

A Design Framework for Experiential History Games

Nicholas Lytle and Mark Floryan

University of Virginia, Department of Computer Science,
85 Engineers Way, Charlottesville, VA 22903
{nal3gc,mfloryan}@virginia.edu

Abstract. This article presents empirical studies of a serious game focusing on various aspects of the American Civil War. We developed and deployed four distinct modules of our game for use within a fifth grade classroom in Virginia, USA. Of the first three modules deployed, only one lead to statistically significant results from pre-test to post-test. We used qualitative information from these first three trials to develop a design framework for experiential serious games of this form. We then developed and tested a fourth module by applying this framework and found significant learning improvements with this fourth module. This paper presents our game, results of empirical studies within a fifth grade classroom, and our proposed design framework identifying key aspects of the learning environment. Our results provide support for our hypothesis that application of this framework leads to increased learning gains. While we do not suggest that our framework is complete or exhaustive, we believe that designers of similar educational games can benefit by employing the principles of this framework directly.

Keywords: Serious Games, Experiential Learning, History Education, Elementary Education

1 Introduction

American students lack of knowledge in United States History is well documented. On the *National Assessment of Educational Progress* in 2014, 82% of U.S. 8th grade students were not considered to be proficient in U.S. History knowledge.¹ In addition, there are well documented gender and racial gaps in history knowledge in elementary education, and the gap between public school students and those in private schools is significant. Comparatively, the number of 8th grade students that reported using computers for history/social studies education in school has increased significantly from 2010 to 2014 (18% to 25%), a promising indication that technology may be able to contribute to the solution to this problem.

While there have been many developments in educational and serious games that focus on STEM knowledge or topics, there have been relatively few developments for history-based games in the United States, especially American

¹ http://www.nationsreportcard.gov/hgc_2014/files/2014_history_appendix.pdf

history. This is countered by the fact that U.S. History is the subject in which the smallest percentage of 8th and 12th grades students were tested as being proficient ².

Agnor Hurt Elementary School, the school where the tests reported in this study were conducted, has a lower passing rate in history tests overall (70%) than both the corresponding district (85%) and state (84%). The school has also scored more poorly on the *fifth Grade Virginia studies test* (scoring 72% vs 77% at Division and 85% at state).³

2 Related Work

While there exist relatively few serious games explicitly focusing on history in the scientific literature, many games are designed to be experiential in nature (including, but certainly not limited to [9][10][11]). These games allow the player to learn through direct experience within a simulated environment. Additionally, many frameworks exist for the design of educational games, but few relate specifically to experiential games. As one example, Akkerman et al. [9] describe a framework for developing a history game, *Frequentie 1550*, employing storification, but this framework is not entirely general in its utility to game designers. While some frameworks are very specific (and clearly useful when applied under the necessary constraints) many are quite general [4][2][5]. For example, Hunicke [3] describes the MDA framework for game design, but does not explicitly include educational purposes and does not provide specific meaningful heuristics for measuring a games application of the framework. Other frameworks [8][2][5], while very useful as general strategies for design, don't recommend explicit design principles, rendering their application somewhat difficult. Dondlinger [12] presents an overview of common strategies employed by successful educational games (goals, narrative, etc.), implying a useful concrete framework. In general, it seems most frameworks are either broadly constructed (useful in their ability to help us with general strategies for educational game design) or fitted to a particular problem (concrete and useful but only within a constrained problem space). One goal of this work is to attempt to provide a meaningful step towards bridging this gap, in which a fairly broad category of game (experiential games) might have a relatively concrete framework applied.

3 Our Game - A Nation Divided

A Nation Divided is a web-based game developed in Javascript using a rendering engine called *PIXI.JS*. The game is played from a top-down perspective as the player takes control of a nameless protagonist (see figure 1, left) navigating the game world. The player interacts with other non-playable characters and entities

² <http://www.nationsreportcard.gov/>

³ <https://p1pe.doe.virginia.gov/reportcard/report.do?division=2&schoolName=1797>

and in many cases, the player is able to choose what response to give to these characters from a list of options.

The game currently contains four distinct modules (more detail below) focusing on distinct aspects of the American Civil War. In each module, the player is placed in a specific location and must complete a variety of tasks. This includes activities such as reading newspapers, communicating with townspeople, and engaging historical figures such as Harriet Tubman, Abraham Lincoln, etc. The quests for each module can be completed in any order, but all must be completed to progress to the ending segment. For many quests, the player must supply a non-playable character with the correct response to a question (or series of questions) to make progress. This helps ensure that the player understands a piece of information within the domain before continuing. If the player is unable to supply the correct answer to that character, there are characters within the vicinity that give hints or help.



Fig. 1. While playing, students walk around town interacting with historical figures and locals (left). Every module contains a unique end game scenario quizzing the students knowledge (right).

After all of the tasks are completed, the final section of a given module is available to the students. These ending scenarios are unique to each module and have their own interaction scheme and win condition. For example, Module 3 (Mid-war) features an end-game segment that tasks players to control a Union blockade and shoot down ships that have Confederate identifiers (figure 1, right). These ending segments usually tie back to the theme or plot that was developed in the previous section.

The game employs a cartoony aesthetic with colorful graphics and light, calming music. While there are certainly aspects of the time period that could have been explored or shown in a more detailed, mature light, the game was specifically designed for an elementary classroom setting and the design was catered around this intended young user. The dialogue, the primary vehicle for delivering information in the game, was not developed to be completely in-line

with the speaking mannerisms of individuals in the 1800s. Instead, the primary goal was to make comprehensible and engaging dialogue that an average 5th grader could read and understand.

As stated earlier, the game contains four distinct modules, details of which follow:

1. **Underground Railroad:** Taking place on a southern cotton plantation in the 1850s, the player must aid Harriet Tubman in collecting supplies, information, and travelers to take on the Underground Railroad. Once ready, the player must complete a turn-based board game moving a group of runaway slaves from city to city navigating away from slavers trying to catch the party. The game is completed when either the slaves are caught (resulting in a game over) or the slaves are able to escape out of the country.
2. **Secession:** Set in an unspecified Virginian plantation in the year 1860, the player is tasked with delivering newspapers containing the news that Abraham Lincoln has been elected president. The player must present the correct newspaper to the non-playable characters (either a northern abolitionist newspaper or a southern newspaper) based off of the information and views that they relate to the player. On completion, the game shows each southern state seceding from the Union in order coupled with information about the beginning of the Civil War.
3. **Mid-war:** Set in 1862 on an unspecified Virginia plantation, the player must help General McClellan with a series of tasks for the Union. This includes relaying messages to Union naval officers, helping wounded soldiers remember which battles they took place in, and aiding Clara Barton and her nursing staff. Once completed, the player takes control of a Union blockade shooting down ships that have concepts associated with the Confederacy (e.g. Virginia, General Lee, Stars and Bars, etc.) and allowing ships that have concepts associated with the Union to pass.
4. **Gettysburg:** Set in Gettysburg, Pennsylvania during the Consecration of the National Cemetery at Gettysburg (1863), the player is tasked by Abraham Lincoln to collect information about what happened during the battle and how people are reacting. The player must piece a timeline of events from General Meade, ask for Frederick Douglass's opinion on the battle and war, and help reporters get the facts of the battle. Afterwards, the player aides Lincoln in constructing the Gettysburg Address by select the correct missing word in a segment of the address given its synonym.

Although similar to one another, each module is unique in exactly how material is presented and game mechanics are designed. The following section attempts to taxonomize the framework of our game.

4 Our Design Framework

Chronologically, we developed a framework for designing modules within our system after the deployment of modules 1-3 (Underground Railroad, Secession,

Mid-war) in the classroom, and before the design and development of module 4 (Gettysburg). The reasoning for this is two fold. Firstly, we observed large variation in the efficacy of the first three modules in the classroom (see methodology and results below) and wished to develop a theoretical understanding of how our various designs affected learning in the classroom. Additionally, we hoped to design the final module (Gettysburg) by applying our framework directly as a more robust test to validate the framework. Thus, one can observe that this framework is somewhat retrospective with regards to our first three modules (designed via qualitative and quantitative results of those interventions), but was applied and validated directly through the final Gettysburg module.

We identified nine design variables and posit that these affect the quality of an educational game, especially those similar in structure to our game.

1. **Spatial Clustering:** Spatial Clustering involves the degree to which related entities in the game are visually (spatially) located near one another. For example, a module that places a Union general (e.g., General Grant) near union soldiers is advantageous. This qualitative measure implies that designers can take advantage of the visual nature of the medium by spatially categorizing related elements for the player, assisting the player in forming a taxonomy of the information. This can also be done with certain groups in highlighting roles and viewpoints. Having all of the slave characters in the cotton fields while the white plantation masters and workers were outside of the fields (Module 1) serves to highlight differences between groups and form relationships between groups and roles.
2. **Temporal Clustering:** High temporal clustering occurs when all of the events chronologically fit into a small time-period. The higher range in time leads to low temporal clustering. For example, a module covering the entire American Civil War will have lower temporal clustering than a module focusing on one year within the war.
3. **Clearly Defined Goals:** Well documented in educational gaming[1][12][6], a module with clearly defined goals makes clear to the student what task is at hand and what must be achieved. Additionally, these goals should be concrete and actionable. Quests in the game that are easily identifiable to the player (e.g., talking to 10 clearly visible wounded soldiers) are easier for a student to accomplish than quests that are not (e.g., finding a canteen hidden on the map but occluded).
4. **Thematically Defined Ending:** This principle states that ending sequences (whether cinematic or interactive) should be closely tied to the activities the player was performing in the main mission to avoid a feeling of separation or exclusion. For example, Module 1's ending in which the player takes control of a runaway party of slaves on the Underground Railroad was a direct extension of the narrative developed from the previous segment in which the player had to collect supplies to get ready for the journey.
5. **Interactive Ending:** This principle states that modules should have a separate game as the ending sequence rather than a passive activity. We observed that modules with this feature tended to be better at engaging the students,

encouraging them to continually play after completion. For example, many students continually played the ending game for Module 3 (the Union blockade game) as it was the most developed and most interactive game of the 4 modules. Conversely, no students went through the slideshow of information about the beginnings of the Civil War found in Module 2 more than once.

6. **Difficulty of Gaming the System:** Gaming the system [13][14] occurs when students attempt to maneuver through an educational (or any) system in an unintended way (usually by leveraging the feedback given). An activity scoring well on this dimension might be a task forcing students to put ten events in history in proper order. Because this task has many combinations (too many to reasonably guess each), it is difficult for a student to game the system (i.e., try every combo until the player can move on without considering the content and/or purpose of the activity). This is best exemplified in our system in Module 4 (Gettysburg) in the segment where the player must aide General Meade in piecing together a timeline of events. Because each required input has 5 different options, most students elected to not attempt a guess and check approach to the quest.
7. **First-Hand Interaction:** This principle involves the degree to which a student can directly interact with any artifacts of interest. For example, a module that discusses a famous historical figure would have less first-hand interaction than a module that allows the player to directly speak with that figure. In examining questions related to identifying people, it was found that the questions that had the subject present in the game performed better. In general, we posit that increased first-hand interaction is likely preferable.
8. **Breadth of Quests:** This refers to the number of unique quests any given module has. For example, Module 1 (The Underground Railroad) had 7 small quests while Module 2 (Secession) had 2 larger quests. While Module 1 was the most successful, we believed that there is most likely a hard limit on how many individual quests a module can contain before a player can become overwhelmed. This limit could also be more related to the number of actions in a module than the number of quests.
9. **Depth of Quests:** In addition to breadth of quests, depth of quests refers to the amount of actions necessary for the completion of a single quest. For example, a quest that requires a student to deliver 10 newspaper is said to be deeper than a quest that only requires the player to speak to 4 naval officers. We posited that there was a hard limit on how deep a quest could be before the task becomes overwhelming or frustrating to the player.

Table 1 below presents the degree to which each of our four modules address the nine principles in our design framework. The lead game designer was asked to identify each module as having low (red in table below), medium (yellow in table below), or high (green in table below) application of the principle in question. A rating of low approximately means that less than 25 percent of the module correctly applies the principle. Yellow was given if between 25 and 75 percent of the module content addresses the principle in question, while green indicates

that most (more than 75 percent) of the content is a direct application of the principle in question.

Table 1. A visual representation of how well each of our four modules applies the various aspects of our design framework. Green implies a strong application, yellow a medium application, and red a weak application of that particular principle.

	<i>Spatial Clustering</i>	<i>Temporal Clustering</i>	<i>Clear Defined Goal</i>	<i>Thematically Defined Ending</i>	<i>Interactive Ending</i>	<i>Hard to Game</i>	<i>First-hand Interaction</i>	<i>Breadth of Quests</i>	<i>Depth of Quests</i>
<i>Underground Railroad</i>	Red	Green	Yellow	Green	Red	Yellow	Yellow	Red	Green
<i>Secession</i>	Red	Green	Yellow	Yellow	Red	Red	Yellow	Yellow	Green
<i>Mid-war</i>	Yellow	Red	Red	Yellow	Green	Red	Green	Yellow	Yellow
<i>Gettysburg</i>	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow

5 Methodology

For each module, students from a small elementary school in Virginia were taken out of the classroom for an hour long period of time. Each student was a 5th grader who was learning the material discussed in the game that year in their classroom. Each student was given a school laptop computer to access the game as well as pencils and pens to answer the pre and post test questions. Some students participated in multiple test sessions while others did not. The number of students that participated in each test session varied from module to module. This information is provided in the Results section under table 2.

The test sessions were divided into three main sections: the pre-test, the game playing session, and the post test. Between each of these sections the students were given instructions on what he or she would be doing by the researchers and the teacher. During each section, the teachers and researchers would provide clarification to any student asking questions or provide help to the student when asked if it did not directly influence their ability to complete the task (e.g. getting paper). Each student was given a unique identifying number that was held constant between all three sections. The teacher was able to know which student was tied to each number, but the researchers did not.

During the pre and post-test sessions, each student was given a test that included a number of multiple choice questions related to the material discussed in the game. These questions were all modeled after Virginia Standard of Learning (SOL) style questions for the 5th grade Virginia Studies SOL and some were direct questions from previous SOL exams. The students were able to answer the questions in any order they chose and were given 15 minutes to complete the assessment. In some sessions, some questions appeared on both the pre and post test for that session and some questions appeared on multiple test sessions.

Between the pre and post test sessions, the students had up to 30 minutes to play the game. Students were allowed to ask the teachers and researchers clarifying questions but were not given information on exactly where to go or how to

complete a task. Any technical difficulties that happened during this period were handled by the researchers. The students were not strictly discouraged from collaborating with other students playing the game unless a student was explicitly telling other students how to complete certain tasks. If a student completed the entirety of the game, he or she was encouraged to continue exploring within the game until the time limit for gameplay was up.

6 Results

Table 2 below summarizes the pre test to post test score changes over all four modules. Three modules led to improvements in scores (modules 1, 3, and 4). Modules 1 and 4 were the only two that show statistically significant results ($p < 0.05$). Both of these modules also contain medium to high effect sizes.

Table 2. Overall results comparing pre test to post test scores across all four test sessions (one test per module)

Module	n	x_1	σ_1^2	x_2	σ_2^2	p	Cohen's D
Underground Railroad	17	0.614	0.176	0.739	0.147	0.012*	0.766
Secession	20	0.611	0.224	0.572	0.208	0.384	0.180
Mid-war	15	0.550	0.210	0.633	0.208	0.191	0.398
Gettysburg	18	0.346	0.225	0.438	0.234	0.008**	0.398

We also performed an analysis of learning gains across various question types, in order to obtain some sense of student learning across categories. Table 3 below summarizes these results. Modules 1 and 4 performed well on questions regarding *people* (important players during various times of the war). Additionally, module 1 displayed improvements on *analysis* questions. Two modules (3 and 4) show significant improvements in questions regarding *events*. There were no significant improvements on questions regarding viewpoints or facts & terms.

Table 3. P-Values of T-Tests from pre to post test broken down by question type. Red highlights indicate that the mean decreased. Bold indicates statistically significant changes.

Module	n	People	Geography	Viewpoints	Facts	Events	Analysis
U. Railroad	17	0.007**	N/A	0.999	0.168	N/A	0.013*
Secession	20	0.090	0.517	0.209	N/A	0.825	N/A
Mid-war	15	0.178	N/A	0.334	N/A	0.006**	N/A
Gettysburg	18	0.003**	0.668	0.104	0.999	0.049*	0.496

7 Conclusions and Future Work

We were pleased to find that overall, the Underground Railroad module and the Gettysburg module displayed the highest learning gains, as they tended to fit our model more closely than the other two modules. We thus believe that our data provides some support for the hypothesis that our framework is a useful tool for designing educational games of this form. We are, however, hesitant to claim that our framework provides sufficient or necessary conditions for effective educational game design. Our intention with this work was to make progress towards useful design principles for game designers constructing experiential games geared towards younger students (elementary in our case). To this effect, we feel that although our framework is not authoritative or complete, it provides useful insight and progress towards a robust set of design principles for educational games. To this end, we note that our most promising result from this work is the increase in learning gains observed with module 1 (Underground Railroad) and module 4 (Gettysburg), the two modules which most closely apply the principles from our framework. Some meaningful insight might be gleaned by re-examining our framework and the degree to which these two modules fit various features (see figure 2 above). In particular, features that were closely applied in our successful modules and not in unsuccessful modules may prove to be more important for design. These features include spatial clustering, providing a clearly defined goal, and providing a thematically defined ending. Thus, these design considerations may very well be the subjects of immediate further inquiry as they have shown promise as a potential indicator of success in the classroom.

Likewise, framework features that are ignored by our two unsuccessful modules might indicate a lowest-common denominator for ineffective design. Features fitting this category include providing a thematically defined ending, having it be difficult to game the system [13], and providing a clearly defined goal. These features, in fact, could prove to be necessary features of any successful design, a conclusion supported by related literature [12][7][6].

Additionally, our analysis breakdown by question might provide some indication of design elements that help reinforce various types of knowledge. For example, identifying important people within the war was most successful in modules with high spatial clustering, a result that may indicate that clustering helps students categorize the major players more accurately. Questions regarding events showed most improvement in modules 3 and 4, potentially indicating that features unique to those modules (first hand interaction, interactive endings) improve learning in this area. Both of these claims require further investigation, but may generalize to other related situations.

In conclusion, the purpose of this work was to investigate the use of educational games for history education. We developed an initial framework based on results of three modules used within a 5th grade classroom, and applied this framework to a fourth module for validity. Our results show that the best pre to post-test learning gains occurred in modules that most closely applied our design framework. While we don't claim that this fully validates our framework, it does provide evidence that the ideas therein are worthy of further exploration

as we attempt to make progress towards a robust set of design principles for educational game design. We thus hope that educational game designers can find utility in thinking about the aspects of our design framework and how these features might affect the efficacy of a resulting game implementation.

Acknowledgements

We would like to thank Andy Barron, David Amin, Zane Laughlin, Adrian Lee Gloria, and the rest of the development team⁴. We would also like to thank Mia Shand, along with all of the faculty, staff, and students of Agnor Hurt Elementary School for their participation in the study.

References

1. Charsky, Dennis. "From edutainment to serious games: A change in the use of game characteristics." *Games and culture* (2010).
2. Westera, Wim, et al. "Serious games for higher education: a framework for reducing design complexity." *Journal of Computer Assisted Learning* 24.5 (2008): 420-432.
3. Hunicke, Robin, Marc LeBlanc, and Robert Zubek. "MDA: A formal approach to game design and game research." *Proceedings of the AAAI Workshop on Challenges in Game AI*. Vol. 4. 2004.
4. Salen, Katie, and Eric Zimmerman. *Rules of play: Game design fundamentals*. MIT press, 2004.
5. Burgos, Daniel, Colin Tattersall, and Rob Koper. "Re-purposing existing generic games and simulations for e-learning." *Computers in Human Behavior* 23.6 (2007): 2656-2667.
6. Prensky, Marc. "Computer games and learning: Digital game-based learning." *Handbook of computer game studies* 18 (2005): 97-122.
7. Gee, James Paul. *Good video games+ good learning: Collected essays on video games, learning, and literacy*. New York: P. Lang, 2007.
8. Kelly, Henry, et al. "How to build serious games." *Communications of the ACM* 50.7 (2007): 44-49.
9. Akkerman, Sanne, Wilfried Admiraal, and Jantina Huizenga. "Storification in History education: A mobile game in and about medieval Amsterdam." *Computers and Education* 52.2 (2009): 449-459.
10. Floryan, Mark, Woolf, Beverly. "Rashi Game: Towards an Effective Educational 3D Gaming Experience". *Proceedings of the IEEE International Conference on Advanced Learning Technologies*. Athens, GA (2011).
11. Froschauer, Josef, et al. "Design and evaluation of a Serious Game for immersive cultural training." *Virtual Systems and Multimedia (VSMM), 2010 16th International Conference on*. IEEE, 2010.
12. Dondlinger, Mary Jo. "Educational video game design: A review of the literature." *Journal of applied educational technology* 4.1 (2007): 21-31.
13. Baker, Ryan, et al. "Why students engage in gaming the system behavior in interactive learning environments." *Journal of Interactive Learning Research* 19.2 (2008): 185-224.
14. Cocea, Mihaela, Arnon HersHKovitz, and Ryan SJD Baker. "The impact of off-task and gaming behaviors on learning: immediate or aggregate?." (2009): 507-514.

⁴ <http://people.virginia.edu/~nal3gc/civilwarcredits>