Routledge
Taylor & Francis Group

Journal of the Learning Sciences

ISSN: 1050-8406 (Print) 1532-7809 (Online) Journal homepage: http://www.tandfonline.com/loi/hlns20

Substantial Integration of Typical Educational Games Into Extended Curricula

Douglas B. Clark, Emily Tanner-Smith, Andrew Hostetler, Aryah Fradkin & Vadim Polikov

To cite this article: Douglas B. Clark, Emily Tanner-Smith, Andrew Hostetler, Aryah Fradkin & Vadim Polikov (2017): Substantial Integration of Typical Educational Games Into Extended Curricula, Journal of the Learning Sciences, DOI: 10.1080/10508406.2017.1333431

To link to this article: http://dx.doi.org/10.1080/10508406.2017.1333431

Accepted author version posted online: 23 May 2017.

Substantial Integration of Typical Educational Games Into Extended Curricula

Douglas B. Clark

Department of Teaching and Learning

Vanderbilt University

Emily Tanner-Smith

Peabody Research Institute

Vanderbilt University

Andrew Hostetler

Department of Teaching and Learning

Vanderbilt University.

Aryah Fradkin

Armistead Gardens Elementary/Middle School

Vadim Polikov

Agora Ventures

Correspondence should be addressed to Douglas Clark, Box 230, 230 Appleton Place, Nashville, TN 37203-5721. E-mail: doug.clark@vanderbilt.edu

Abstract.

Much research focuses on what might be possible with digital games in the classroom. This study focuses on what is currently probable and typical. It uses a controlled quasi-experimental design to compare outcomes for students of 13 teachers in 10 diverse urban, suburban, and rural schools. The teachers

integrated a set of 55 typical educational games into their curricula on Jacksonian democracy. Teachers reported strong engagement benefits for the game condition and a strong interest in using games of this type in the future in their surveys and interviews. Each teacher taught at least one classroom with the games and at least one classroom without the games. When the one teacher who reported a failed implementation was dropped from the analysis, the results showed significantly higher gains for the game condition in terms of multiple-choice, open-response factual outcomes, evidentiary depth, and student engagement outcomes. When the failed implementation was included, the game condition demonstrated improved scores for all outcomes, but only student engagement and evidentiary depth remained significant. Moderator analyses highlighted the role of teacher experience and student engagement in the efficacy of the game condition, indicating that game instruction was particularly beneficial for Special Education students.

Introduction

Research raises three distinctions that are addressed in the current study. First, there is an important distinction between studies that examine what might be possible with specialized or advanced game designs and curricula versus what is probable in terms of typical current educational games and curricula. This study focuses on the latter in terms of what is currently probable for typical educational games and curricula.

Second, a challenge in any research on games for learning involves representing a subgenre of games broadly enough to be generalizable beyond the design of single game or developer while including a sample of games homogenous enough to arguably represent something more specific than the omnibus term "digital game." A digital game by definition can

span from multi-million dollar complex titles to "games" that involve only rudimentary gamification of a typical school worksheet. This study focuses on a subgenre that we define as "typical educational games" to constrain and define a subset of games specific enough to address our focus on probable/typical and begin to address our distinction of specificity in terms of a subgenre of games that might be envisioned for classrooms.

Third, recent meta-analyses by Wouters et al. (2013) and Clark, Tanner-Smith, and Killingsworth (2016) suggest that extended interaction with digital games for learning is central to their efficacy. This study thus focuses on curricula designed and implemented by each teacher with durations of approximately three weeks.

Beyond these three specific distinctions, studies are needed that compare multiple extended implementations of specific subgenres of games in a manner that allows comparison of the outcomes across implementations and contexts. The current study attempts to support these goals of generalizability by comparing outcomes for students across the classes of 13 social studies teachers in 10 diverse urban, suburban, and rural schools across 6 states and the District of Columbia. The teachers integrated a set of 55 typical educational games into their own curricula on Jacksonian democracy.

Digital Games and Learning

Digital games are influential and ubiquitous presences in the lives of young learners. A 2008 study by the Pew Internet and American Life Project found that 97% of teens aged 12-17 played computer, web, portable, or console games, and 50% of them reported daily or nearly

daily play (Lenhart et al., 2008). Beyond youth, a 2015 Pew study on adults and video games reported that 77% of men and 57% of women ages 18-29 (and 50% of men and 48% of women overall) reported playing video games (Duggan, 2015). A 2012 Pew study reported that digital games generated $25 billion in sales in 2010 (Anderson & Rainie, 2012).

With this massive interest in digital games, aspects of recreational games have spread across education, commerce, training, and social media in terms of educational games, serious games, and gamification. Scholars refer to this as the "rise of a gameful world" (Walz & Deterding, 2015, p. 7). This rate of growth is only anticipated to grow. A 2012 Pew survey of 1021 technology stakeholders and critics found that 53% agreed with the statement: "By 2020, there will have been significant advances in the adoption and use of gamification. It will be making waves on the communications scene and will have been implemented in many new ways for education, health, work, and other aspects of human connection and it will play a role in the everyday activities of many of the people who are actively using communications networks in their daily lives" (Anderson & Rainie, 2012, p. 3).

Contributing to this expanding presence of games, educational researchers are increasingly interested in the affordances of digital games as a medium for learning. Investigation into the use of games for learning has grown from a small niche area to a major focus of research over the past 15 years (e.g., Dieterle, 2010; Gee, 2007; Tobias & Fletcher, 2011). Studies have demonstrated the potential of digital games to support learning in terms of conceptual understanding (e.g., Barab et al., 2009; Klopfer, Scheintaub, Huang, Wendel, & Roque, 2009), process skills and practices (e.g., Kafai, Quintero, & Feldon, 2010; Steinkuehler

& Duncan, 2008), epistemological understanding (e.g., Squire & Jan, 2007; Squire & Klopfer, 2007), and players' attitudes, identity, and engagement (e.g., Barab et al., 2007; Dieterle, 2009; Ketelhut, 2007). Reports by the National Research Council and others (e.g., NRC, 2011; Martinez-Garza, Clark, & Nelson, 2013; Young et al., 2012) have acknowledged this potential, but also acknowledge the unevenness of systematic evidence for games as learning tools. Although several recent meta-analyses have demonstrated that games can successfully support learning, these reviews have underscored the wide range of design choices and game genres that make comparisons and generalizations difficult (Clark et al., 2015; Sitzman, 2011; Vogel et al., 2006; Wouters, van Nimwegen, van Oostendorp, and van der Spek, 2013).

Best Practices and Technology Integration in History Education

Researchers and educators across the fields of History Education and Social Studies Education have identified key elements of high quality and rigorous history learning. Chief among these elements are the sourcing, contextualizing, and inquiring that, through careful reading and analysis of primary sources and documents, constitute the work of a historian (NRC, 2005; VanSledright, 2011; Wineburg, 2001). Teachers then, ought to be using primary source materials to engage students in using evidence from these texts to support historical arguments or narratives that answer questions historians might ask (NCSS, 2013). Typically, history is learned through the presentation of facts in the form of dates, events, and people via a textbook or lecture positioning the nature of history, at best, as an exciting story (NRC, 2005). Doing the work of a

historian supports students' learning history more deeply and more rigorously than the typical story presentation in two ways. First, rather than receiving information from a teacher or textbook, students are positioned as active learners engaged in interpretive and cognitively demanding work while reading and aligning evidence with claims to answer questions. Second, students would be working with history much like a historian does, countering inadequate views of history as a stagnant collection of facts and repositioning history as a process of making meaning of past phenomena by engaging in practices.

In terms of technology, while reform efforts highlight an important role for technology in social studies education (NCSS, 2013; NRC, 2005), social studies education has generally been slow to integrate technology into classroom practices (e.g., Swan & Hoffer, 2008; Whitworth & Bearson, 2003). Most of the research and reform efforts have focused on technology in terms of identifying, accessing, critiquing, and communicating around source materials (e.g., Friedman & Heafner, 2007; Hicks, Lee, Berson, Bolick, & Diem, 2014; Saye & Brush, 2002). Interestingly, while *Oregon Trail* and *Civilization* are possibly the two most famous examples discussed in terms of digital games and education, relatively little research has been conducted on games in social studies compared to games in science, math, and reading. As Metzger and Paxton (2016) outline, building on the core research on *Civilization* (e.g., Lee & Probert, 2010; Pagnotti & Russell, 2012; Squire, 2006; Voorhees, 2009), the majority of the research that has been conducted on games and social studies education has focused largely on "simulation" games and the analysis of mass-market entertainment titles that can be applied to classroom contexts (e.g., Kee & Graham, 2014; McCall, 2014; Wright-Maley, 2015). Research has not focused, however,

on the types of simpler educational games that are more typical of what students encounter in school contexts.

Rationale and Research Questions for the Current Study

The research highlighted above and elsewhere (e.g., Gee, 2007; Squire, 2011) raises three key distinctions that are addressed in the current study: (a) distinguishing between what might be possible with specialized or advanced game designs and curricula versus what is probable in terms of typical educational games and curricula; (b) representing a genre broadly enough to be generalizable beyond the design of single game or a single developer while including a sample of games homogenous enough to represent something more specific than the omnibus term "digital game;" (c) taking into account the importance of extended interaction with the games when studying games for learning (Clark et al., 2015; Wouters et al., 2013), and (d) taking into account how the game and non-game curricula as implemented by the teachers reflect best practices for history learning as outlined in research, reform, and standards documents (NCSS, 2013; NRC, 2005; Vansledright, 2011).

In terms of the first distinction of possible versus probable/typical, the current study focuses on what is probable/typical. Focusing on what is possible is also crucial, and the work of the lead author of this study has predominantly focused most on exploring what might be possible with games in schools by designing and exploring advanced designs of games and curricular structures (e.g., Clark, Nelson, Chang, Martinez-Garza, Slack, & D'Angelo, 2011;

Clark, Sengupta, Brady, Martinez-Garza, & Killingsworth, 2015; Clark, Virk, Sengupta et al., 2016). That said, it is valuable to explore what is currently probable/typical as a foundation for understanding the current realities of schools and educational games, knowledge that can then inform research into what is possible.

We define "educational games" as games intended to support learning of academic content. The games could be referred to as "serious games" in the sense "that the objective of the computer game is not to entertain the player, which would be an added value, but to use the entertaining quality for training, education, health, public policy, and strategic communication objectives" (Wouters et al., 2013, p. 2). By "typical," we refer to the genre of educational games that are (a) representative of the norm for educational games currently found in schools, (b) relatively inexpensive to create, (c) relatively short to play (3-30 minutes), (c) specifically designed to support the learning of specific academic content, (d) inclusive of a range of features and mechanics from basic recreational games, and (e) structurally advanced beyond the minimalist threshold of bare gamification that involves only a thin overlay of points or badges (Anderson & Rainie, 2012; Walz & Deterding, 2015). By focusing on this definition of typical educational games, we constrain and delineate a subset of games specific enough to address our focus on probable/typical and begin to address our distinction of specificity in terms of a subgenre of games that might be envisioned for classrooms.

To address the challenge of generalizability beyond the design of a single game or a single developer, the current study recruited 16 different game studios and independent game developers to create games for the study. In an effort to support comparisons and analyses, all

games were developed around one of five subtopics about Jacksonian democracy. To standardize the development budgets for each game, the studios and developers were paid a moderate flat fee to build the games and offered cash prizes if they were among the ten highest rated games by students and teachers. The developers were free to create as many or as few games as they wished within a four-month time period. Ultimately, the studios and developers created and submitted 55 games for the project. Thus, the current study addresses the second distinction in terms of representing the defined subgenre of typical educational games beyond the design of a single game or a single developer.

To address the third distinction of extended curricula versus brief interventions, all teachers in the current study were asked to teach the curriculum for durations of at least three weeks of class time. Each teacher taught at least one classroom that did not use the games and at least one classroom that did use the games. Teachers were asked to use the games at least 50% of the time in their games classrooms, but other aspects of the curricula for the games and non-games classrooms were left to the teacher so that the results of the study could be generalized in terms of typical classroom practices.

To address the fourth distinction of the degree to which game and non-game curricula implemented by typical teachers might reflect best practices for history learning, we interviewed the participating teachers about their implementations as outlined above.

Building on the framing outlined above, the goals of this study are to explore and support generalizable claims about the efficacy of typical educational games integrated across extended curricula. More specifically, the current study explores the following research questions:

1. What are the characteristics of the cross-section of typical educational games created by developers for the study?

2. How effectively do the teachers' curricula with the games support student learning compared to their curricula without the games? Are there differences in learning outcomes in terms of factual knowledge as measured by multiple-choice versus open-response questions? Are there differences in the evidentiary depth of students' open-response answers?

3. How do teachers integrate the games into their curricula? To what degree does this integration reflect best practices for history learning as outlined in research, reform, and standards documents?

4. How engaging are the games for students, how favorably do the teachers view using such games in retrospect, and what types of games do the students and teachers favor?

Method

This study employed a quasi-experimental comparison group design to evaluate the efficacy of using digital history games for teaching historical concepts to middle- and high-school students. This non-equivalent posttest-only comparison group design (Shadish, Cook, & Campbell, 2002) was used to contrast the means of posttest outcome measures for the game and non-game conditions. Classrooms were the unit of assignment to conditions, thus this design can also be referred to as a cluster-quasi-experimental design. Baseline differences between intervention and comparison groups are one of the primary threats to interval validity in a quasi-

experimental design (i.e., due to potential selection bias in assignment to conditions). Therefore, and as described in greater detail below, we matched students in the game and non-game conditions on prior history grades and ELL, IEP, and Special Education status to ensure baseline equivalence between the groups. This quantitative design was supplemented with analysis of qualitative implementation data from teacher interviews.

Sample

The original sample included 1,080 students enrolled in 46 classrooms taught by 15 teachers located in 10 public, private, or charter schools in the United States (2 in Alabama, 2 in the District of Columbia, 1 in Kentucky, 2 in Maryland, 1 in Missouri, 1 in New Jersey, 1 in Rhode Island). The schools are located in urban (54%), suburban (23%), and rural (23%) contexts. Teachers were instructed to teach at least one of their history classes using the digital games, and at least one other class using (non-game) practice-as-usual teaching techniques. Teachers did not randomly allocate classrooms to conditions. Instead, teachers were encouraged to balance between the game and non-game classes to the degree possible when selecting which classrooms received the game conditions.

Given the potential selection bias associated with non-random allocation of classrooms to conditions, the final analytic sample included a smaller subset of classrooms and students that were matched and established baseline equivalence on key student characteristics. One teacher with 11 students in the game condition and 15 students in the non-game condition was dropped from the analytic sample because they were located in a private school, which was deemed potentially incomparable to the other public and charter schools in the sample. One teacher with

25 students in the game condition was dropped because that teacher provided the game instruction to her own classroom, but allowed a different teacher to provide the non-game instruction to the 27 students in the comparison classroom. The data were dropped in this case because the experimental and comparison conditions were not taught by the same teacher.

For the remaining 1,002 students, we balanced students in the game and non-game conditions on measures of prior achievement and special status. All balancing was done within teachers. First, students with Special Education, 504/disability, IEP, or ESL/ELL designations were matched to students in the non-game condition with the same designation who were taught by the same teacher and had similar prior grades in the history class. Any students with ELL, Special Education, or IEP status who could not be matched were dropped from the analytic sample. Next, within each teacher, we examined the baseline equivalence between game and non-game conditions on students' prior grades in the history class. If students in a teacher's game and non-game classes were non-equivalent on prior grades (defined as independent t values > 1.96 or Cohen's d effect sizes > 0.25), we trimmed the high/low performers from the analytic sample until the groups were equivalent. This yielded a final analytic sample of 848 students (440 in game condition; 408 in non-game condition) who were well balanced and equivalent on prior academic achievement and special status.

Curricular Topic

While the current study wished to support generalizability of claims, a disciplinary focus was needed. Social studies was chosen for the richness of the content and the flexibility it presented for the design of a range of different types of games. The topic for the unit chosen was

Jacksonian Democracy. Jacksonian Democracy is central to the middle school social studies standards across most states. In the Tennessee state standards, for example: "8.55 Explain the events and impact of the presidency of Andrew Jackson, including the "corrupt bargain," the advent of Jacksonian Democracy, his use of the spoils system and the veto, his battle with the Bank of the United States, the Nullification Crisis and the Indian removal. (C, E, G, H, P, TN)" (Tennessee Dept of Education, 2016). The topic was envisioned in terms of five microunits: (a) Andrew Jackson's personal history; (b) Jackson Populism, the Kitchen Cabinet, Spoils Systems, and Political Parties; (c) the Nullification Crisis; (d) the battle with the Bank of the United States; and (e) the Indian Removal Act and the Trail of Tears.

Material Provided to Game Developers and Teachers

Developers were provided with a sample textbook chapter, regular teaching materials students might typically encounter (e.g., PowerPoints and handouts), and sample tests on Jacksonian democracy, but they were not provided with the actual posttest to ensure that the games would not be over-aligned with the outcome assessment. Developers were encouraged to draw upon any other source materials and reference materials. Developers were informed that the Common Core standards endorse (and teachers commonly use) primary source documents like political cartoons and newspaper extracts to promote student learning.

Teachers were provided with all of the materials that were provided to the game developers to support planning, but teachers were encouraged to incorporate whatever materials and activities they wished into their curricula. Teachers were also provided with the final unit test (unlike the developers) so that they could plan instruction around the unit test. We provide

qualitative analyses of the nature of non-game curricula and the game curricula in the results section and compare these implementations to the best practices proposed by reform and standards documents.

Historical Knowledge Outcomes

The unit posttest included four multiple-choice questions and seven short essay questions. Four social studies teachers not involved in the pilot were recruited to assist with coding. Student tests for all teachers and conditions were presented to coders for coding in a randomized order to minimize any coding drift. Tests were labeled only with anonymous ID numbers. Coders were blind to students' membership in the game or non-game conditions. The research team was not provided with the key connecting ID numbers to experimental condition until after all coding was completed. We examined four outcomes measuring students' history knowledge, which were collected as part of the teacher-administered end-of-unit posttest: (a) multiple-choice factual knowledge, (b) open-response factual knowledge, (c) open-response evidentiary depth, and (d) word count of open-response answers.

Multiple-choice factual knowledge. This outcome was measured as the number of correct responses from 0 to 4 on the four multiple-choice questions assessing students' knowledge about Andrew Jackson's Democratic Party, Spoils Systems and Kitchen Cabinet, and attitudes toward the Federal Bank.

Level of open-response factual knowledge. This outcome was measured on a scale of 0 to 10, indicating level of factual accuracy based on open-text responses assessing students' knowledge about Andrew Jackson's presidency on the seven multipart open-response questions.

Each short-response posttest question was deconstructed into parts based on what the prompts asked students. Appendix A presents the rubric used for Question 11 as an example. Table 1 presents examples of students' answers and their scores on the scale of 0-10 for Question 11. Explicit knowledge and information were identified based on what possible answers would be considered correct for the parts of each prompt in each question. Each short response question had a detailed and comprehensive rubric for each subsection or element of the answer. Each coder scored one or two open-response questions for all of the data (and the coder who scored only one open-response question also coded the multiple-choice answers for all students). Throughout the coding process, coders checked each other's work, provided advice, and added notes to the rubrics to detail what they counted for each rating and their reasoning for giving specific ratings. A random subsample of 100 items was double-coded to assess inter-rater reliability on the level of factual knowledge ratings. Inter-rater reliability was assessed using intra-class correlations (*ICC*s) to assess the consistency of coder ratings. The resulting *ICC*s were in the good/excellent range, average *ICC* = .80 (range .63 - .95) (Cicchetti, 1994), indicating that coders had a high degree of agreement and reliability.

Level of open-response evidentiary depth. This outcome was measured on a scale of 0 to 10, indicating level of contextualization based on open-text responses assessing students' evidentiary depth and contextualization about Andrew Jackson's presidency. Specifically, the identified attributes were broken into two categories, the thinking that was prompted explicitly in the question, and the thinking that was implicitly encouraged through the prompts. The explicitly prompted thinking included evidence, example, and explanation. The rubric for these explicit critical thinking attributes was duplicated for questions that asked for more than one piece of

evidence, example, or explanation with each rated separately. Appendix B presents the rubric used for all questions for evidentiary depth. Table 1 presents examples of students' answers and their evidentiary scores on the scale of 0-10 for Question 11. The implicit thinking attributes included contextualization and implications. This was rated for each prompt response in relation to the extent the student answered beyond what was asked of them in the prompt. A random subsample of 100 items was double-coded to assess inter-rater reliability on the level of contextualization ratings. The resulting *ICC*s were in the good/excellent range, average *ICC* = .81 (range .65 - .95), indicating that coders had a high degree of agreement.

Length of responses. This outcome was measured as the average number of words used in the seven open-response questions on the posttest. Table 1 also includes word counts for the example answers for Question 11. This was collected as a potential measure to capture increased student engagement in the curriculum.

Student Engagement Survey Outcomes

We also examined two outcomes measuring students' interest and confidence in U.S. history, which were collected as part of the teacher-administered and student-completed survey.

Student interest in unit. This outcome was measured as the average of two items on a 1-10 scale (Cronbach's α = .72), which asked students how much they liked learning about Andrew Jackson's presidency, and how interesting they found this unit compared to other U.S. history units covered prior in the school year.

Student confidence. This outcome was measured as the average of two items on a 1-10 scale (α = .67), which asked students how good they think they are at history class, and how much they like learning about U.S. history.

Potential Moderators of Intervention Effects

The study collected data on several school, teacher, and student characteristics that we examined as potential moderators of the effect of the game intervention on students' outcomes.

School characteristics. We examined five characteristics of the schools as potential moderators of the effect of the intervention. School proficiency rating (range 0-100%); the percentage of free and reduced price lunch students in the school (range 0-100%); percent of White students in the school (range 0-100%); overall school quality grade from SchoolGrades.org (range A-F); and overall school rating from GreatSchools.org (range 1-10).

Teacher characteristics. We examined five characteristics of the teachers and their implementation of the games as potential moderators of the effect of the intervention. Teacher experience was coded as the number of total years of teaching experience. Teachers' perspectives on the games' effectiveness were measured as the average of three items on a 1-5 scale [strongly disagree to strongly agree], indicating whether teachers believed games improved the effectiveness of instructional time, engaged students in more rigorous discussions, and improved student performance (α = .74). Teachers also reported whether they perceived games to increase focus among students who were typically off-task (range from 1 [strongly disagree] to 5 [strongly agree]). Finally, teachers' interest in using games to teach U.S. history was measured as the average of two items on a 1-5 scale [strongly disagree to strongly agree], indicating whether

teachers would recommend games to other teachers or would be interested in using games in the future (α = .63). Finally, we measured the total minutes of instruction reported by the teachers.

Student characteristics. We collected data on five additional characteristics of the students, which were examined as potential moderators of the effect of the intervention. Students' baseline performance in history class was measured via teacher-reported student grades (range 0-100%). Students' with special status were designated via teacher report in terms of Special Education, individualized education program (IEP), or English as a second language/English language learner (ELL). Students also self-reported whether they thought video games would make history class better (range from 1 [no] to 10 [yes]).

Quantitative Analytic Strategy

We used multilevel linear regression models to examine the overall effectiveness of the game condition for improving students' historical knowledge and engagement. All models were estimated separately for the six outcomes, and statistically controlled for students' baseline history grades, grade level, and ELL, Special Education, and IEP status. To account for clustering of students within classrooms and schools, all models were estimated using three-level models that included random intercepts for classrooms and schools. Further, to account for potential variability in the effectiveness of the game intervention across schools, all models included random coefficients for the intervention effect, thereby permitting the intervention effect to vary across schools.[1] These models can be written as:

[1]Fit statistics indicated that the models including random coefficients did not significantly improve model fit above the models including only random intercepts (with χ^2 (df = 2) ranging

$$y_{ijk} = \beta_1 + \beta_2 GAME_{ijk} + \beta_3 x_{ijk} + \cdots \beta_p x_{ijk} + \zeta_{jk}^{(2)} + \zeta_k^{(3)} GAME_{ij} + \varepsilon_{ijk}$$

where y_{ijk} is the outcome response for student *i* in classroom *j* in school *k*; β_1 is the average intercept, β_2 is the average slope for the intervention effect across schools; $\beta_3 \ldots \beta_p$ are the effects of the *p* control variables in the model (prior grade, special statuses, and grade level); $\zeta_{jk}^{(2)}$ is the random intercept for classrooms, which represents the deviation of classroom *j*'s intercept from the mean intercept β_1; $\zeta_k^{(3)}$ is the random slope for the intervention effect, which represents the deviation of school *k*'s slope from the mean slope β_2; and ε_{ijk} are the residuals.

In addition to the main effects models examining the overall effectiveness of the game condition, the study explores whether any of the school, teacher, or student characteristics outlined above moderated the effects of the intervention. These moderation effects were examined using multiplicative interaction terms, with all models estimated examining the effect of one moderator at a time. To facilitate interpretation of results, all intervention effects are translated into standardized mean difference effect sizes (Cohen's *d*), estimated as the difference in adjusted posttest means (i.e., β_2 from the multilevel model specified above) divided by the unadjusted pooled standard deviation.

Although there were no missing data on any of the covariate control or effect size moderators, there was a small amount of missing outcome data due to student non-response on

---

from 2.0 to 4.88). Nevertheless, we elected to report estimates from the models including both random intercepts and random coefficients, given that these models provided the most conservative test of the game condition by permitting intervention effects to vary across schools.

surveys. Missing outcome data were not imputed, but rather all analyses were conducted using listwise deletion.

Qualitative Analysis of Teacher Interviews

In addition to the quantitative analyses, we conducted qualitative analysis of interviews with teacher-participants after the posttests. The goal of these interviews involved exploring teachers' experiences implementing the games, overall perspectives on the effectiveness of the games, and details on how the teachers viewed students' use and success learning with the games. To analyze the qualitative responses from these interviews, a short list of codes was developed based on initial analyses of the interviews. The codes were selected because they offered an added dimension to the quantitative data analysis already conducted. We narratively describe the results from these interviews, highlighting teachers' implementation, teachers' reactions and perspectives, the nature of the games and depth of thinking supported, gender differences and academic performance status of the students, and teachers' views on student engagement.

Results

School, Teacher, and Student Characteristics

Table 2 provides descriptive statistics for the characteristics of the schools, teachers, and students in the analytic sample. The schools represented a range of urban (54%), suburban (23%), and rural (23%) contexts. Schools' proficiency ratings from SchoolGrades.org ranged

from .19 - .71 (*M* = .40, *SD* = 0.18), and schools' grades from GreaterSchools.org ranged from A (22%) to D (33%). The average percentage of students in schools receiving free/reduced price lunch was .56 (*SD* = 0.35), and the average percentage of White students in schools was .51 (*SD* = 0.35). Most of the teachers used the games in their 8th grade classrooms, although one teacher used the game in her 6th grade classrooms and one in his 10th grade class. Teachers' experience ranged from one to 25 years in the classroom. The teachers in the sample had an average of 12.08 (*SD* = 7.62) years of experience.

Teachers were asked to spend at least three weeks of class time on the game and non-game versions of their curricula. Teachers reported spending an average of 949 (*SD* = 146) minutes on game and non-game versions of their curricula (this equates to approximately 19 class periods of 50 minutes each, but we report in minutes because teachers had different block schedules and times). Table 3 presents further details from teachers about implementation quality and acceptability. Because the student samples were matched on prior history grades and special status, students in the game and non-game conditions were similar at baseline by design in terms of their prior history grades (overall *M* = 81.41, *SD* = 15.31) and special statuses (3% ELLs; 3% Special Education; 4% IEP).

Nature of the Games

The first aim of the study was to investigate the nature of the typical educational games that game studios and independent developers would tend to create. The following sections describe these five categories in the order of their increasing sophistication of game mechanics and pedagogical design: (a) basic ungated games, (b) gated progress games, (c) deeper content

games, (d) interactive stories, and (e) intrinsic design games. Student, teacher, and researcher ratings for each category are also discussed (ranging from 1 = lowest rating to 5 = highest rating).

Basic Ungated Games are reminiscent of the simplest recreational games that one might find in an app store with some academic content added in a manner such that mastery of that content does not affect progress in the game. In each of these games, students encounter historical information and/or source materials during game play, but mastery is not integrated into mechanics or progression. Many of these games also had multiple-choice assessments built in after the game or at the end of a game level, but these were also not connected to progression or game play. These basic ungated games were the least sophisticated category of games created by the developers. Overall, there were 22 games in this category, contributed by four developers, with an average estimated play time of 7.1 minutes. The largest subcategory of these games were based on "runner" or "jumper" mechanics (Figure 1 Top Right) where the player runs continuously from left to right on the screen, jumping over and sliding under obstacles along the way, or continually jumps up to platforms higher on the screen. The second subcategory focused on "memory" mechanics where the player needs to flip over two tiles at a time while trying to match the pictures on the back of each tile (Figure 1 Top Left). The third subcategory involved "slicing" (or "fruit ninja") mechanics wherein the player attempts to move her finger across the surface of the tablet so as to "slice" through tokens as they move around the screen. The final subcategory involved games that incorporated multiple basic mini-games including memory, slice, jigsaw puzzle, and picture comparison (Figure 1 Bottom).

One "runner" game won a teacher-student award as a result of the beautiful production values of the game and the content sprinkled within (e.g., paintings and other source materials with explanatory text). Most of the other the games in the runner, memory, and slice subcategories, however, were less remarkable in their execution and more simplistic in the nature of the academic content (Figure 1 Top). As shown in the graphs of Figure 2, students generally rated the games in the Basic Ungated Games category among the most highly, with the teachers less favorable, and the researchers the least favorable. We had assumed that the students would quickly tire of these basic mechanics because the mechanics are not as sophisticated as the mechanics of AAA recreational game titles, or even as sophisticated as many common recreational app games, but apparently students also find these basic mechanics very satisfying. Many of the teachers suggested changing these ungated mechanics into gated mechanics (discussed in the next category), remarking that many students often ignored the pedagogical content in the ungated games. It would be easy to make this change, and the underlying games are certainly simple to create.

The final compilation subcategory of basic ungated games incorporated source materials, such as paintings and political cartoons, for the backgrounds and components of the various minigames. These games were among the teachers' favorites and students also liked them, garnering four of the ten teacher-student awards. The researchers liked the subcategory better than the other three subcategories of basic ungated games, but the researchers did not see same pedagogical or recreational value for these games as the teachers and students (Figure 2). That said, the compilation games engaged students in looking at source materials as part of identifying elements in a scene, putting together jigsaw puzzles from pieces of a painting, and identifying

subtle differences between two versions of the same political cartoon. The players therefore likely learned more about the source materials and paintings than in the other basic ungated games.

Gated Progress Games involved game mechanics where progress in the game was gated with multiple-choice questions about Jacksonian content. There were 11 games in this category, contributed by four developers, with an average estimated play time of 13.8 minutes. One subcategory focused on quiz show games wherein the players were participants in a television show. Only two games were created in this subcategory, both by the same developer. The games entailed beautiful and humorous production values: Andrew Jackson as the host, politicians of the days as panelists, and the opportunity to win a variety of humorous prizes. A second subcategory focused on a simple role-playing mechanic wherein the player explored an island, fought battles to get treasure, and used the treasure to buy better equipment. The context of the game had nothing to do with Andrew Jackson, but the player needed to answer multiple-choice questions about Jacksonian content in order to effectively attack opponents in battles (Figure 3 Top). A third subcategory engaged the player in exploring a museum, wherein each exhibit presented text or source materials. The player answered multiple-choice questions to convince the docent to allow the player to move to the next room. The fourth subcategory involved a single "tower defense" game wherein the player could answer multiple-choice questions during battles to obtain additional resources (Figure 3 Bottom).

As shown in Figure 2, the ratings for the gated progress games were the most consistent of any category across teacher, student, and researcher ratings. In terms of subcategories, the teachers preferred the tower defense and museum games, the students preferred the tower

defense and quiz show games, and the researchers preferred the museum and quiz show games. Overall, the tower defense game won one of the ten teacher-student awards.

Deeper Content Games were organized around answering questions about content in a manner beyond decontextualized or isolated rote mechanics. The deeper content games generally involved greater contextualization and/or interpretation and connection in an interesting or immersive contextualized manner. There were 10 games in this category, contributed by four developers, with an average estimated play time of 7.9 minutes. Examples included collecting clues in a mini-game and then connecting the clues into an extended coherent descriptive narrative on a Jacksonian topic (Figure 4 Top), taking the role of Jackson in a debate and choosing his answers to most closely align with his political and social views (Figure 4 Bottom), or interpreting political cartoons of the era to answer questions about the era.

As shown in Figure 2, the teachers and researchers rated these deeper content games higher than the basic ungated games and the gated progress games, whereas the students rated them lower, particularly for the subcategory that engaged students in interpreting source materials. The subcategory that focused on collecting clues in a minigame and connecting the clues into an extended coherent descriptive narrative, however, was fairly popular with the students and very popular with the teachers. Two of these games won teacher-student awards.

Interactive Stories engaged players in clicking on locations to enter various scenes where the player could click on characters to read what they had to say and click on responses to engage in simple dialog with the characters (Figure 5). These interactive stories were arguably not games in the sense of having a "win" state. Instead they provided opportunities for the players to explore contexts appropriate to the Jacksonian era and hear different perspectives on

the circumstances and politics of the era. Sometimes these interactive stories were followed by multiple-choice questions. In a subset of the games, these multiple-choice questions were integrated into a narrative of creating a newspaper article that the player had been investigating in the interactive story. There were six games in this category, contributed by two developers, with an average estimated play time of 5.0 minutes. As shown in Figure 2, students rated the interactive stories lower than they rated any other category of games, whereas the researchers and teachers rated them higher. As a result of the low student ratings, none of these games won awards. We would guess that students gave low ratings because the interactive stories required substantial reading and because the interactive stories did not provide clear feedback regarding the player's relative degree of success in terms of winning or losing. As discussed later in terms of the teacher interviews, teachers reported similar observations from the classrooms, particularly in terms of lower performing students.

Intrinsic Design Games involved more complex game mechanics that directly explored the content through means beyond multiple-choice questions. There were six games in this category, contributed by four developers, with an average estimated play time of 19.2 minutes. One subcategory involved simple simulations where farmers needed to buy and manage land and supplies within the context of shifting political and economic factors of the era (Figure 6). The ostensible pedagogical goals involved helping players understand and identify with the perspectives of farmers and common people struggling to get by. The second subcategory involved two card games where the abilities of the cards were inspired by political organizations and maneuvers of the era. The third subcategory included a single game reminiscent of a simplified version of the classic *Oregon Trail* game where players bought supplies and made other choices as they tried to navigate the hardships of the Trail of Tears.

Overall, as shown in Figure 2, students rated the simulation and Oregon Trail games roughly equal to their ratings of the basic ungated and gated progress games, whereas the teachers and researchers gave them higher ratings. As a result, the Oregon Trail game and one of the simulations won two of the teacher-student awards. The ratings for the intrinsic design card games were much lower for the teachers, students, and researchers. In fact, students rated the card games the lowest of any subcategory of games, while teachers rated them at the bottom along with the basic ungated slice games and the gated role-playing games. The researchers rated card games higher than the runner, memory, and slice basic ungated games, but agreed that the execution of the card games fell short of their interesting and creative potential.

Correlations of Teacher, Student, and Researcher Ratings. Teachers and researchers tended to rate the games similarly (r = .38). The researchers' ratings were meant to emphasize pedagogical value as implemented (as opposed to pedagogical potential of the general style of mechanic), and teachers' ratings generally seem to follow this orientation. As seen in the graphs of Figure 2, the major divergence between the ratings of the researchers and teachers involved the basic ungated games, where the teachers tended to rate the games higher, particularly in terms of the compilation subcategory. Students, by contrast, diverged substantially. Whereas teachers and researchers rated the deeper content, interactive stories, and intrinsic design games higher than the basic and gated games, students tended to rate the basic and gated games the most highly (with teacher-student ratings correlated r = .33; researcher-student ratings correlated r = -.40). It is important to note that teachers often reported differences in game preferences between high and low performing students, and it is therefore possible that higher performing students rated games more similarly to the researchers.

Implementation Quality and Acceptability Based on Teacher Surveys

Table 3 summarizes the teacher survey results in terms of implementation quality and acceptability of using games to teach historical concepts to middle- and high-school students. Teachers were asked to use the games for at least 50% of the curricular time in their games classes. Of the total time used to teach this course unit for the games classes, teachers reported using the game for 50% of classroom time (range 25-80%). On average, teachers agreed that they found it easy to restructure their class material to include the games (M = 4.08; SD = 1.12).

Teachers were less confident that students could play the games and interpret the content with minimal support from the teacher (M = 3.62; SD = 1.26). Interviews with the teachers (which we will discuss later) suggested that this was largely around the card games and simulations in terms of students needing for teacher support. In terms of student engagement, the teachers rated students' engagement in the game classrooms (M = 4.31; SD = 0.63) higher than the students' engagement in the non-game classrooms (M = 3.23; SD = 0.73) to a significant degree (t = 4.07, p = .002). Essentially, nine teachers rated the game condition classes with higher engagement levels than they rated their non-game classes, four gave equal ratings, and none rated students' engagement in the game classes lower than they rated their non-game classes. Furthermore, the interview data suggests that this statistic underreports the degree of teachers' perceptions of increased engagement for the students in the game condition. Of the four teachers who provided equal ratings to their game and non-game classrooms, (Booker, Steve, Lawrence, and Ann), Booker explained in his extended interview that all of his classes are always very engaged and stayed that way, Steve said the games raised the engagement of his lower class (which he had assigned to the games condition) from a 2 to a 4, which made it equivalent to his higher level class (which he had assigned to the non-game comparison condition), Lawrence said the games engaged the students differently but not more, while Ann reported no difference in engagement (3 and 3 out of 5) as a function of general morale issues at the school that we will explore in greater detail later.

In terms of the future, 92% of the teachers strongly agreed or agreed with the survey question that they would like to use games like these in the future (54% strongly agree, 38% agree, 8% unsure, 0% disagree, 0% strongly disagree). In the follow-up interviews, teachers on

average expressed that they would want to use games for 32% of their instructional time in the future (ranging from 17% to 50%). Thus, overall, teachers completed their experience with the games highly favorable toward continuing to use similar games in their curricula if available in the future.

Nature of the Effects of Game Instruction on Student Outcomes

Main effects. Table 4 presents the overall effects of game instruction on students' historical knowledge and engagement outcomes. The results indicated that although students in the game condition had higher levels of factual knowledge on both the multiple-choice (d = 0.05, b = 0.05, 95% CI [-0.11, 0.21]) and open-response (d = 0.08, b = 0.12, 95% CI [-0.03, 0.27]) relative to students in the non-game condition, these differences were not statistically significant. Similarly, students in the game condition had longer open-response answers compared to students in the non-game condition (d = 0.15, b = 1.83, 95% CI [-0.43, 4.09]), but this difference was not statistically significant. Students in the game condition demonstrated significantly higher levels of evidentiary depth relative to students in the non-game condition (d = 0.13, b = 0.12, 95% CI [0.01, 0.22]). The results also indicated that students in the game condition reported significantly greater interest in this historical unit in the class, compared to students in the non-game condition (d = 0.28, b = 0.52, 95% CI [0.24, 0.80]). This finding of higher student engagement in the game condition based on student surveys was corroborated in teacher interviews as we will discuss later. Finally, the results indicated that students in the game condition expressed lower, but not significantly lower, levels of confidence in their history class relative to students in the non-game condition (d = -0.17, b = -0.34, 95% CI [-0.75, 0.07]).

Main effects excluding the teacher who reported a failed implementation. In the extended follow-up interviews, one teacher was an outlier in terms of her experiences. Whereas all of the other teachers expressed reactions and experiences that ranged from somewhat positive to very positive emotionally and analytically, this outlier teacher – whom we will call Ann – was noticeably emotionally upset in her interview after the completion of her unit on Jacksonian democracy. Ann had been very enthusiastic and excited about using the games with her students and had planned carefully, but reported rebellion from the students. Ann's students' results were the most negative in terms of learning outcomes and engagement.

Without Ann's results, the outcomes for the study would look quite different. With Ann's classes removed from the quantitative analyses, the positive gains of multiple-choice (d = 0.15, b = 0.14, 95% CI [0.02, 0.26]) and open-response factual (d = 0.12, b = 0.16, 95% CI [0.00, 0.32]) knowledge would have demonstrated statistical significance in addition to the significant gains for evidentiary depth (d = 0.15, b = 0.14, 95% CI [0.03, 0.24]) and student interest (d = 0.29, b = 0.54, 95% CI [0.26, 0.83]) (Table 4). We have included Ann in the reported the results for the study in the preceding section and in the following moderators section because outwardly her implementation fit within parameters, but we wish to explore what might have occurred in her classrooms that was so different. We explore her experience and the experiences of the rest the teachers through qualitative analyses of their post interviews in the following sections.

Moderators of effects. The results from the moderator analyses, including Ann as well as the other teachers, indicated that the effects of the game condition on student outcomes were remarkably consistent across the various student, teacher, and school characteristics that we examined. Indeed, the effects reported in Table 4 were consistent for students in the top and

bottom quartiles of the distribution relative to their comparison condition counterparts, and there was also no evidence that the intervention effects reported in Table 4 varied in schools with different racial/ethnic or socioeconomic compositions. Seven interactions were statistically significant, however, and we discuss those next in turn.

First, while the main effect model indicated that students in the game condition had higher multiple-choice scores than students in the non-game condition (as shown in Table 4), the moderator analyses indicated that this effect varied depending on teachers' perceptions of whether the games increased student focus and also teachers' interest in implementing the game in the future. Students in the game condition demonstrated significantly higher levels of multiple-choice factual knowledge relative to students in the non-game condition for teachers who strongly agreed on the survey that "students who would otherwise be off-task were more focused during game-play" (d = 0.23 when teachers strongly agreed; d = 0.04 when teachers agreed, d = -0.32 when teachers strongly disagreed). Similarly, students in the game condition demonstrated significantly higher levels of multiple-choice factual knowledge relative to students in the non-game condition when teachers expressed greater interest in implementing the games in the future (d = 0.34 when teachers had the most interest; d = -0.15 when teachers had the least interest). Students' multiple-choice factual knowledge is therefore higher for the games classes relative to the non-game classes for the teachers who reported better experiences with the games and more positive attitudes toward the games.

The results also indicated that the effect of the game condition on students' evidentiary depth varied depending on teachers' perceptions of whether the games increased student focus. Students in the game condition demonstrated significantly higher levels of evidentiary depth

when teachers strongly agreed that students who would otherwise be off-task were more focused during game-play (d = 0.22 when teachers strongly agreed; d = 0.15 when teachers agreed, d = -0.04 when teachers strongly disagreed). Thus, again, the game condition improved students' outcomes for teachers who reported better experiences with the games and more positive attitudes about the games.

The moderator analyses also indicated that the effect of the game condition on students' response length on the open-response questions varied for students based on their Special Education status, as well as the school's grade from SchoolGrades.org. Special Education students in the game condition had significantly longer response lengths than Special Education students in the non-game condition (d = 0.76), whereas non-Special Education students showed modest differences in response length between conditions (d = 0.14). Furthermore, students in the game condition tended to have longer response lengths than students in the non-game condition in the higher rated schools (d = 0.34 for A schools; d = 0.20 for B schools; d = 0.14 for C schools; d = -0.07 for D schools).

Finally, the moderator analyses indicated that the effect of the game condition on student confidence in history varied according to students' Special Education status as well as teacher experience. Special Education students in the game condition had significantly higher confidence in history than Special Education students in the non-game condition (d = 1.22); whereas non-Special Education students in the game condition showed lower confidence in history compared to those in the non-game condition (d = -0.18). Furthermore, students in the game condition expressed significantly higher confidence in history relative to their comparison condition counterparts when taught by more experienced teachers (d = 0.10 for teachers with 25 years of

experience), but significantly lower confidence relative to their comparison condition counterparts when taught by less experienced teachers (d = -0.40 for teachers with 1 year of experience).

Correlations between ratings of game genres and student outcomes. In addition to moderator analyses, we also calculated correlations by teacher between average test outcomes and average student ratings of the five game genres. While it would have been particularly interesting to analyze individual student outcomes in terms of individual student game ratings or individual game play progressions, we can look at student game ratings averaged only by teacher as a result of the protocols to anonymize student identities for the current study. Correlation analyses show significant positive correlations between higher average test outcomes and higher average student ratings for the "deeper content" category of games (Table 5). The correlations are significant for open response factual knowledge and evidentiary depth, marginally significant for open response length, but non-significant for multiple choice. We did not see similar strong relationships by teacher between average learning outcomes and average student ratings for any of the other game categories. We similarly did not see relationships between teachers' ratings of the game categories and their average student outcomes.

Implementation of Games Based on Post-Interviews

Two strategies for implementation were typical for all of the teachers except Ann. The most common strategy involved teaching a lesson the first half of the class period and then having students play the games the second half of the class period. "Most days, I just talked for the first half of class. Then games for the second half of class" (Abe). A few teachers followed a

second strategy exemplified by Frank, a teacher from an urban middle school, who used "…the material of the games to help them form their knowledge of the content enough to do the assignments" (Frank). Frank had students take notes on game play and use the notes to write raps, poems, and skits to perform. Related approaches included using the games to reinforce or enhance disciplinary skills. "I would introduce a concept; trail of tears, a short PowerPoint, we did a couple of primary sources. I do that for the other class but in a much more intense level. So we would have to read a little bit more about it and then usually we would use the games for the second half kind of to reinforce and enhance…" (Willis).

Regardless of approach, teachers in these first two strategies split the class period time evenly over the span of each instructional day. The one exception was Ann, the outlier teacher discussed earlier. Ann described clustering game play across a couple of class periods a week.

My original plan was to have them playing the game for two of the four days. On the first day it would be they would read something about whatever the topic was and then on Thursday we would take an assessment, but then we found that actually giving the game group…52 minutes to play a game, it actually did not work very well. I had to make it a little bit more structured. (Ann)

As a result, Ann ultimately adjusted to play games a little each day with more game play on Wednesdays. By comparison, the implementation strategies of the rest of the teachers avoided the challenges of full game days with evenly divided instruction and game time on most days.

Another problem negotiated by the teachers was the degree to which students could choose the games they played each day versus being told to play certain games. "I boxed them

off according to subject and I let them decide because I feel like choices at times in teaching is a powerful thing" (Beatriz). Even teachers like Beatriz, who placed high value on student choice, however, clustered games together according to subject matter connections with the unit of study (for example, games related to Nullification Crisis only). Some encouraged students to "go back and pick the ones [they] like[d]" when they finished the cluster (Nancy). On the opposite end of the spectrum, some of the other teachers directed students to specific games, limiting choices for that day to a handful of options. "I'm going to give you 30 minutes to play the games and here are the games I think you should focus on during the class period" (Frank). Ann expressed the tightest levels of control:

I didn't give them completely free reign, I would usually give them a choice of a couple of games because [many of the basic ungated] games all had the same kind of content but they were presented in three different ways, that was the collector [slice], memory, and jumper games… For that, I thought it would be best to give the kids the choice of which one of those to play. If you really like fruit ninja, then go ahead and play fruit ninja... For those I would give them a choice and then for the other games I kind of divided them up into games that I thought were more review of content versus teaching of content. I asked them to play the ones that I thought were more teaching of content first, and then to play the ones that were more review content after that... At the beginning, it was like all right, you need to play these three games on Tuesday and these three games on Wednesday or whatever it was, but then toward the end of the pilot especially in that low class, I said you have 10 minutes to play these three games, pick one, go. All right, you have 15 minutes to play these two games, pick one, go. I was a lot more structured with it, because I felt they needed more structure. (Ann)

Several teachers discussed trade-offs and tensions in their interviews about deciding the manner and degree of choice offered to students. Teachers previewed the games and identified games that required more reading or had greater educational value in terms of depth. They noticed their students avoiding these games when given choice. Abe shared:

I think some of the games with the best game play were the ones with the least educational value. I think Nullies was a great example of that. While Nullies was a great game if the kids read what they were supposed to read, they just didn't do that. They [said] get this newspaper out of my face because I want to go back to playing. In that way, I would have liked to maybe direct them to say, "You have to play these two games before you play what you like." (Abe)

Recognizing this tension, other teachers provided students with more latitude. "I prefaced it with, if you know that you need extra work in some of these areas because you're going to be assessed on them, then you should find the games that are close aligned to what you need. I think some of them did" (Lawrence). According to the teachers, choice influenced student engagement and motivation to play. Balancing the student learning needs with student interest in the games was a central concern for most of the teachers. Ann's rationale for control, however, was based in pedagogical logic:

I think it at least allowed me to manage and be more aware of whether they were on task or off task. One of the things I noticed was happening when I wasn't specific about which game they should play would be that they would open a game and then play it for 30 seconds and then press the power button and go back and then they're putting another game in and they're going back and forth. They were just playing the same game over and over and over again, and not getting

the rest of the games, which for some of the units maybe that would work fine but for some of the other units, you needed to play all of them to get all the content. (Ann)

Ann thus provided the least choice to her students.

One final point for clarification in terms of the manner in which teachers implemented the games relates to our earlier analysis of students' ratings of the games and genres. More specifically, readers might express concerns about whether students were exposed to all five game genres before they did their ratings of each genre (i.e., basic ungated games, gated progress games, deeper content games, interactive stories, and intrinsic design games). As described in the current section, teachers introduced games by curricular topic (e.g., Nullification Crisis) and generally collected rankings as part of each curricular topic. Each curricular topic included games across the genres. Students therefore encountered all of the genres before doing the vast majority of their ratings.

Comparison of Implementations in Non-Game Versus Game Versions of Curricula

Across teachers, the curricula that the teachers implemented for their non-game classrooms generally reflected the same types of non-game activities that they included in their game classrooms. The teachers simply tended to include more of those same kinds of activities. References to instruction in the non-game groups indicated what the teachers described as "traditional instruction" in their interviews. Classes using the games were typically divided into 20-30 minutes of game time and 20-30 minutes of lecture, discussion, or what teachers described

vaguely as an "activity." Typical methods of instruction for non-game groups and non-game time in game groups included lecturing, question-and-answer style discussion, and independent worksheet time. Common materials included PowerPoint lectures with notes on events, dates, policies, and people that the students were expected to know. These lectures were accompanied by handouts with graphic organizers or questions to be answered for students to complete in small groups, with a partner, or individually. One participant described using poetry and music, and highlighted her effort to be creative and to engage students in that way on a regular basis, but while the vehicles of poetry and music may have been more engaging for her students, the focus of instruction remained on events, dates, policies, and people the students were expected to know. One teacher noted that he used primary source materials with students in both classes frequently, but he was definitely an exception to the common patterns outlined above for the rest of the teachers.

Overall, the teachers therefore did not indicate that they were engaging students in the kinds of practices or depths of learning history that experts in the fields of history education or social studies education have advocated in recent decades. As outlined in the introduction, researchers and educators across the fields of History Education and Social Studies Education have identified sourcing, contextualizing, and inquiring through careful reading and analysis of primary sources and documents as key best practices for students that constitute the work of a historian (NRC, 2005; VanSledright, 2011; Wineburg, 2001). Teachers, from this perspective, should be using primary source materials to engage students in using evidence from these texts to support historical arguments or narratives that answer questions historians might ask (NCSS, 2013).

Typically, however, research shows that history is learned through the presentation of facts in the form of dates, events, and people via a textbook or lecture positioning the nature of history, at best, as an exciting story (NRC, 2005). Thus, the teachers in the current study tended to implement the games into traditional approaches to teaching history focusing on the presentation of facts in the form of dates, events, and people via a textbook, lecture, and worksheet approaches to teaching history. These are the same practices that the teachers used in their non-game classrooms.

In terms of the current study, therefore, both the non-game and game classrooms involved traditional curricula, with games replacing some portion of these traditional curricula in the game-classrooms. From the perspective of the study, the control classrooms are therefore appropriate baseline points of comparison in the sense that the conditions compare games+traditional in the games classrooms to extended traditional in the non-game classrooms.

From the perspective of best practices advised by reform and standards documents, both the game and non-game classrooms focused on typical and traditional approaches to history education. The teachers' implementations of their curricula therefore align with the current study's emphasis on what is typical and probable in today's classrooms, but the teachers' implementations unfortunately do not demonstrate that integrating the games into the teachers' curricula transformed the teachers' overarching approach or perspectives on their goals for history education. What the teachers described is typical of history teaching in schools today.

Teacher Reactions and Perspectives Based on Surveys and Interviews

As shown in Table 2, most teachers thought the games were effective in their classrooms, agreed that games helped promote student engagement, and strongly agreed that they would like to use the games in the future. Responses to questions of overall impressions were consistently positive for all of the teachers except Ann. Several teachers noted general interest or positive perspectives with comments like "Overall, I really enjoyed it. What I enjoyed the most was the kids enjoying it" (Irina). As Abe explained:

The games were just a great way to engage all of your students. Like I said earlier there were a handful of students who think that for whatever reason they just didn't feel like playing the games that day. Some of the kids were a really caught up in a book that they were reading or they just wanted to talk to their friends or whatever. But for the most part the kids were really engaged and just playing the games. I would say that I would strongly agree for any of the kids not just the ones that typically engage. I will say that 95% of the kids were strongly engaged with the games. (Abe)

Most of the teachers expressed comments focused on student interest and time on task with the games. "They very much looked forward to playing the games each day, and looked forward to playing the next group of games" (Abe). Many of the teachers also described examples of student interest in relationship to the content changing. "Once we started, they were definitely more excited about coming to class and doing the material because they knew that half way

through class they would be able to play games" (Keith). One teacher noted that, despite teacher interest in the subject matter, students usually do not make relevance-based connections to ideas from Jacksonian Democracy. He felt that the games helped change that:

Jacksonian America is very important. Usually over quite a few teachers enjoy the subject. In my years, the students usually had a harder time relating to it. I think this helped give it a little bit more vigor. (Willis)

Willis is an example of the typical insights shared in the interview. Teachers thought students were more interested in the content because of the games.

There were exceptions to these positive experiences with increases in student interest. A noteworthy example involved the students being more excited about the games than the accompanying instruction. Viewed as positive interest in most cases, Celeste shared an alternative perspective:

[My class of high performers] weren't really looking at my teaching as much as 'let's just get through it, so we can play the game.' I was a little disappointed with them, because of that. I don't think they were valuing what I was teaching them as much as 'let's just get to the tablet.' (Celeste)

For Celeste, this was only one of her classes (the one she termed as normally higher performing), but marks the potential for teachers to notice increased motivation toward the games counterbalanced by decreased interest during other instructional time. Importantly, Celeste reported the opposite for her low performing class:

My period five, which is a majority of the Special Ed [students], they truly got a lot out of the games as an additional resource to what I'm teaching them for the lesson. I think they really got the most out of the games. They really used that as an additional tool for them. (Celeste)

As discussed, Ann was the most extreme outlier. Ann had high hopes for introducing digital games into her classroom. In fact, the high expectations she established before the trial may have been a primary factor in the failed implementation as the students' disappointment led to lower engagement. She thought she would see engagement and interest in the game class heighten substantially compared to her comparison class. She reported, however, that the games group and the comparison group she taught did not differ in terms of excitement or engagement:

I was really hoping that there would be a bigger difference, I guess. I was hoping that there would be a huge difference between the games and the regular instruction and that the kids would be so excited to get their hands on the tablet and do something different, that they would suddenly be super engaged and what I found was that actually wasn't the case. The kids who are engaged in school every day anyway, come to school on time and want to learn, they thought it was really cool but then to the other kids it was just another thing to blow off. It was kind of disappointing, I think. (Ann)

Ann's high hopes for improved engagement led to disappointment for her. In her game classes, students who typically engaged continued engaging, and those who did not typically engage in class activities continued not engaging.

Specific Aspects of Student Engagement and Collaboration

While the section above outlines teachers' overarching thoughts about engagement, this section explores specific observations about student engagement and collaboration. One important insight from these specific observations in terms of engagement is that some games were viewed as substantially harder than others. Games that required more reading, problem solving, or narrative building were viewed as more difficult. Willis, for example, stated that higher performing students also enjoyed reading more, which changed the kinds of games they enjoyed:

Which makes sense. Because some of [the games] required more reading and what have you. So it stands to reason that someone who enjoys reading more [will like those games more]. I would say the one in the museum also, definitely the higher performing students really liked that one. And like the lower performing students... There is less movement going on and you have to read a lot of the clues. [Another one] was a saloon [one of the immersive stories] I believe it was… I really enjoyed that one. My higher performing students really like that one, but lower performing students not so much. (Willis)

These are the games that the researchers had categorized as deeper content games and immersive stories. Teachers reported that these games supported greater depth or enhancement of learning, while quiz style games were viewed as good for review or reinforcing what students were learning.

Heightened student engagement also had interesting implications for classroom interaction and collaboration from the teachers' perspectives. In particular, some teachers noticed differences in discussion and the nature of cooperation in the classroom. "I think it facilitated more discussion in the classroom because the kids…could share their experiences playing the game with each other…the discussions seemed more excited [than the comparison group]" (Frank). Despite students playing individually, they were playing the same games and these shared experiences prompted conversations about the games, which turned to conversations about concepts, facts, and ideas from the unit. When asked what kinds of conversations students were having, Frank said:

They would say… stuff like 'I didn't know that about Andrew Jackson.' [And] they want[ed] to help each other. That was really nice to see them be 'Come on man, I'm going to show you where this is. It's over here.' That was good. There was just like organic cooperation. (Frank)

The emerging discussions and collaboration were prompted by game play and further facilitated by Frank's use of guiding questions. For this teacher, scaffolding student learning by setting a clear learning goal, being able to answer a particular question, supported enthusiastic collaboration among students. These observations of productive collaboration align with research on games (Clark et al., 2016; Wouters et al. 2013) showing that games are more effective in collaborative settings.

Other teachers experienced effects of the games in similar ways. "Students who…normally wouldn't be engaged in discussion I felt were more confident after playing the games because the information was so repetitive" (Beatriz). While repetition might have worn

thin on the novelty of the games, the variety of game formats apparently mitigated this effect. Repetition of content and concepts in different forms built a foundational understanding and confidence in students according to the teachers. Friendly competition was another mitigating factor of the potential effects of repetition on student interest or learning. Irina shared an insight on the relationship between interest, repetition, confidence, and competition:

Yes, because the act of playing the game, and because they thought the games were really cool, they usually remembered more information from that moment, and they would care more, because it ultimately became a competition thing. Like 'I know more information because that means I did better on the game. I got a higher score. So let me show you what I know.' (Irina)

Fostering inter-intra class competition could potentially lead to a tension between collaboration in learning and success in the competition, but teachers who noticed competition as a phenomenon acknowledged several positive outcomes related to increased motivation among students. Steve even suggested that one way to make the games better "…would be [to include] games that are a little bit more competitive." Considering potential drawbacks of increased competition in or around the games, Frank explained "…more competition, but I don't think they're going to totally stop helping each other because they want to win." In addition, competition served as motivation for student engagement with the games. Keith noted, "…[students] wanted me to leave the games on because their bases were doing so well and they were having a competition to see who could get to the highest level."

In addition to the productive interactions and collaboration catalyzed by friendly informal competition, the teachers also noted other productive collaborative interactions. When sharing

about typical non-game instructional collaboration, Booker noted that "…lower level kids really get helped by the higher and middle level kids, but sometimes they have the ability to sit back and really listen." From this teacher's perspective, student engagement typically defined as verbal interaction was reframed as a more equitable hands-on engagement with games. As he noted "the games gave [lower-level students] the opportunity to be just as hands-on as everybody else." In addition, teachers noted that during reflective debriefing or review time there were differences between game and non-game groups. Malloy described, "[Non-game group students] seemed like they didn't discuss as much, or the same few kids would answer the questions. I got more responses and kids elaborated more on the topics that we discussed [among game group students]." This teacher noticed far more game group students engaging in discussion with peers and offering rich and elaborative responses on discussion topics. As a result of their experiences with the games, collaboration outside of game time was richer from this teacher's perspective. Reviewing the unit with students who played games left teachers thinking that learning with the games deepened students' understanding of the content and historical context and made them more confident in their ability to share that learning with the class.

Academic Performance Status and Gender

Teachers expressed varied perspectives on differences in learning across students they identified as academically high performers or low performers. The quantitative results reflected this ambiguity in terms of the homogeneity of relative benefits between experimental and comparison conditions for the upper quartile and lower quartile of students based on prior grades in their classes. Most teachers indicated there were differences between high and low performers,

similar to Willis's earlier quote about higher performing students liking to read more and therefore liking certain games more, but the nature of the differences described by teachers varied. Broad statements like "I think some of the higher performing students may have preferred different games" (Brian) were typical in teachers' initial responses, but teachers also expressed more nuanced perspectives for other specific subsets of students. The major consistent finding involved Special Education students, a pattern that correlates well with the patterns in the quantitative data demonstrating larger benefits for Special Education students. "I have some Special Ed students who really struggle even with sociability with other kids and they were completely interacting with the other students no matter what level" (Nancy). Some teachers reported that their students with ADHD or ADD also benefited from increased "focus" (Brian and Celeste). Although the quantitative data did not demonstrate larger benefits for ELL students compared to non-ELL students, one teacher noted the games seemed particularly effective in supporting ELL students:

I think they were able to interact with them pretty well given their limited English. And it definitely served as a motivational tool for them. [The ELL teachers] usually paired them up with a bilingual student to help them out and also help read the questions to them, or what have you. But [the ELL students] definitely enjoyed them. As soon as they would come in, they would ask if they were going to be using them. So I feel it was very useful for them. (Willis)

In terms of students whom the teachers considered as typically low performing, some teachers noted improved confidence and foundational knowledge to participate more substantively in class. This was indicated by comments from teachers like, "Now when we were

doing our discussion it's not like all of the sudden they wanted to talk more, but I could tell that their base of knowledge had improved based on what they were learning from the games" (Keith) and "I think for some of my lower functioning students, it definitely helped them get some more information that they probably otherwise wouldn't have" (Lawrence). Related to the tension caused when giving students choices, several teachers noted that students they identified as lower performing "…would just start tapping through" (Frank) the more challenging games or avoid them altogether. This insight is clarified further through reports from teachers that "the lower level kids liked the more interactive content-driven games. Whereas some of the high level students really liked the challenging ones…Some of my lower level students liked the collector games and the memory-type games" (Booker). The teachers interviewed thus shared clear but varied benefits for different learners depending on perceived ability.

In terms of gender, most teachers (60%) saw no differences in engagement by gender, although some (40%) did report a slight preference for the games by boys. Typical perspectives included: "I didn't see gender differences. I think boys and girls were equally engaged" (Lawrence), "I think my class was relatively equal" (Booker), and "It was pretty universally loved I would say. There definitely were not any clear-cut boys and girls liking them more than the other. I do not think. It may be slightly more in favor of the boys" (Willis). Of the six teachers who noted gender differences, some observations were general: "I think, typically, the boys tend to be more excited and be more willing to play the games" (Abe). Some teachers offered a more specific insight related to gender differences and the games, hinting at ideas that the games were of interest to the boys because they were struggling to read. "My boys especially the ones who were struggling to read were the ones that were coming up and asking me to play

the game so it was kind of nice to see. The girls were a little slightly apprehensive at first but once we got going they were better with it. There was definitely a demographic there that these games reached" (Beatriz). Finally, some teachers noted differences in the specific types of games preferred by gender (e.g., boys liking the tower defense game best), but noted that all students liked playing the games.

Nature of the Games and Depth of Thinking Supported

In addition, teachers commented about the varying depths of thinking supported by the games. In terms of retention of information, several teachers noted improvements in their games class. For example, Celeste observed that the games "…corresponded with the material very well…it surely would have just been a very good reinforcement to what I was teaching…use it for review at the end of a unit" (Celeste). This made review and discussions richer productions in their classrooms. "I want to attribute to the tablets where the kids retained some of the information and were able to answer questions and discuss better than the ones who didn't play the games if I would do a review" (Malloy). Similarly, Keith reported, "As a teacher, I could definitely tell that they were picking up information and facts from the games. I think better than they had picked up information before…because it's visual and it's audio" (Keith). Teachers attributed improved retention of information to the variance in modality across games and between the games and the traditional texts used in class. From the teachers' perspectives, students' ability to recall was enhanced either because of (a) variation in mode and format directly or (b) increased interest and motivation to learn that were supported by that variation in mode and format.

In addition to increases in retention, teachers noticed that learning with the games led to deeper understandings. Willis shared that the games opened up spaces for questions about content that got at deeper conceptual understanding. Digging deeper into basic information supported students in making connections to similar phenomena today:

It definitely opened up some things. I got a lot of questions about the Trail of Tears. I got a lot of questions about... some of them just wanted to know more about Andrew Jackson as a person. I think... you know, rather than get just the basic information, they just felt it was... You know, a lot of questions. How would he react if he were running for president today and Donald Trump insulted him or his wife? You know. Would he challenge him to a duel? Just a lot of questions like that would come out. More than from the [control] class, I think even. So, I think finding out some little extra bits of information just peaked their interest to dig a little deeper. (Willis)

In a similar example Celeste shared that "[students] went to town on the Kitchen Cabinet part. They were questioning each other about it, talking about, in the election that's coming up, they were comparing how he would have picked people to, well, if Trump wins, who is he going to pick? It was a more in-depth discussion relating to today's president or the upcoming president" (Celeste). The students' game play not only strengthened explicit knowledge and information retention but built a foundation that supports thematic connections across time. From the teachers' perspectives this push to deeper thinking was in part because:

They could visualize, I think, the context more easily than they had been able to with the Comparison group; they weren't as able to visualize it as well because they didn't have the

visual, the cartoon, the animation. Their experience of being at the exact time, being some character from that time period (Frank).

Here Frank highlighted visualization similarly to Keith, but noted the way player agency and being in the game deepens student learning in support of contextual understanding.

Discussion

Several topics merit further discussion. The most important of these involves the teacher/student implementation dialectic, as highlighted by Ann's case. Other important issues are those related to the self-confidence findings and the positive results for Special Education students.

Teacher/ Student Implementation Dialectic

Ann's case makes clear that subtle factors in the interplay between the teacher and the students dramatically impact the efficacy of the games. Ann was disappointed by the lack of increased engagement among students in her game classes relative to her comparison classes. She had high hopes for the potential of digital games to get her students excited about learning, and noticed no changes in her students' interest, excitement, or engagement. She attributed this to broader issues of morale at the school:

That's why I'm saying it's not a kid problem; it's a school culture problem. They've been allowed to not care, and as a school, we can't seem to find a care big enough or effective enough to get students to be motivated to care on a school wide basis. This has been the case for the

three years that I've been there certainly, so I don't think it's a function of them being inner city, minority, high poverty students, so much as it is a function of when you allow students to create a culture where learning is not the reason you come to school every day. It doesn't matter what you give them and how many hoops the teacher jumps through and provides for them. There's just one more thing that is a struggle. That's why I really do think my results prove that, because the kids who were going to be engaged, they were engaged whether they were in the pilot unit or the control unit. The kids who were going to learn learned, and the kids who didn't, didn't. (Ann)

In addition to potential challenges with the way school morale or school culture might shape students' engagement, which none of the other teachers reported, three additional key differences between Ann and other teachers provide insights into the implementation strategies and the role of the teacher-student relationship in successful learning with the games. These key implementation differences were student agency, task structuring, and teacher framing.

First, in terms of student agency, Ann demonstrated some of the deepest and most detailed thinking and planning of any of the teachers, but she also implemented some of the tightest controls over students' choices in terms of game playing. There was substantial consensus across the other teachers about the need to balance the tensions of control and choice in terms of student engagement. It is possible that Ann's students, who appear to already have had low morale, would have benefited from a greater flexibility of choice. Self-determination theory (e.g., Deci & Ryan, 2010; Ryan & Deci, 2000), for example would suggested that constraining student agency would diminish students' motivation to learn. The other teachers varied in the degree of choice they provided to students, from teachers like Beatriz who only

grouped games by subject matter and then let kids choose within those groups, to teachers like Frank who directed students to a handful of options upon which to focus each day. Regardless, all of the teachers provided students with higher levels of choice and agency than did Ann.

Second, in terms of task structuring, while all of the other teachers balanced instructional time and gameplay time each day, Ann was the only teacher to cluster game play into fewer full class periods each week rather than dividing instructional time and game play more evenly each day. One could posit that consistent and even integration of the games into the curriculum would be more productive, which Ann's interview quotes suggests she discovered over time. By comparison, the other teachers typically engaged students in game play about 50% of the time spread across all days. Most commonly, teachers either started the class session or ended the class sessions with time for game play. As discussed earlier in the Implementation section, the most common strategy involved teaching a lesson the first half of the class period and then having students play the games the second half of the class period. "Most days, I just talked for the first half of class. Then games for the second half of class" (Abe). A few teachers followed a second strategy exemplified by Frank, who used "…the material of the games to help them form their knowledge of the content enough to do the assignments" (Frank). As a result, from Frank's perspective, students in his game classes demonstrated a deep understanding of the historical context and could situate their understandings in that context.

Third, in terms of framing expectations, it is possible that Ann set her students' expectations too high about the nature of the games before the students were actually able to play them.

In the low-level games class, there's a kid who I talked to his mother ahead of time, like we're doing this game thing, I really think it's going to help your son and she was like, oh that sounds like something he would totally be into. It's almost as if he knew that I was hoping that it would work for him, and he completely refused to play any of the games. He sat there with the tablet on his desk, refusing to play it. (Ann)

When asked if she had tried to get other students excited about the games in advance, Ann confirmed that she had:

The games classes I tried to build it up a little bit, like guess what we're going to do next, guys? We're going to do these things with the games… (Ann)

As a result, the students may have been disappointed by the games as they engaged with them because the games did not live up to students' expectations. By contrast, the other teachers tended to frame the games as part of school, albeit a fun part of school. Frank, for example, explained that:

Most class periods I started by introducing a topic and give them almost like a brief, main lesson on sometime like the bank crisis and then I would say to them, "I'm going to give you 30 minutes to play the games and here are the games I think you should focus on during the class period. I think these games give you the best information on the bank crisis and the most interesting to play at the same time…I would definitely play these games first." (Frank)

By integrating the games firmly within the structure of the curriculum in this manner, it seems possible that the games were framed for the students as part of the curriculum rather than being

framed as recreational games, which Ann's approach may inadvertently have done, which in turn may have decreased students' satisfaction with the games.

We only have interview data from Ann upon which to base these conjectures, and clearly morale was a big challenge at her school, but Ann's initial enthusiasm for the games, along with her palpable disappointment in the post interviews, provide a cautionary tale about the delicate teacher-student implementation dialectic in the successful integration of games into the curriculum.

The moderator analyses from the quantitative analyses reinforced these patterns in terms of the importance of (a) teacher and student attitudes toward the games and (b) teacher experience and school quality. The results indicated that students' multiple-choice factual knowledge outcomes were substantially stronger for the game condition relative to the comparison condition for the teachers who expressed the strongest interest in using the games in the future and for the teachers who reported that students who would otherwise be off-task were more likely to be on task with the games. Similarly, students in games classes demonstrated greater evidentiary depth for teachers who agreed or strongly agreed that their off-task students were being motivated. This variance in teachers' perceptions of whether or not the games were motivational for off-task students, and the variance's relationship to student outcomes, suggests a range of buy-in across classrooms. Similarly, the strong relationship between the extent of teachers' interest in using the games again in the future and outcomes also suggests that there was a range of student buy-in across classrooms. These patterns indicate a correlation between

success and teacher implementation, specifically the dialectic between the teacher and the students.

The quantitative analyses further reinforced this interpretation regarding the roles of teacher experience and school quality. In terms of students' self-confidence, students in the games classrooms demonstrated a small positive effect size of increased self-confidence compared to their respective comparison classrooms for the most experienced teachers (25 years of experience) and a small but increasingly negative effect size for teachers with less experience. Similarly, students in games classrooms demonstrated substantially greater response lengths in their open-response answers relative their respective comparison classrooms in higher-rated schools, but not in schools with low ratings (based on the ratings from SchoolGrades.org).

The results of the current study thus indicate that the role of the teacher, the nature of implementation, and cultural variables of the school and classroom are all critical. Research on digital games for learning often focuses only on the interactions of players with the games (Martinez & Clark, 2013). The findings of the current study underscore the importance of attending to the role of the teacher as well as to overarching socio-cultural and community variables in thinking about digital games for the classroom.

Student Confidence Ratings

Most of the quantitative analyses suggested some range of benefit, from minor to small but significant, for the games classrooms relative to the comparison classrooms. The exception involved students' self-confidence ratings about learning history, which was measured as the

average of two student responses to questions of how good they think they are in history class and how much they like learning about U.S. history. Although there was no overall difference in confidence for students in the game versus non-game conditions, results from the moderator analyses indicated that this effect varied by teacher experience levels. Namely, students in the game conditions had significantly lower confidence than those in the non-game condition when taught by less experienced teachers (d = 0.10 for teachers with 25 years of experience, d = -0.40 for teachers with 1 year of experience). A benign interpretation might be that students were interpreting "how good they think they are at history class" and "how much they like learning about U.S. history" in terms of the status quo of their typical history classroom practices, which might seem less appealing after having experienced the game-enhanced unit. It is also possible, however, that the experience of feeling effective in the game-enhanced curriculum actually made them feel less effective or optimistic about the confidence upon returning to the normal curriculum. As discussed regarding the teacher/student implementation dialectic in the preceding section, the teacher's approach to integrating the games into their curriculum seems critical.

Increased Confidence for Special Education Students

The one strong exception in terms of student confidence involved Special Education students. Special Education students in games classrooms demonstrated much higher self-confidence with a quite large effect size (d = 1.22) relative to Special Education students in comparison classrooms. Similarly, Special Education students in the game condition wrote significantly longer open-response answers than Special Education students in the non-game condition (d = 0.76). These are the two largest effect sizes in the study. These findings were

strongly supported in the teacher interviews, where many teachers described especially good learning, engagement, and social outcomes for their Special Education students. Teachers reported positive outcomes for Special Education students that included: (a) it put students on an equitable playing field in terms of task structure, (b) students interacted cooperatively across social or academic groups, and (c) Special Education students benefited academically from the visual and audio support in the games. Adding the games seems to have been particularly valuable and successful for these students, not only in terms of engagement and confidence, but also the students' integration into the social context of their classrooms.

Caveats and Limitations

Although the preceding sections have highlighted the importance of the teacher/student implementation dialectic to the efficacy of the games, it is important to remember that this study focuses on what is currently probable/typical for typical educational games and curricula rather than what might be possible with specialized or advanced game designs and curricula. The games in the current study did not target the deeper and more ambitious learning goals central to advanced recreational and educational game designs and mechanics (e.g., Gee, 2007; Squire, 2011). Therefore, the findings of the current study should not be taken as an upper limit of what might be possible from that perspective. As Clark, Tanner-Smith, and Killingsworth (2016) make clear in their meta-analysis on games for learning, design plays a far greater role in efficacy and outcomes than does the broad category of the medium.

Furthermore, the typical educational games in the current study left substantial room for additional optimization and refinement even within the category as we have defined it. Many of the games created for this study were excellent, but all of them could have been improved at a relatively low cost based on even a single round of feedback from teachers and researchers.

Finally, while the previous two caveats suggest a higher up limit to what is possible, a third caveat should temper predictions. None of the teachers in the current study regularly used games in their curriculum. Although this inexperience limited the teachers' ability to integrate the games into their curricula as effectively as they might with more experience, the novelty of the games likely increased student engagement beyond what we might be reasonably expected if games became a regular and more typical component in the classroom. Therefore, while some caveats of the current study suggest a potentially higher ceiling for what might be possible, we also cannot discount the potential of novelty effects, or even Hawthorne effects, that would predict lessened engagement if students encountered games more regularly in the curriculum.

With increased saturation in the curriculum, we might also anticipate that students' expectations and thresholds for what they considered a "good" game might also shift upward toward the thresholds that players expect from the production values and mechanics of the recreational games they enjoy. That said, there are four factors that reduce concerns about novelty effects. First, the students played the games for three weeks, by the end of which novelty effects would have attenuated to some degree. Second, the teachers were asked to use the games at a very high level of usage (50%). At a more moderate level of usage, perhaps the 32% of instructional time that teachers on average estimated for how much instructional time they would

want to devote in the future, novelty and variety would still exist. Third, new games in each new unit would enhance ongoing novelty and variety. Finally, we were surprised by how highly the students rated the simplest of the game mechanics (the basic ungated games with the runner, memory, jigsaw, difference detection, and slice mechanics), suggesting that the students find these mechanics engaging even though the students have undoubtedly encountered these mechanics extensively in past and even though these mechanics don't match AAA recreational game titles in sophistication or complexity. Thus we might anticipate some attenuation, but not disappearance, of the added value of novelty and variety for ongoing integration of games into the curriculum. Furthermore, teachers would likely become more experienced and proficient in successfully integrating games into their curriculum.

Conclusions and Final Thoughts

We used a cluster-quasi-experimental design to examine the efficacy of substantial integration of typical digital educational games into extended curricula, supplemented with qualitative implementation data from teacher interviews. The main findings from the quantitative analysis can be interpreted in two ways. If the one teacher who reported a failed implementation is dropped, the results demonstrate significantly higher gains for the game condition in terms of multiple-choice, open-response factual outcomes, evidentiary depth, and student engagement outcomes. If the failed implementation case is included in the analyses, all learning outcomes were higher for students in the game condition, but only student engagement and evidentiary depth were significantly higher. Moderator analyses further highlighted the role of teacher experience and student engagement in the efficacy of the game condition. Although the

beneficial effects of game instruction on students' outcomes were remarkably consistent for students across levels of prior achievement in history, the results indicated that the game instruction may be particularly beneficial for Special Education students.

All of these results from the quantitative analyses must be interpreted cautiously given that the quasi-experimental design precludes any strong causal inferences. Nonetheless, these findings highlight the potential of typical games for enhancing instruction, particularly when combined with the data from the teacher surveys. In terms of student engagement, the teachers rated students' engagement in the game classrooms as significantly higher than students' engagement in the non-game classrooms. In terms of future usage, all of the teachers except Ann "strongly agreed" or "agreed" that they would like to use games like these in the future (and Ann marked "unsure" rather than "disagree" or "strongly disagree"). Teachers on average expressed that they would want to use games like these in the future for 32% of their instructional time. The teachers thus viewed these games as a highly beneficial curricular activity.

The success of the implementation of games into the teachers' curricula appears to depend on a dialectic between the teacher and the students that determines student buy-in and attitude about the games. As discussed, in classrooms where this buy-in was high, results for the games classrooms relative to their control classrooms were impressive. Thus even with typical educational games, the role of the teacher and quality of implementation as well as cultural variables of the school and classroom are critical. Research and development of digital games for learning often focuses only on the interactions of players with the game (Martinez & Clark, 2013). The findings of the current study underscore the importance of attending to the role of the

teacher as well as to overarching socio-cultural and community variables in thinking about digital games for the classroom.

Finally, it is important to remember that this study focuses on what is currently probable for typical educational games and curricula without additional training or curricula for the teachers. The typical educational games of the current study do not target the more ambitious learning goals targeted by advanced game designs and curricula (e.g., Gee, 2007; Squire, 2011), the teachers received no additional training, and the typical educational games in the current study left substantial room for further optimization and refinement even as we have defined them. The positive outcomes therefore suggest substantial potential for educational games as a valuable addition to teachers' repertoire of curricular tools and resources, particularly when one considers the high engagement levels reported by the students and teachers. Even if the games had led only to the same level of learning, the games could be considered successful given the high engagement overall and the supports for Special Education students. Yet they performed better than that. Augmenting regular instruction with the games for three weeks significantly improved student learning.

References

Anderson, J. & Rainie, L. (2012). *Gamification: Experts expect 'game layers' to expand in the future, with positive and negative results*. Retrieved from http://www.pewinternet.org/2012/05/18/the-future-of-gamification/

Barab, S. A., Sadler, T., Heiselt, C., Hickey, D., Zuiker, S. (2007). Relating narrative, inquiry, and inscriptions: A framework for socio-scientific inquiry. *Journal of Science Education and Technology, 16*(1), 59-82.

Barab, S. A., Scott, B., Siyahhan, S., Goldstone, R., Ingram-Goble, A., Zuiker, S., & Warrant, S. (2009). Transformational play as a curricular scaffold: Using videogames to support science education. *Journal of Science Education and Technology 18*, 305-320.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284-290.

Clark, D. B., & Martinez-Garza, M. (2012). Prediction and explanation as design mechanics in conceptually-integrated digital games to help players articulate the tacit understandings they build through gameplay. In C. Steinkuhler, K. Squire, & S. Barab (Eds.), *Games, Learning, And Society: Learning And Meaning In The Digital Age* (pp. 279-305). Cambridge, UK: Cambridge University Press.

Clark, D. B., Nelson, B. C., Chang, H.-Y., Martinez-Garza, M., Slack, K., & D'Angelo, C. M. (2011). Exploring Newtonian mechanics in a conceptually-integrated digital game: Comparison of learning and affective outcomes for students in Taiwan and the United States. *Computers & Education*, *57*(3), 2178-2195.

Clark, D. B., Sengupta, P., Brady, C., Martinez-Garza, M., & Killingsworth, S. (2015). Disciplinary Integration in Digital Games for Science Learning. *International STEM Education Journal, 2*(2), 1-21. doi:10.1186/s40594-014-0014-4

Clark, D. B., Tanner-Smith, E., & Killingsworth, S. (2016). Digital games, design, and learning: A systematic review and meta-analysis. *Review of Educational Research, 86*(1), 79-122. doi:10.3102/003465431558206

Clark, D. B., Virk, S., Sengupta, P., Brady, C., Martinez-Garza, M., Krinks, K., Killingsworth, S., Kinnebrew, J., Biswas, G., Barnes, J., Minstrell, J., Nelson, B., Slack, K., & D'Angelo, C. (2016). SURGE's evolution deeper into formal representations: The siren's call of popular game-play mechanics. *International Journal of Designs for Learning, 7*(1), 107-146.

Deci, E. L., & Ryan, R. M. (2010). *Self-determination*. Hoboken, NJ: John Wiley & Sons, Inc.

Dieterle, E. (2010). Games for science education. In A. Hirumi (Ed.), *Playing Games in School: Video Games and Simulations for Primary and Secondary Education*. Eugene, OR: International Society for Technology in Education, pp. 89-112.

Duggan, M. (2015). Gaming and gamers. *Pew Research Center*. Retrieved from http://www.pewinternet.org/2015/12/15/gaming-and-gamers/

Federation of American Scientists. (2006). *Report: Summit on Educational Games: Harnessing the Power of Video Games for Learning*. Washington, D.C.

Friedman, A. M., & Heafner, T.L. (2007). You think for me, so I don't have to. *Contemporary Issues in Technology and Teacher Education*, *7*(3), 199-216.

Gee, J. P. (2003/2007). What video games have to teach us about learning and literacy. New York, NY: Palgrave Macmillan.

Gee, J. P. (2007). *What video games have to teach us about learning and literacy* (2nd ed.). New York, NY: Palgrave Macmillan.

Habgood, M. P. J., & Ainsworth, S. E. (2011). Motivating children to learn effectively: Exploring the value of intrinsic integration in educational games. *Journal of the Learning Sciences, 20*, 169–206. doi:10.1080/10508406.2010.508029

Hicks, D., Lee, J., Berson, M., Bolick, C., & Diem, R. (2014). Guidelines for using technology to prepare social studies teachers. *Contemporary Issues in Technology and Teacher Education, 14*(4), 433-450.

Hines, P. J., Jasny, B. R., & Merris, J. (2009). Adding a T to the three R's. *Science, 323*, 53. doi:10.1126/science.323.5910.53a

Kafai, Y. B., Quintero, M., & Feldon, D. (2010). Investigating the "'Why"' in Whypox: Casual and systematic explorations of a virtual epidemic. *Games and Culture*, *5*, 116–135. doi:10.1177/1555412009351265

Kee, K., & Graham, S. (2014). Teaching history in an age of pervasive computing: The case for games in the high school and undergraduate classroom. In K. Kee (Ed.), *Pastplay: Teaching and learning history with technology* (pp. 270–291). Ann Arbor, MI: University of Michigan Press.

Ketelhut, D. J. (2007). The impact of student self-efficacy on scientific inquiry skills: An exploratory investigation in River City, a multi-user virtual environment. *The Journal of Science Education and Technology, 16*(1), 99–111.

Klopfer, E., Scheintaub, H., Huang, W., Wendal, D. & Roque, R. (2009). The Simulation Cycle: combining games, simulations, engineering and science using StarLogo TNG, *E-Learning, 6*(1), 71-96.

Lee, J. K., & Probert, J. (2010). Civilization and whole-class play in high school social studies. *The Journal of Social Studies Research*, *34*, 1–28.

Lenhart, A., Kahne, J., Middaugh, E., Macgill, A., Evans, C., & Vitak, J. (2008). *Teens, video games, and civics* (p. 64). Pew Internet & American Life Project. Retrieved from http://www.pewinternet.org/Reports/2008/Teens-Video-Games-and-Civics.aspx

Martinez-Garza, M., & Clark, D. B. (2013). Teachers and teaching in game-based learning theory and practice. In M. Khine & I. Saleh (Eds.), *Approaches and Strategies in Next Generation Science Learning* (pp. 147-163). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-2809-0.ch008

Martinez-Garza, M., Clark, D. B. & Nelson, B. (2013). Digital games and the US National Research Council's science proficiency goals. *Studies in Science Education, 49,* 170-208. doi:10.1080/03057267.2013.839372

Mayo, M. J. (2009). Video games: A route to large-scale STEM education? *Science, 323*, 79-82.

McCall, J. (2014). Simulation games and the study of the past. In K. Kee (Ed.), *Pastplay: Teaching and learning history with technology* (pp. 228–253). Ann Arbor, MI: University of Michigan Press.

Metzger, S. A., & Paxton, R. J. (2016). Gaming history: A framework for what video games teach about the past. *Theory & Research in Social Education*, *44*(4), 532-564.

National Council for the Social Studies (NCSS). (2013). *The college, career, and civic Life (C3) framework for social studies state standards: Guidance for enhancing the rigor of K-12 civics, economics, geography, and history.* Silver Spring, MD: NCSS.

National Research Council (NRC). (2005). *How Students Learn: History, Mathematics, and Science in the Classroom.* Committee on *How People Learn*, A Targeted Report for Teachers, M.S. Donovan and J.D. Bransford, Editors. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Research Council. 2011. *Learning Science Through Computer Games and Simulations*. Washington, DC: The National Academies Press. doi:10.17226/13078

Pagnotti, J., & Russell, III, W. B. (2012). Using Civilization IV to engage students in world history content. *The Social Studies*, *103*, 39–48. doi:10.1080/00377996.2011.558940

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist*, *55*(1), 68-78.

Dieterle, E. (2009). Neomillennial learning styles and River City. *Children, Youth and Environments, 19*(1), 245–278.

Saye, J. W., & Brush, T. (2002). Scaffolding critical reasoning about history and social issues in multimedia-supported learning environments. *Educational Research, Technology, and Development, 50*(3), 77-96.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston, MA: Houghton Mifflin.

Sitzmann T. (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology, 64*, 489–528. doi:10.1111/j.1744-6570.2011.01190.x

Squire, K. (2006). From content to context: Videogames as designed experience. *Educational Researcher*, *35*(8), 19–29. doi:10.3102/0013189X035008019

Squire, K. (2011). *Video games and learning: Teaching and participatory culture in the digital age.* New York, NY: Teachers College Press.

Squire, K. D. & Jan, M. (2007). Mad City Mystery: Developing scientific argumentation skills with a place-based augmented reality game on handheld computers. *Journal of Science Education and Technology, 16*(1) 5-29.

Squire, K.D., & Klopfer, E. (2007). Augmented reality simulations on handheld computers. *Journal of the Learning Sciences, 16*(3), 371-413.

Steinkuehler, D., & Duncan, S. (2008). Scientific habits of mind in virtual worlds. *Journal of Science Education and Technology, 17*(6), 530-543.

Steinkuehler, D., & Duncan, S. (2008). Scientific habits of mind in virtual worlds. *Journal of Science Education and Technology, 17*, 530-543. doi:10.1007/s10956-008-9120-8

Swan, K. O., & Hoffer, M. (2008). Technology and social studies. In L. S. Levstick, and C. A. Tyson (Eds.), *Handbook of research in social studies education,* pp. 307-328. New York, NY: Routledge.

Tennessee Department of Education (2016). Eighth Grade History Standards. Retrieved from https://www.tn.gov/assets/entities/education/attachments/std_ss_gr_8.pdf

Tobias, S., & Fletcher, J. D. (Eds.) (2011). *Computer games and instruction*. Charlotte, NC: Information Age Publishing.

Vansledright, B. (2011). *The challenge of rethinking history education: On practices, theories, and policy*. New York, NY: Routledge.

Vogel J.J., Vogel D.S., Cannon-Bowers J., Bowers C.A., Muse K. & Wright M. (2006) Computer gaming and interactive simulations for learning: A meta-analysis. *Journal of Educational Computing Research 34*, 229–243. doi:10.2190/FLHV-K4WA-WPVQ-H0YM

Voorhees, G. A. (2009). I play therefore I am: Sid Meier's Civilization, turnbased strategy games, and the cogito. *Games and Culture*, *4*, 254–275. doi:10.1177/1555412009339728

Walz, S. P. & Deterding, S. (2015): An Introduction to the Gameful World, in Walz, S. P. & Deterding, S. (eds.), The Gameful World: Approaches, Issues, Applications (pp. 1–14). Cambridge, MA: MIT Press.

Whitworth, S. A., & Berson, M. J. (2003). Computer technology in the social studies: An examination of the effectiveness literature (1996-2001). *Contemporary Issues in Technology and Teacher Education, 2*(4), 472-509.

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*, 249-265. doi:10.1037/a0031311

Wright-Maley, C. (2015). Beyond the "Babel problem": Defining simulations for the social studies. *The Journal of Social Studies Research*, *39*, 63–77. doi:10.1016/j.jssr.2014.10.001

Young, M. F., Slota, S., Cutter, A. B., Jalette, G., Mullin, G., Lai, B., Simeoni, Z., Tran, M., Yukhymenko, M. (2012). Our Princess Is in Another Castle: A Review of Trends in Serious Gaming for Education. *Review of Educational Research, 82*(1), 61–89. doi:10.3102/0034654312436980

Appendix A

Coding Rubric for Level of Open-Response Factual Knowledge for Question 11

The open response questions were all coded in terms of level of factual knowledge on a scale of 1-10 points. Each question included its own rubric. The rubric for Question 11 is included here as an example. Example answers to Question 11 are included in the Methods section.

Question 11. How did Andrew Jackson stop South Carolina from leaving the country? Explain all the ways he did so and what the final result was.

PART 11FA: (6 points). Add up points for any and all of the following OR other accurate examples of the ways he stopped South Carolina from leaving the union (up to a maximum possible 6 points -- could be 1 or 2 points for each up to a total of 6 points)

- Put Federal troops on alert in S.C.
- Jackson issued a proclamation to the people of South Carolina that declared 1) union cannot be dissolved and…
- For stating that the proclamation also 2) that disunion by force was treason.
- In 1833 asked Congress to collect the tariff by force if necessary

- At the same time he supported a compromise bill to lower tariffs
- He got both bills passed congress and nullification was avoided
- Students could respond with other appropriate answers for "what Jackson did"

For each one

- 1 point for terse/minimal statement. For example: "He sent in the military"
- 2 points for some mostly accurate detail or explanation. For example: "He passed the Force Bill which allowed him to send military to force them to pay the tariff."

PART 11FB: (4 points). – Stating the final result 2 points for each of the following or other accurate results (up to a maximum possible 4 points -- could be 1 point for each of 6 or any combination of 1s and 2s up to 4)

- South Carolina stayed in the union
- South Carolina repealed the tariff nullification /backed down on nullification (1 point for agreed to follow rules)
- South Carolina didn't have support from enough other states
- Students could respond with other appropriate answers for "final result"

For each one

- 1 point for terse statement

- 2 points For some mostly accurate detail

Appendix B

Coding Rubric for all Questions for Level of Open-Response Evidentiary Depth

The open response questions were all coded in terms of level of evidentiary depth on a scale of 1-10 points. All questions employed the same rubric for evidentiary depth as outlined below. Example answers to Question 11 are included in the Methods section.

In addition to Factual Accuracy, code each open item in terms of evidentiary depth. Prompts explicitly (or in the case of #10 implicitly) call for students to explain, provide evidence, or give examples from the text (text is used here to describe any resource they have to interact with and then recall or make sense of to answer the prompts). In addition, we want to know if there were any moves toward contextualization or implications as ways of considering deeper cognitive demand.

Score Range 0-10

If there are more than two E3 (examples, evidence, explanation), then just choose the strongest two.

PART HA

- E1 (Example, Evidence/Reasons, or Explanations)
  - 0 - Absent

- 1 - Mostly Inaccurate (either in terms of relevance or language function or factually inaccurate or inappropriate) or Incomplete
- 2 - Mostly Accurate but terse (name or term only without elaboration)
- 3 - Mostly Accurate and elaborated
- 4 - Accurate and elaborated with explicit connections to other parts of question

PART HB

- For E2 (Example, Evidence/Reasons, or Explanations)
  - 0 - Absent
  - 1 - Mostly Inaccurate (either in terms of relevance or language function or factually inaccurate or inappropriate) or Incomplete
  - 2 - Mostly Accurate but terse (name or term only without elaboration)
  - 3 - Mostly Accurate and elaborated
  - 4 - Accurate and elaborated with explicit connections to other parts of question

PART HC

- Contextualization - Implications
  - 0 - No contextualization or implications beyond the E3

- 1 - Some minimal contextualization or implications beyond the E3
- 2 - Contextualization or implications beyond the E3

Figure 1. Basic Ungated Games included basic mechanics like flipping tiles to match (top left), running and jumping (top right), and comparing pictures (bottom). In all cases, content was encountered but demonstrating understanding was not part of mechanics.

Figure 2. Teacher, student, and researcher ratings of games (1 = lowest rating; 5 = highest rating)

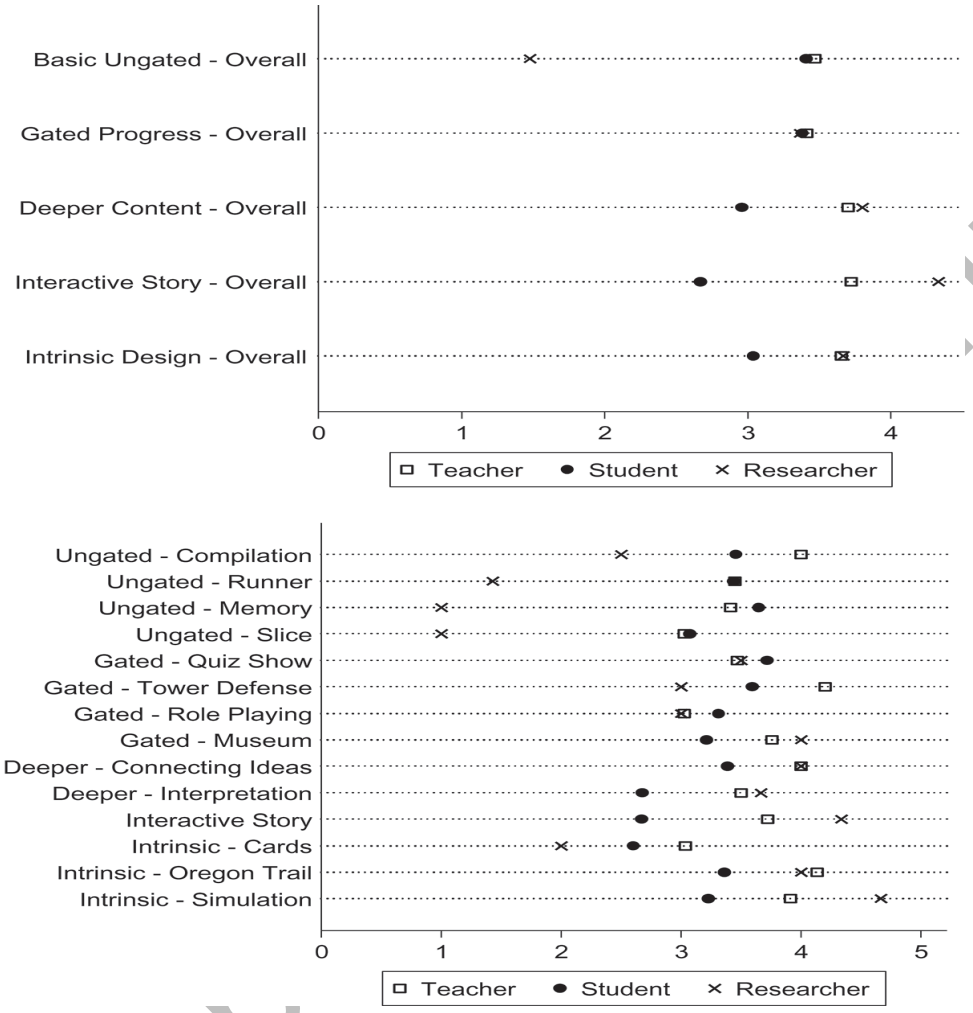

Figure 3. Gated Progress Games involved recreational game mechanics where progress was affected by performance on multiple-choice questions that would pop up (e.g., in order to attack in the top game or to earn extra resources in the bottom tower defense game).

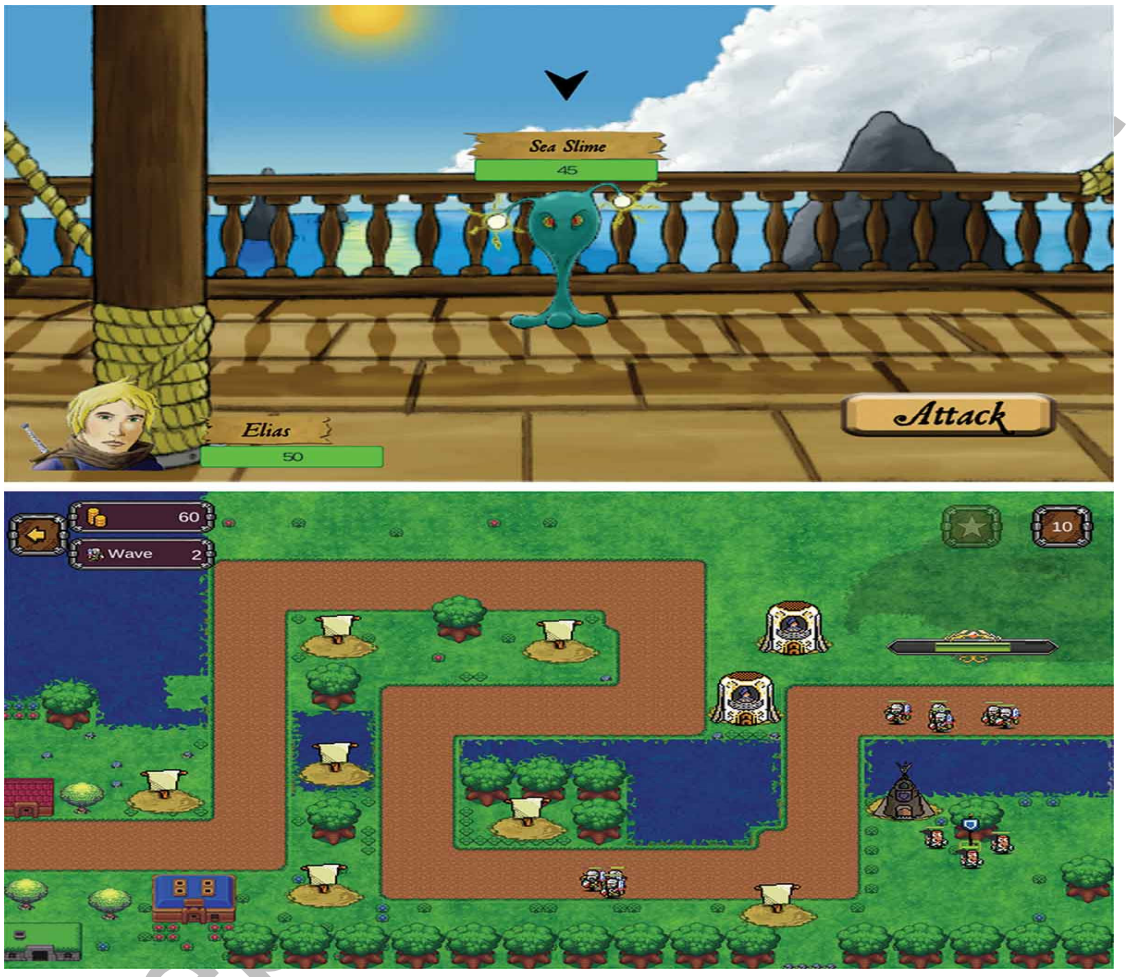

Figure 4. Deeper Content Games involved greater contextualization and/or involved interpretation and connection in an interesting or immersive contextualized manner. In the bottom picture, the player chooses answers that Jackson would say in a debate.

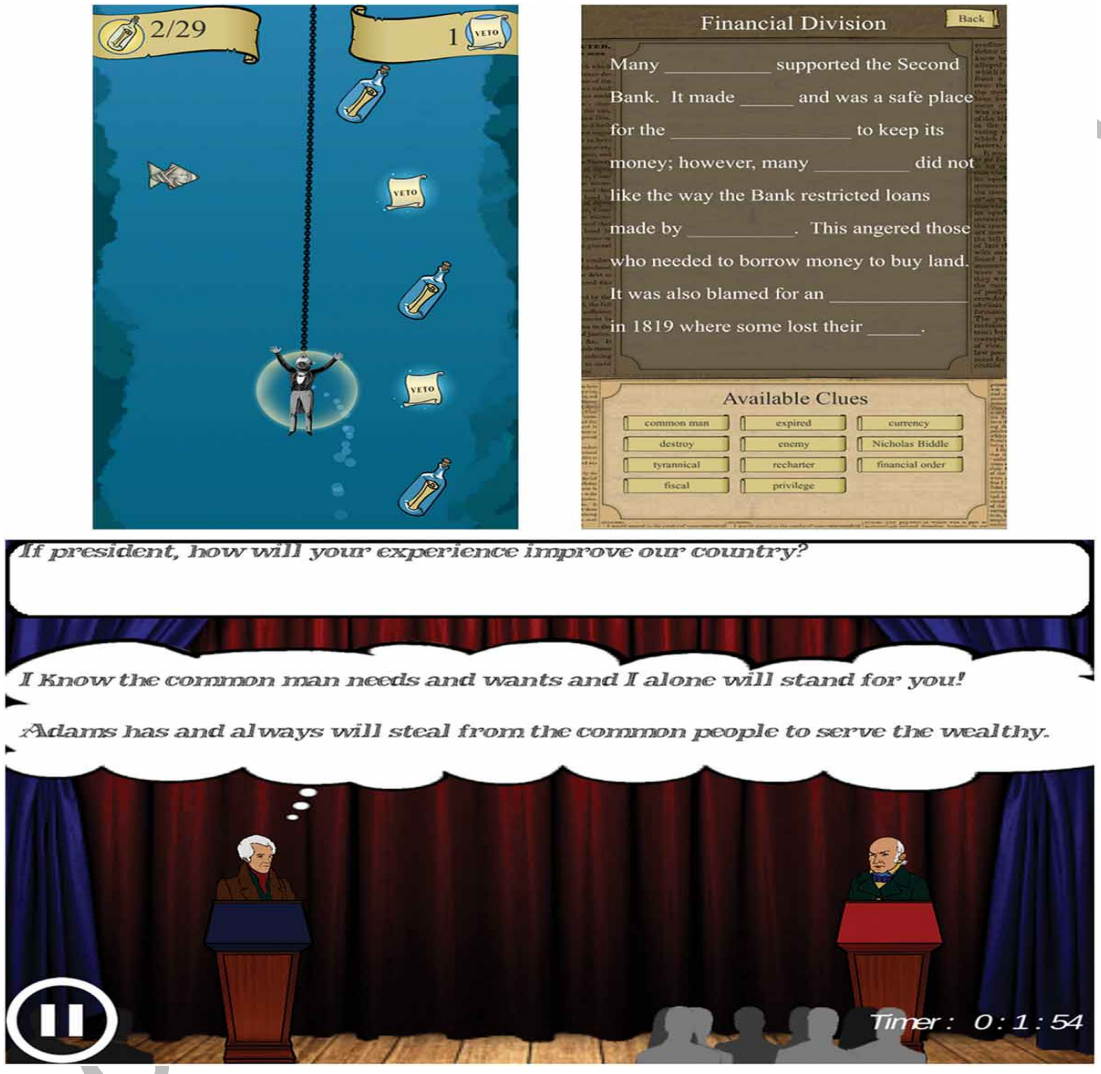

Figure 5. Interactive Story Games engaged the player in clicking on locations to enter various scenes, click on characters to read what they had to say, and click on responses to engage in simple dialog with the characters.

Figure 6. Intrinsic Design Games involved more complex game mechanics to explore content through means other than multiple-choice quizzes. The simulation game example in these screenshots engaged the player as a farmer farming in the south.

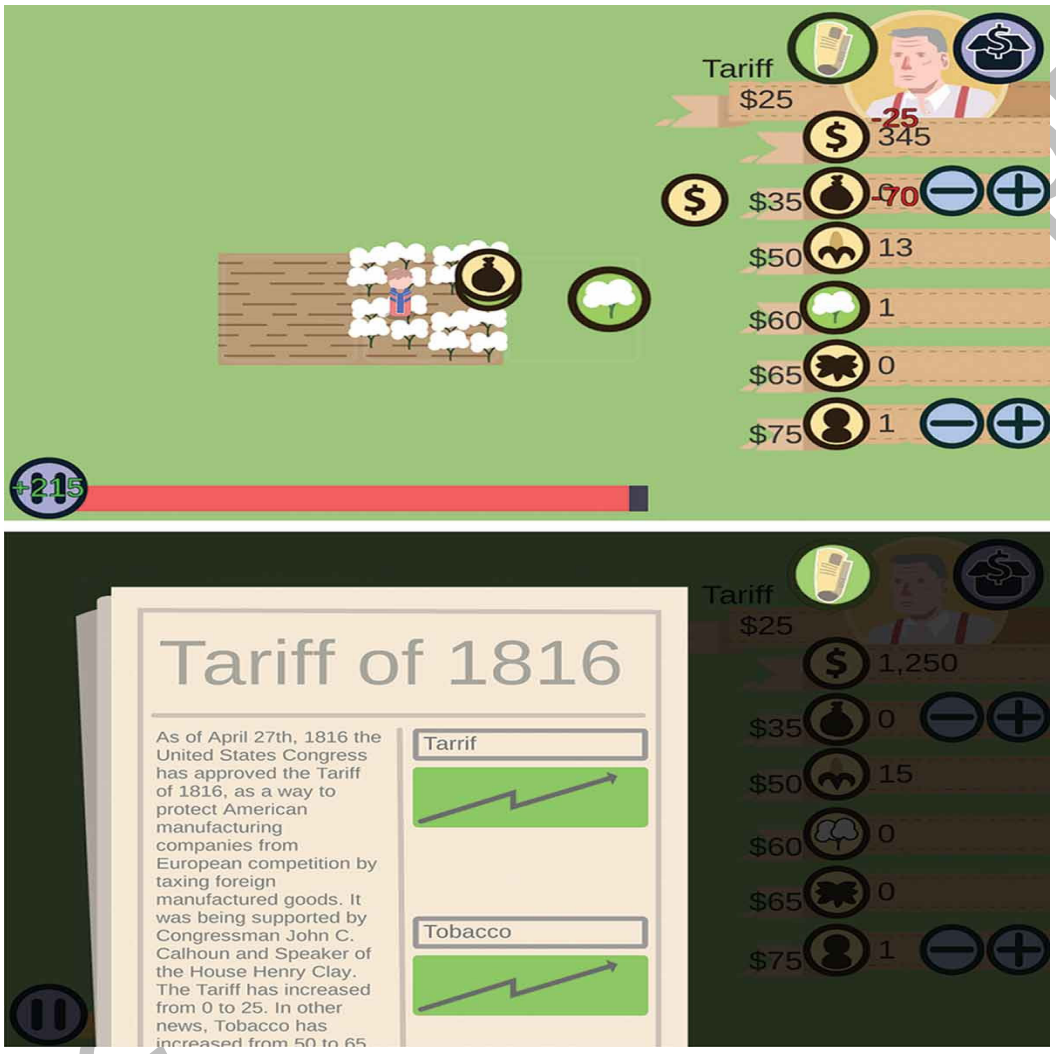

TABLE 1 Example Student Answers With Coding and Relative Frequency for Question 11

Coding & Frequency	*Student Answers with Grammar and Spelling as Written*
Factual Knowledge: 10 Evidentiary Depth: 8 Word Count: 82 Frequency: Rare	Andrew Jackson stop Carolina from leaving the cantry by saying if South Carolina was to leave this Union which is the United States it would be considered treason. He also sent the federal military to be on guard at South Carolina. He asked Congress to allow federal government to use force to collect the tarrif. He also supported the law of making the tarrif lower. The result was no other states supported South Carolina so South Carolina had to repeal their nullification.
Factual Knowledge: 8 Evidentiary Depth: 6 Word Count: 53	Andrew Jackson stopped South Carolina from leaving the country by threatening to use force to collect the tariff. He even sent troops to South Carolina to show this. He also passed a compromise bill that lowered the tariff. As a result, South Carolina repealed their demand for nullification and stayed with the US.

Frequency: Uncommon	
Factual Knowledge: 2 Evidentiary Depth: 2 Word Count: 45 Frequency: Common	Andrew Jackson stoped South Carolina from leaving the country by lowering the tax on things made with cotton. They ended up staying and making their own things with cotton. He gave them special government jobs. He gave other people than just rich people government jobs.
Factual Knowledge: 1 Evidentiary Depth: 1 Word Count: 24 Frequency: Common	Jackson stop them from leaving by cutting down their taxes 20%, and of how they settled the conflict in peace without starting a war.

Factual Knowledge: 0 “Blank”

Evidentiary Depth: 0

Word Count: 0

Frequency: Uncommon

TABLE 2 Characteristics of the Students, Teachers, and Schools in the Analytic Sample

	Game Condition (n = 440)		Non-Game Condition (n = 408)		
	Mean	*(SD)*	*Mean*	*(SD)*	*Range*
Student Outcomes					
Multiple-choice factual knowledge	2.44	(0.98)	2.34	(0.99)	0 - 4
Open-response factual knowledge	3.16	(1.43)	3.11	(1.42)	0 - 8.14
Open-response evidentiary depth	1.99	(0.94)	1.93	(0.90)	0 - 7.14
Length of responses (word count)	25.41	(12.72)	24.86	(12.01)	0 - 84.43
Interest in unit	7.10	(1.72)	6.62	(2.03)	1 - 10
Self-confidence in history	6.95	(2.05)	7.20	(1.99)	1 - 10

Student Characteristics					
Prior history grades	82.18	(15.39)	80.58	(15.20)	20 - 100
Special education student (%)	2.95%		3.18%		0 - 1
Individualized education plan (%)	2.50%		4.90%		0 - 1
English language learner (%)	3.18%		3.19%		0 - 1
Think games would make history better	7.79	(2.49)	7.87	(2.43)	1 - 10
School Characteristics					
Proficiency rating from SchoolGrades.org	40%	(18%)			19%-71%
School grade from SchoolGrades.org (%)					
A	22%				
B	22%				

C	22%		
D	33%		
School rating from GreatSchools.org	5.60	(2.99)	2 - 10
School Setting			
Urban	54%		
Suburban	23%		
Rural	23%		
Percent free/reduced price lunch	56%	(35%)	8%-95%
Percent White	51%	(35%)	3%-97%
Teacher Characteristics			
Grade level using game instruction	8.00	(0.94)	6 - 10

Years of teaching experience	12.08	(7.62)	1 - 25
Minutes of instruction	949.54	(146.44)	704 - 1224
Percentage of curricular time for games in games classrooms	50%	(15%)	25%-80%

Note: *SD* = standard deviation.

TABLE 3 Teacher-Reported Implementation Quality and Acceptability

	Mean	*(SD)*	*Range*
1 = Strongly Disagree to 5 = Strongly Agree			
Games allowed me to use instruction time more effectively	3.69	0.85	2 - 5
Students engaged in more rigorous discussions	3.46	1.13	1 - 5
Students performed better on daily assessments	3.69	0.75	2 - 5
Students who are normally off task were more focused	4.23	1.09	2 - 5
Easy to restructure curriculum to include games	4.08	1.12	2 - 5
Students can play all games with minimal support	3.62	1.26	2 - 5
I would recommend using games like these to others	4.69	0.48	4 - 5
I would like to use games like these on regular basis	4.46	0.66	3 - 5
Percentage curricular time I would like to use games regularly	32%	13%	17-50%

1 = Very Low Engagement to 5 = Very High Engagement			
Rate student engagement in game classrooms	4.31	0.63	3 - 5
Rate student engagement in non-game classrooms	3.23	0.73	2 - 4

Note: *SD* = standard deviation.

TABLE 4 Effects of Game Condition on Historical Knowledge and Student Engagement Outcomes

					Game		Non-Game	
Outcome		*b*	95% CI	Cohen's *d*	*Mean*	*n*	*Mean*	*n*
Original Analysis – All Teachers Included								
Multiple-choice knowledge	factual	0.05	[-0.11, 0.21]	0.05	2.45	434	2.34	405
Open-response knowledge	factual	0.12	[-0.03, 0.27]	0.08	3.17	434	3.11	405
Open-response depth	evidentiary	0.12*	[0.01, 0.22]	0.13	2.00	434	1.93	405
Length of responses		1.83	[-0.43, 4.09]	0.15	25.46	434	24.85	405
Interest in unit		0.52*	[0.24, 0.80]	0.28	7.11	388	6.65	315

Self-confidence in history		-0.34	[-0.75, 0.07]	-0.17	6.95	388	7.19	315
Sensitivity Analysis – Ann Dropped								
Multiple-choice knowledge	factual	0.14*	[0.02, 0.26]	0.15	2.48	398	2.33	367
Open-response knowledge	factual	0.16*	[0.00, 0.32]	0.12	3.19	398	3.13	367
Open-response depth	evidentiary	0.14*	[0.03, 0.24]	0.15	2.00	398	1.94	367
Length of responses		2.17	[-0.29, 4.63]	0.18	25.36	398	24.85	367
Interest in unit		0.54*	[0.26, 0.83]	0.29	7.10	357	6.63	278
Self-confidence in history		-0.35	[-0.80, 0.10]	-0.18	7.02	357	7.29	357

Notes: *b* = unstandardized regression coefficient from three-level linear regression models. Cohen's *d* = standardized mean difference effect size contrasting game versus non-game condition. *Mean* = average marginal [adjusted] mean, estimated from multilevel linear regression models. *n* = number of students in condition; sample sizes vary across outcomes due to student survey non-response.

*$p < .05$.

TABLE 5 Bivariate Correlations Between Student Outcomes and Ratings of Games by Category

	Basic Ungated		Gated Progress		Deeper Content		Interactive Story		Intrinsic Design	
	r	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*
Multiple choice factual	-.04	.89	-.01	.97	.12	.72	-.08	.81	.10	.77
Open-response factual	.15	.64	-.13	.68	.57	.05	.18	.57	.04	.91
Evidentiary depth	.24	.45	-.04	.90	.61	.04	.19	.54	.07	.84
Length of responses	.35	.27	.01	.97	.56	.06	.35	.26	.25	.43