Chen, W. et al. (Eds.) (2016). Proceedings of the 24th International Conference on Computers in Education. India: Asia-Pacific Society for Computers in Education

Usability Study of an Augmented Reality Game for Philippine History

Ma. Mercedes RODRIGO*, Eric Cesar E. VIDAL Jr.†, Nicko CALUYA‡, Jenilyn AGAPITO° & Walfrido David DIY¤
Ateneo Laboratory for the Learning Sciences, Ateneo de Manila University, Philippines
* mrodrigo@ateneo.edu, † ericvids@gmail.com, ‡ ncaluya@ateneo.edu,
° jen.agapito@gmail.com, ¤ wdiy@ateneo.edu

Abstract: We investigate the extent to which Augmented Reality can contribute to effectiveness and enjoyment in the context of history-game-based learning. This paper presents a series of usability tests of a mobile Augmented Reality game application for learning Philippine history, Igpaw: Intramuros, with representatives of the game's target audience. These tests discovered issues that compromised player enjoyment, including fatigue during gameplay, hurdles in learning AR-specific game mechanics, game repetitiveness, and outdoor safety concerns, in addition to minor technical issues. Nevertheless, the tests show the game's high effectivity in teaching Philippine history.

Keywords: Augmented Reality, Philippine History, Educational Game Design, Usability

1. Introduction

Many students equate the learning of history with rote memorization of names, dates, and places to reach a common account of past events, but they fail to appreciate patterns of change, instead amassing disjoint, unrelated facts (Squire & Barab, 2004). Students therefore find history boring, and may pass a semester-long course without an appreciation of the historical narrative (e.g. Fitzhugh, 2004; McMichael, 2007). To mitigate this, educators have integrated computer games among their teaching materials, arguing that games can be excellent carriers of educational content that can deepen students' understanding while keeping them engaged (Schell, 2014). The extents to which these theories and assumptions translate to outcomes, and the factors that contribute to or hinder games' success as educational tools, are the subject of ongoing investigation (Schmitz, Klemke, and Specht, 2012; Arias, 2014; Backlund & Hendrix, 2013; Wouters et al 2013), but for history education, Mortara et al. (2014) believe games can raise cultural awareness by depicting folklore, rules of behavior, customs, traditions, and spiritual beliefs, and can create immersive experiences where students can examine artifacts or archeological locations as they were when they were originally constructed.

History-related games tested in classrooms have been shown to have a positive effect on learning (Backlund & Hendrix, 2013). However, the effect of games on engagement and motivation is less definite. Squire & Barab (2004) found that user engagement is difficult to measure as it is influenced by the individual's goals, the game's affordances, and the general learning environment among others. Along a similar vein, Pujol et al. (2013) said that the affordances of a game played in a historical location such as a museum may impede player's experiences, as they required players to stay in the same spot for a long time. It is necessary, therefore, for history games to undergo user experience testing (Keil, et al, 2013). As it is, much of the literature we reviewed regarding the use of games in history education were generous with technical details but seldom included user feedback (e.g. Sedano, et al., 2013; Ardito & Lanzilotti, 2008; Furió, et al., 2013; Keil, et al, 2011). Thus, the purpose of this paper is to describe one such effort to evaluate the user experience. We describe Igpaw: Intramuros, a mobile augmented reality game for Philippine history, and detail our play testing methods and results.

2. Introduction to Igpaw: Intramuros and Augmented Reality

Augmented Reality (AR) refers to the superimposition of digital information (such as virtual characters) on top of real settings in 3D space, providing the user with situated knowledge (Azuma &

Bishop, 1997). AR seems to lend itself quite naturally to historical themes: artifacts can be brought to life, ruins can be reconstructed (Choudary et al., 2009), and players can live out history within actual contexts (Ferdinand, et al., 2005; Herbst, et al., 2008). Igpaw: Intramuros builds on these successes in the context of a developing country, the Philippines. While a previous paper expounds on game design choices and constraints (Rodrigo et al, 2015), this paper chiefly discusses a series of usability / play tests and debriefings with two main goals: to identify game design problems and issues, and to determine the extent to which the users enjoyed the experience.

Igpaw is set in Intramuros, Manila, Philippines. Intramuros was the Philippines' seat of government, church, and culture throughout the Spanish and American periods. Igpaw utilizes a story narrative wherein a malevolent force has displaced well-known Philippine historical personages (such as Jose Rizal, the national hero, Gen. Douglas MacArthur, and others) from their supposed time periods, and it is the player's job to unravel this mystery. To do so, the player needs to visit one of several real historical markers scattered around Intramuros. Focusing the camera on one of these markers will trigger a virtual scene inhabited by a few of these personages, and the player can talk to them to learn about their predicament, as well as use virtual historical items that can be picked up within the scene. As the player solves each personage's personal predicament, s/he also learns about the history of the said personage as well as that of the historical location where the player is standing.

Figure 1. Screenshots showing (a) a historical marker, (b) an AR virtual guide, and (c) in-game map

The word "igpaw" is a Filipino word meaning "to jump, to go beyond"; the game uses this word to exhort players to "go beyond the physical walls of Intramuros", i.e., discover its hidden history. The game is divided into three modules, each composed of three scenes, with a virtual guide appearing in every scene for consistency. Upon opening the game for the very first time, a simple tutorial is presented to orient the user with the game's orient-and-select user interface—the user points the camera at a virtual object so that buttons will appear and the user can select a button to perform an action. Every scene interweaves a variety of historical narratives from different time periods in order to minimize player travel time; a strict chronological narrative would force the player to backtrack between scenes multiple times. Also, a scene may even include fantasy elements, so long as these elements are related in some manner to the characters within that location and their fantastic nature is made clear to the player.

3. Usability Testing Methodology

The usability test's goals were to identify issues in the game design and to determine the extent to which users enjoyed playing the game. To these ends, we performed three iterations of user tests.

Members of the testing team include a facilitator, several observers, and several testers. The facilitator was responsible for welcoming the testers, observer-to-tester assignment and debriefing. The observers monitored the testers while they played the game to completion. The testers were high school or college students who took or are currently taking Philippine History classes; a total of 8 testers, 3 female and 5 male (3 high school and 5 college students, aged 16 to 22) were present in this study. The testers were each provided with a fully-charged mobile phone or tablet for the duration of the test.

Prior to field testing, the facilitator and observers familiarized themselves with the plot, objectives, and design of Igpaw, as well as preparing testing materials consisting of (a) Observation Sheet (see Table 1), (b) Debriefing Sheet, and (c) the actual mobile devices, installed with a fresh copy of the then-current game build. Each observer would record on the observation sheet the device name and model, battery percentages at the start and end of scenes, and task observation results for each test.

The tasks identified were based on specific in-game actions the players had to carry out to successfully complete the game. For every task, the observer had to take note of whether the tester

successfully accomplished the task ('Task Success') and the time it took in minutes/seconds ('Task Time'). The 'Comments' column listed observations and verbal comments by the testers while playing the game. The observation sheet was updated to reflect any game changes between each testing session.

Table 1: Sample of an Observation Sheet

	Tutorial			
Battery Life at Start of Scene: %	Tasks	Task Success	Task Time	Comments
1)	Calibrate Device			
2)	Talk to Daligmata (virtual guide)			
3)	Examine Daligmata			
4)	Locate & Pick Up Magic Wand			
5)	Give Magic Wand to Daligmata			
6)	Return to Map			
Battery Life at End of Scene:___%				

The debriefing sheet, given after testers completed the game, was a two-part written questionnaire that asked about their game experience. The first part consisted of the System Usability Scale (SUS), a standard tool for measuring users' perceived system usability as well as learnability (Brooke, 1996). Various researchers have shown the tool to be reliable and valid even with a sample size of 8 – 12 participants (Tulis & Stetson, 2004) and that it can be used for a vast range of systems and technologies (Bangor, et al., 2008; Sauro, 2011). The SUS is composed of 10 questions with 5 response options, from Strongly Disagree to Strongly Agree, averaged to a single number representing overall system usability. Any number above a score of 68 is considered above average. The second part of the debriefing sheet consisted of questions taken from Desurvire & Wiberg (2009) that elicit feedback regarding game play, immersion, and usability and game mechanics, using a similar 5-point scale.

Observers also had to gather as much qualitative information as they could through tester observation and note-taking. The observers were directed to attentively watch the device's battery life, whether the GPS tracking was working or not, and the accuracy of locations as depicted in the map (previously shown in Figure 1c). Furthermore, they had to monitor whether testers' in-game actions correspond with the intended sequences of tasks. If this is not the case, observers ought to ascertain the cause. Apart from verbatim comments, facial expressions and bodily responses were also noted.

During the first two sessions, participants played the game in the same order, i.e. Module 1, 2, then 3. For the last session, the game was played in three different sequences in order to test whether the in-game map screen is reliable enough to guide players to their destinations regardless of module order.

4. Testing and Debriefing Results

We divide the major test findings into two sections: findings related to the game's usability, and affective findings. Technical problems that emerged during the testing were mostly limited to minor bugs related to buggy animation and interface issues, although it is worthwhile to note that despite the average play time of 19 minutes per module and 98 minutes for the entire game (including travel time), the game consumed more battery life than the developers had anticipated due to AR consuming more power, guiding the decision to cut down the length of the game from the original four modules to three.

4.1 Usability Debriefing

We summarize the responses from the debriefing questionnaire as a SUS score computation of the users' responses across all usability sessions (see Table 2). The average SUS across all sessions was 71.56. These numbers are better interpreted as percentiles; thus, the 71.56 average falls roughly at the 69th percentile. Per SUS evaluation, Igpaw is 69% usable (grade B), implying that the game was generally usable and learnable, but can be improved. Additional questions taken from Desurvire & Wiberg (2009) revealed players' thoughts about game play, immersion, and game mechanics (see Table 3). The first five items under game play measures the player's enjoyment. While most of the items were average or above average (borderline enjoyable), players gave item 1 an average score of 2.83, that is, players found the game to be borderline repetitive or boring, which may be attributed to repetitive activities (e.g., talking to all characters, picking up all items). Also, the high standard deviation in item 5 (1.47) implies that some (though not all) players found the game physically tiring.

Table 2: SUS Scores

	Session 1		Session 2		Session 3	
	Total Score	Converted Score	Total Score	Converted Score	Total Score	Converted Score
Tester 1	30	75	26	65	28	70
Tester 2	29	72.5	24	60	31	77.5
Tester 3	-	-	29	72.5	30	75

Table 3: Game usability heuristics

	Game play	Ave.	Stdev
1	The game was fun, no repetitive or boring tasks.	2.83	0.75
2	The game provided me clear goals for every task.	3.83	0.98
3	The game world provides the player with a sense of control.	3.17	0.41
4	I felt safe while playing the game.	3.00	0.00
5	I felt some fatigue while playing the game.	3.17	1.47
	Immersion	Ave.	Stdev
1	The game story is informative and encourages immersion.	3.83	0.75
2	The game is educational and historically accurate.	4.17	0.41

	Usability and Game Mechanics	Ave.	Stdev
1	The controls are unconventional and hard to use.	2.00	0.63
2	I need to read a manual before playing the game.	2.17	1.60
3	The tutorial was necessary for me.	3.50	1.38
4	The tutorial gave me enough information.	4.00	1.10
5	The game provides me with enough help as to not get stuck in any level.	3.50	0.84
6	The screen layout is visually pleasing, the user interface is consistent, and the art is recognizable.	3.67	0.82
7	The objects in the game are recognizable and speak to their function.	3.83	0.41

In terms of immersion, the participants strongly agree that the conversations/dialogs are informative and educational, and that the story would most likely capture the engagement of a history enthusiast. In terms of usability, however, the high standard deviations of items 2, 3, and 4 meant that there was high divergence across users' experiences; indeed, while some participants needed help to play the game to completion, others had no problem finishing the game by themselves. The next section on affective feedback substantiates the findings from this written questionnaire.

Players also reported physical fatigue and physical discomfort. To play the game, the players needed to hold the device perpendicular to the ground, but for prolonged periods of time, this strained their arms. Weather also affected the player's level of discomfort. Sessions 1 and 2 had sunny weather while the Session 3's weather was cloudy and breezy. Players during the first two sessions felt uncomfortable staying in the heat and humidity for a prolonged period. During the third session, the cloudy weather distorted the GPS tracking of some of the devices and led to some confusion.

4.2 Affective Feedback

During game play, the observers noted that the testers experienced a wide range of affective states. In order of discussion, these states include delight and surprise, confusion and frustration, boredom, and competition between engaged concentration and anxiety.

Players were delighted and/or surprised as the virtual guides welcomed them to the adventure. The novelty of seeing 3-D characters on the camera feed led to smiling/laughing, gasping, and verbal reactions (e.g. "Wow!"). However, these reactions were markedly less frequent as the game progressed.

Confusion and frustration for some participants were caused in part by the AR game's orient-and-select approach. These players tended to tap a character or item to make options appear (which may cause objects to disappear from view due to the camera orientation) instead of using the device's camera to focus on the object. During the second session, one player asked the other player what to do and how to proceed. Some players were also confused when focusing on some game objects, such as a cannon and a headless priest. The camera had to be aimed at the rear end of the canon and on the "head" area of the priest, while the players intuitively focused at the center of the cannon's frame and the body of the priest. Nonetheless, they were able to learn where to focus and proceeded with the game. (Subsequent tests had the bounding box for the cannon and the headless priest enlarged and moved closer to their centers.) One player did not realize that she can actually look around with her device to see other virtual objects, prompting the developers to revise the tutorial instructions for accuracy. Furthermore, when several markers were within a module's location, Session 1 players were confused as to which marker was the correct one, and one tester tried to scan the marker from afar. To address this, an in-game picture of the correct marker for each scene was added for the second session. Some of the markers were above eye level, and players standing right in front of the marker obstructed the view of other players or non-playing tourists. The system was redesigned to adapt to the current real

position of the player, so that players can scan the marker from any angle and don't have to give way to other people. Finally, during session 1, one of the selected markers turned out to be in a restricted site, with security personnel approaching the players. This was replaced with another marker for session 2.

As for boredom, the players found the in-game dialog to be too long, resulting in a tendency to skip most of the dialog, with some players missing important details for game completion. Subsequent tests used color-enhanced dialog to highlight these details, facilitating this kind of player behavior.

Most importantly, the players' engaged concentration elicited safety concerns. Engrossed and focused players had a lack of regard for their surroundings. Holding a mobile device up for prolonged periods in a public place potentially exposed the players to crimes of opportunity (although no such incidents actually occurred). One of the sessions was conducted during a national holiday, when the foot traffic in Intramuros was heavy. Crowds of tourists were taking photos of the historical sites, vendors were selling their goods, and street children were begging for money. Furthermore, when walking from one location to another, the players tended to focus on the in-game map instead of their surroundings. The facilitators had to call the players' attention before they fell off the sidewalk or stepped into oncoming traffic. Finally, players got anxious when other people stared at them, either curious about what the players were doing or suspecting that the players were photographing them. During the tests, the observers watched out for the players' safety, but actual players might have no such backup. To address this, the tutorial of the game was redesigned to remind players regarding safety. The tutorial also encouraged players to go out in groups rather than solo.

5. Conclusions

History education lends itself quite naturally to augmented reality games. By necessitating the exploration of actual historical locations, history-based games draw players into an immersive experience that is engaging and educational. The success of these types of games, though, rests largely on the quality of the overall user experience. Following Mortara et al.'s (2014) educational objectives, the design of Igpaw: Intramuros is meant to raise players' cultural and historical awareness and immerse players in local folklore and literature while being at actual historical locations. Our usability testing with 8 players across three different sessions found that the game is highly effective in its goal of teaching Philippine history, confirming the strength of our game design.

However, this design is not without caveats. While there were technical problems that were trivial to correct, issues regarding how enjoyable the game was were more substantial. Players complained of fatigue from holding up the device. Environmental factors such as heat, humidity, traffic, and overall safety influenced players' physical and emotional comfort levels. Game mechanics had to be learned, after which, game play tended to be repetitive. Developers and educators hoping to create more games of this genre should attempt to address these issues. Players may, for example, feel more comfortable and secure in enclosed, air-conditioned, contained environments rather than outdoors. Game storylines and dialogs need to be more compelling and varied. Opportunities for rest to mitigate fatigue should also be built in.

This paper contributes to the literature in that it highlights the need for user experience testing. We designed this game with several constraints in mind (Rodrigo et al., 2015), as other developers have. In this paper, we discussed the impact of our design decisions on our players—something not often found in similar work. No other paper discussed safety concern factors (resulting from the choice of outdoor game locations) affecting the user experience. These issues notwithstanding, the development of these types of games has the potential to contribute substantially to history education, AR game design, and development of related technologies. These games can make history come to life and draw learners in, helping them come to a deeper appreciation of the past and its connection with life today.

Igpaw: Intramuros is available as a free download on Google Play Store and Apple App Store.

Acknowledgements

We thank all the members and sponsors of our development team, as listed in Igpaw: Intramuros's credits page. This work was made possible through a grant from the Philippines Commission on Higher Education entitled "Design, Development and Deployment of An Augmented Reality Game for Philippine History."

References

Ardito, C., & Lanzilotti, R. (2008). Isn't this archaeological site exciting!: a mobile system enhancing school trips. In Proceedings of the working conference on Advanced visual interfaces (pp. 488-489). ACM.

Arias, M. (2014). Using Video Games in Education. Journal of Mason Graduate Research, 1(2), 49-69.

Azuma, R., and Bishop, G. (1994). "Improving Static and Dynamic Registration in an Optical See-through HMD." In Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques, 197–204. ACM.

Backlund, P., & Hendrix, M. (2013). Educational games-are they worth the effort? A literature survey of the effectiveness of serious games. InGames and virtual worlds for serious applications (VS-GAMES), 2013 5th international conference on (pp. 1-8). IEEE.

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the System Usability Scale. International Journal of Human-Computer Interaction, 24(6), 574-594.

Brooke, J. (1996). SUS - A Quick and Dirty Usability Scale. Usability Evaluation in Industry, 189(194), pp. 531-540.

Choudary, O., Charvillat, V., Grigoras, R., & Gurdjos, P. (2009). MARCH: mobile augmented reality for cultural heritage. In Proceedings of the 17th ACM international conference on Multimedia (pp. 1023-1024). ACM.

Desurvire, H. & Wiberg, C. (2009). Game Usability Heuristics (PLAY) for Evaluating and Designing Better Games: The Next Iteration. In Proceedings of the 3rd International Conference on Online Communities and Social Computing: Held as Part of HCI International 2009 (OCSC '09), A. Ant Ozok and Panayiotis Zaphiris (Eds.). Springer-Verlag, Berlin, Heidelberg, 557-566.

Ferdinand, P., Müller, S., Ritschel, T., & Wechselberger, U. (2005). The Eduventure-a new approach of digital game based learning combining virtual and mobile augmented reality games episodes. In Pre-Conference Workshop "Game based Learning" of DeLFI 2005 and GMW 2005 Conference, Rostock (Vol. 13).

Fitzhugh, W. (2004). History is Fun. In Concord Review Society Newsletter and Forum (Vol. 2, pp. 10-11).

Furió, D., González-Gancedo, S., Juan, M. C., Seguí, I., & Costa, M. (2013). The effects of the size and weight of a mobile device on an educational game.Computers & Education, 64, 24-41.

Herbst, I., Braun, A. K., McCall, R., & Broll, W. (2008). TimeWarp: interactive time travel with a mobile mixed reality game. In Proceedings of the 10th international conference on Human computer interaction with mobile devices and services (pp. 235-244). ACM.

Keil, J., Pujol, L., Roussou, M., Engelke, T., Schmitt, M., Bockholt, U., & Eleftheratou, S. (2013). A digital look at physical museum exhibits: Designing personalized stories with handheld Augmented Reality in museums. In Digital Heritage International Congress (DigitalHeritage), 2013(Vol. 2, pp. 685-688). IEEE.

Keil, J., Zollner, M., Becker, M., Wientapper, F., Engelke, T., & Wuest, H. (2011). The House of Olbrich—An augmented reality tour through architectural history. In Mixed and Augmented Reality-Arts, Media, and Humanities (ISMAR-AMH), 2011 IEEE International Symposium On (pp. 15-18). IEEE.

McMichael, A. (2007). PC Games and the Teaching of History. The History Teacher, 203-218.

Mortara, M., Catalano, C. E., Bellotti, F., Fiucci, G., Houry-Panchetti, M., & Petridis, P. (2014). Learning cultural heritage by serious games. Journal of Cultural Heritage, 15(3), 318-325.

Normand, J.M., Servières, M. & Moreau, G. (2012). A new typology of augmented reality applications. In Proceedings of the 3rd Augmented Human International Conference (AH '12). ACM, New York, NY, USA, Article 18, 8 pages. DOI=10.1145/2160125.2160143

Pujol, L., Katifori, A., Vayanou, M., Roussou, M., Karvounis, M., Kyriakidi, M., & Ioannidis, Y. E. (2013). From personalization to adaptivity-Creating immersive visits through interactive digital storytelling at the Acropolis Museum. In Intelligent Environments (Workshops) (pp. 541-554).

Rodrigo, M. M. T., Caluya, N. R., Diy, W. D. A. & Vidal, E. C. E. (2015). Igpaw: Intramuros: Design of an augmented reality game for Philippine history. In Proceedings of the 23rd International Conference for Computers in Education. Hangzhou, China.

Sauro, J. (2011). A practical guide to the System Usability Scale: Background, benchmarks, & best practices. Denver, CO: Measuring Usability LLC.

Sedano, C. I., Leendertz, V., Vinni, M., Sutinen, E., & Ellis, S. (2013). Hypercontextualized Learning Games Fantasy, Motivation, and Engagement in Reality. Simulation & Gaming, 44(6), 821-845.

Schmitz, B., Klemke, R., & Specht, M. (2012). Mobile gaming patterns and their impact on learning outcomes: A literature review. In 21st Century Learning for 21st Century Skills (pp. 419-424). Springer Berlin Heidelberg.

Schell, J. (2014). The Art of Game Design: A book of lenses. CRC Press.

Squire, K., & Barab, S. (2004). Replaying history: engaging urban underserved students in learning world history through computer simulation games. In Proceedings of the 6th International Conference on Learning Sciences (pp. 505-512).

Wouters, P., Van Nimwegen, C., Van Oostendorp, H., & Van Der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. Journal of Educational Psychology, 105(2), 249.