRAPHITE '04*. ACM Press, 2004. [Online]. Available: https://doi.org/10.1145%2F988834.988873

[26] R. T. Azuma, "A Survey of Augmented Reality," *Presence: Teleoperators and Virtual Environments*, vol. 6, no. 4, pp. 355–385, aug 1997. [Online]. Available: https://doi.org/10.1162%2Fpres.1997.6.4.355

[27] B. Katherine, *Emerging Research and Trends in Interactivity and the Human-Computer Interface*. IGI Global, oct 2013.

[28] ——, *Emerging Research and Trends in Interactivity and the Human-Computer Interface*. IGI Global, oct 2013.

[29] *Statistical Power Analysis for the Behavioral Sciences*. Routledge, jul 1988. [Online]. Available: https://doi.org/10.4324%2F9780203771587

[30] J. Earthy, B. S. Jones, and N. Bevan, "ISO Standards for User-Centered Design and the Specification of Usability," in *Usability in Government Systems*. Elsevier, 2012, pp. 267–283. [Online]. Available: https://doi.org/10.1016%2Fb978-0-12-391063-9.00049-3