Commercial Video Games as Preparation for Future Learning

Dylan Arena, Stanford University, darena@alumni.stanford.edu

Abstract: To examine the learning benefits on traditional school content of recreational play of commercial games, 102 community-college participants were randomly assigned to play *Call of Duty 2*, *Civilization IV*, or no game at home for at least 15 hours over 5 weeks. All participants then took a short multiple-choice test about World War II history; heard a 20-minute lecture about World War II; and then took another multiple-choice test about World War II history. Results (using an intention-to-treat analysis) showed no differences on the pre-lecture test but a positive effect (Cohen's d = .27) of gameplay on the post-lecture test, suggesting that recreational gameplay had prepared participants to learn from the lecture. These findings suggest a new role for games in learning contexts, in which the games—instead of carrying the educational load alone—provide compelling experiences that are coupled with the powerful explanatory structures of a formal curriculum.

Introduction

In the field of digital game-based learning there has been much focus on the development of educational games—games designed specifically to teach. Attention has also been paid to commercial, off-the-shelf (COTS) games, but that work has focused mostly on the motivation, engagement, and community participation engendered by such games. To complement such work, in this paper I hope to demonstrate that recreational COTS gameplay can, under the right circumstances, lead to learning gains on even the traditional, fact-based curricular instruments criticized by many in the educational research community but used widely in schools today. By showing this learning benefit of COTS games, I hope to strengthen the argument for digital game-based learning more broadly.

It is not surprising that there has been relatively little research demonstrating the benefits that might accrue on traditional fact-based tests from simply playing COTS games in one's leisure time, because there is no reason to expect that games created purely to entertain would produce learning benefits on school content that are measurable by traditional assessments. A novel assessment framework called Preparation for Future Learning (PFL), however, is designed specifically to measure immature forms of knowledge that traditional assessments miss. It does this by incorporating learning resources into the assessment process to determine what test takers' immature knowledge has prepared them to learn (Bransford & Schwartz, 1999).

In this study, I used the PFL framework to investigate whether being randomly assigned to receive and play one of two COTS games (*Call of Duty 2*, n = 34, or *Civilization IV*, n = 35) at home over the course of five weeks would prepare community-college students to learn from a lecture about World War II compared to a control condition that received no game (n = 33), as measured by performance on pre- and post-lecture multiple-choice tests. In a more qualitative vein, I also examined whether participants' gameplay experiences would influence their responses to open-ended questions about scenarios from World War II that were not mentioned in the lecture.

Methods Design

The study was an experimental field trial with non-random selection of participants from a convenience sample but random assignment of participants to three conditions. Two of these conditions' participants were given one of two commercial video games, *Call of Duty 2* (*CoD2*) or *Civilization IV* (*Civ4*), which they were assigned to play at home for at least 15 hours over a period of five weeks; the third condition's participants were given no game and assigned no gameplay. After this five-week period, participants from all three conditions came to a room on their community-college campus, took a 16-item pretest about World War II history, watched a 20-minute video of a narrated-slideshow lecture about World War II history, and then took a 36-item posttest and a brief survey. After completion of the session, participants were given research credit, and participants in the condition that had not already received a game were given one and asked to play it for at least 15 hours over a five-week period. (Although the post-instructional gameplay of these control-condition participants was not of theoretical interest for this study, I did not want control participants to participate less fully because of being assigned to a condition that received neither a game nor the opportunity for a gift card, so I explained in the initial information session that participants in all three conditions would receive games to play for compensation and that the only condition difference would be the timing of that gameplay.) All participants who completed sufficient gameplay (as determined by examination of save-game files) were e-mailed digital $75 Amazon.com gift cards. Figure 1 details this design.

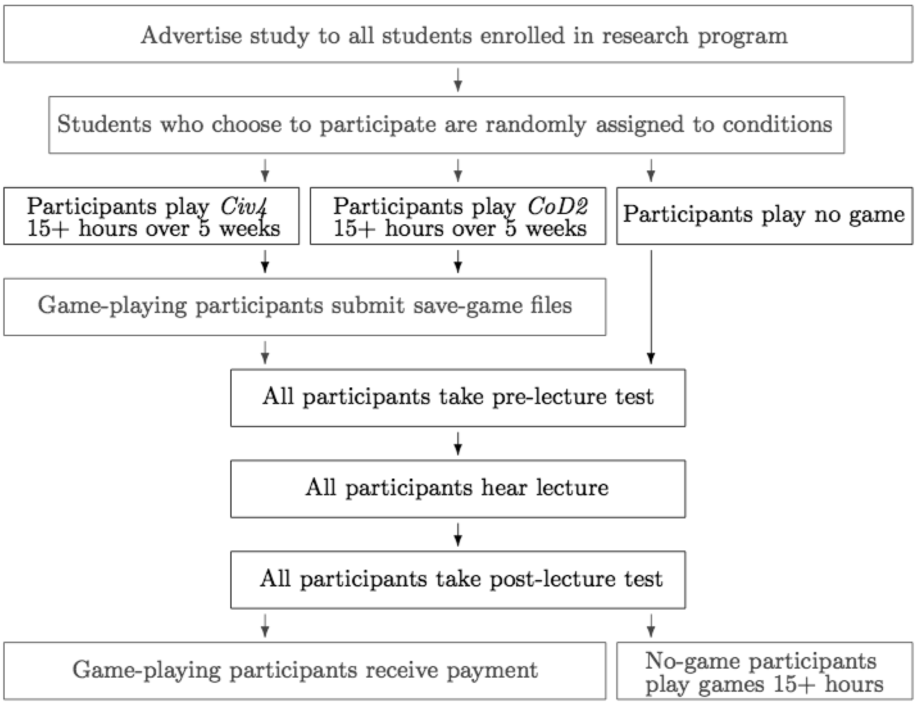

***Figure 1*: Experimental design and procedure.**

Participants

I recruited 119 participants and obtained usable data from 102(1), all of whom were students at a local community college who were enrolled in introductory social-science classes that required research participation for course credit. The final participant sample exhibited high diversity in ethnicity, nationality, socioeconomic status, and prior digital gameplay. Participants ranged in age from 16 to 42 years, with a median age of 20, and 66 of the 102 participants were female.

Materials

Each participant was given a new, shrink-wrapped copy of either *CoD2* or *Civ4*, installable on the participant's personal computer (both Windows and Macintosh versions were available).

The lecture took the form of my narration of a 24-slide presentation that discussed the events of World War II while focusing on the two themes of *Nations* and *Battles*, which I hypothesized to be more relevant to the players of *Civ4* and *CoD2*, respectively. Although the lecture was written so that players of the two games would retain information from different parts of it, I never mentioned either game in the lecture.

The multiple-choice-tests, which were intended to look for benefits of recreational gameplay even on the most traditional, "schoolish" measures, comprised items from the National Assessment of Educational Progress (NAEP: National Center for Education Statistics, 2011), the California Standards Test (CST: California Department of Education, 2009a and 2009b), and a purpose-built test from a World-War-II study guide produced by the company SparkNotes (SparkNotes Editors, 2005), which also served as a primary source for the lecture.

I also administered a pre-experimental questionnaire and an exit survey to (a) screen out anyone who had played either game before, (b) collect demographic data, and (c) solicit feedback about game enjoyment and learning behaviors.

As another way to examine effects of gameplay on learning, I asked participants open-ended questions about two scenarios presented in the exit survey after the post-lecture test. The scenarios had not been mentioned in the lecture but were intended to build upon the lecture's two themes of *Nations* and *Battles*, respectively. In the *Nations* scenario, British ships fired on their French allies in 1940 off the coast of Algeria. In the *Battles* scenario, American soldiers scaled a cliff under heavy German fire to disable an artillery battery. After presenting each scenario, I asked participants what they would want to learn to better understand the scenario.

Results

My primary analysis protocol was as follows. I first identified a set of variables that might predict performance on the pre- and post-lecture tests, as shown in Table 1.

	Control	*CoD2*	*Civ4*
Age in years			
M (SD)	22.76 (6.77)	22.65 (6.84)	22.60 (5.94)
English proficiency (reading/speaking composite)			
M (SD)	3.17 (1.23)	3.53 (1.71)	3.31 (1.56)
Game enjoyment			
1, 2, 3, 4, 5 (on Likert scale)	0, 0, 33, 0, 0	3, 5, 7, 11, 8	2, 6, 8, 15, 4
M (SD) [if treated as continuous]	3 (0)	3.47 (1.26)	3.37 (1.09)
Gender			
Females, Males	21, 12	21, 13	24, 11
Prior digital gameplay			
Never, 1-2 times, 3-6 times, > 6 times	5, 4, 7, 17	6, 7, 3, 18	3, 2, 4, 26
Prior social-studies interest			
1, 2, 3, 4, 5 (on Likert scale)	2, 1, 8, 13, 8	2, 3, 8, 15, 6	2, 4, 8, 11, 10
M (SD) [if treated as continuous]	3.73 (1.07)	3.59 (1.08)	3.66 (1.19)
Quarter			
Winter, Spring, Autumn	17, 7, 9	17, 7, 10	13, 13, 9

***Table 1*: Candidate predictor variables.**

Next, I constructed Analysis of Covariance (ANCOVA) models with the pre- and post-lecture-test scores as dependent variables and the variables listed in Table 1 as predictors, along with a gameplay-condition factor (because my primary interest for these analyses was the effect of gameplay rather than of each specific game, this two-level factor contrasts *CoD2* and *Civ4* participants with Control participants). For the post-lecture test, I also included pre-lecture-test score. My first step for each analysis was to test a model containing main effects of all predictors against a model that also contained all one-way interactions with the gameplay-condition factor, but in neither case did the marginal explanatory power of the interaction model reach statistical significance. Therefore, neither of the models discussed contains any interaction terms. Using the saturated main-effects model as a starting point, I then performed an all-possible-subsets model selection (2) to find the model with the highest R^2_{adj}, with the constraint that all predictors in models under consideration be at least marginally significant (p < .1). I will present these "parsimonious" models below.

Pre-lecture test

As shown in Figure 2, mean scores on the pre-lecture test were low for all three conditions.

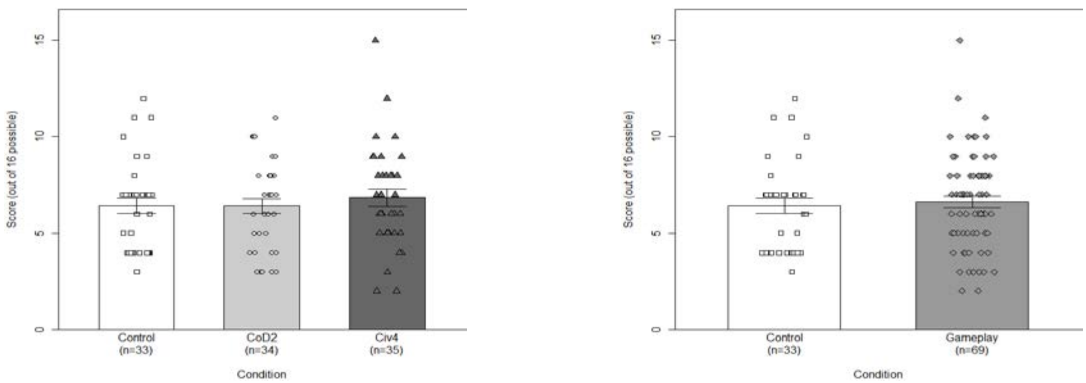

***Figure 2*: Pre-lecture-test scores by condition and gameplay.**

The test also had low reliability for this sample (Cronbach's α = .43), most likely because many participants were guessing on many of their response, thus introducing random noise to the measurement. The sole predictor chosen by the all-possible-subsets selection procedure for the parsimonious ANCOVA model for pre-lecture-test scores was English proficiency, as shown in Table 2. Notably, gameplay condition was not predictive.

	df	$SS_{TypeIII}$	***F***	η^2	***p***
English proficiency	1	70.86	13.32	.12	.00042***
Residuals	100	532.16			
R^2_{adj} = .11, $F(1, 100)$ = 13.32, p = .00042					

***Table 2*: Pre-lecture test ANCOVA.**

Post-lecture test

As shown in Figure 3, scores on the post-lecture test were much higher than on the pre-lecture test, indicating that participants in all conditions learned from the lecture. The test's reliability for this sample was also much higher than was the pre-lecture test's (Cronbach's α = .86). Mean scores in the gameplay conditions were slightly higher than in the control condition, suggesting a small benefit of gameplay (Cohen's d = .27).

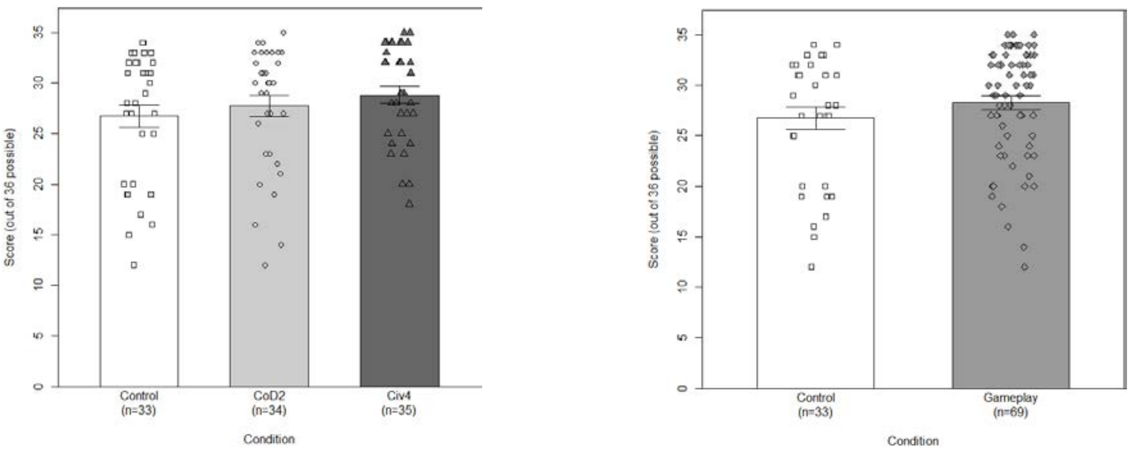

***Figure 3*: Post-lecture-test scores by condition and gameplay.**

The all-possible-subsets selection procedure for the ANCOVA model with post-lecture-test score as dependent variable produced a model with seven predictors, as shown in Table 3. Notably, gameplay condition (denoted by the "Received a game" variable) was selected in this model, albeit with a small effect size (η^2 = .026). Age, English proficiency, game enjoyment, pre-lecture-test score, and prior social-studies interest were also positively associated with post-lecture-test score. (Quarter of participation was marginally significant, reflecting cohort or seasonality effects that are not of theoretical interest but that contribute construct-irrelevant variance.)

	df	$SS_{TypeIII}$	***F***	η^2	***p***
Age	1	123.55	5.63	.036	.020*
English proficiency	1	70.86	13.32	.12	.00042***
Game enjoyment	4	239.07	2.72	.069	.035*
Pre-lecture-test score	1	145.08	6.60	.042	.012*
Prior social-studies interest	4	334.12	3.80	.096	.035*
Quarter	2	125.40	2.85	.036	.063.
Received a game	1	88.97	4.05	.026	.047*
Residuals	87	1910.95			
R^2_{adj} = .36, $F(14, 87)$ = 5.06, p < .0001					

***Table 3*: Post-lecture test ANCOVA.**

Open-ended questions

Because I had intended the *Nations* theme to resonate more with *Civ4* players and the *Battles* theme to resonate more with *CoD2* participants, I predicted that the *Nations* scenario would elicit more responses focused on *Nations* themes from *Civ4* players, and likewise the *Battles* scenario would elicit more *Battles*-focused responses from *CoD2* participants. Figure 4 shows results by condition for the two scenarios. In these graphs, correctness was determined by whether the questions participants asked about the scenario reflected a focus on the appropriate theme for that scenario: e.g., for the *Nations* scenario, a question about whether the French ships being fired upon were controlled by Germany would be scored as reflecting a *Nations* focus, whereas a question about whether the French commanders had insulted the British commanders would not. The error bars in each graph represent the 68% confidence intervals for the proportions (corresponding to roughly +/- 1 SD, assuming normality).

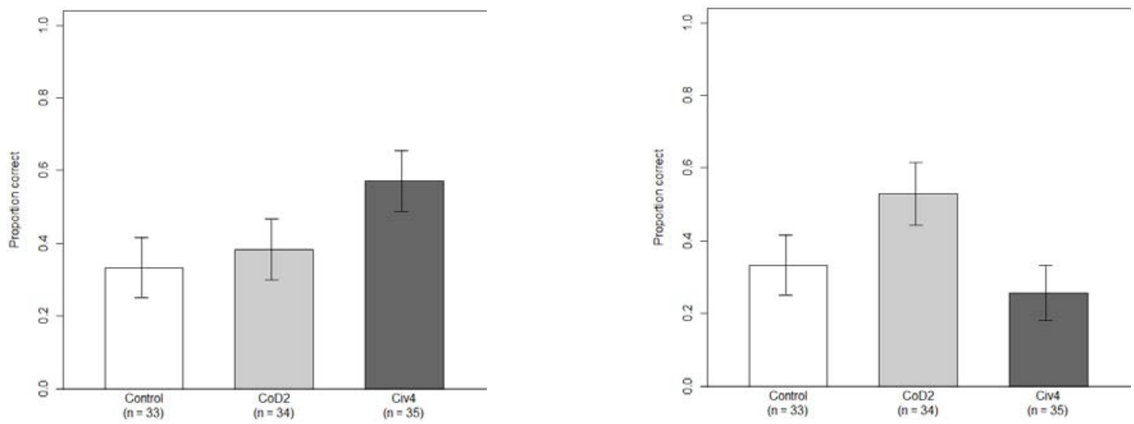

***Figure 4*: Open-ended responses for *Nations* (left) and *Battles* (right) scenarios.**

Fisher's exact test of proportions for the *Nations* scenario was marginally significant, p = .058; for the *Battles* scenario, the test was significant, p = .030.

Discussion

Taken together, the results of this study support the claim that playing enjoyable video games at home can help both male and female students learn in school, if the formal instruction leverages the students' gameplay experiences. (The strong predictive effect of prior social-studies interest for the post-lecture test shows the importance of also leveraging students' interests.) The multiple-choice-test results showed only a small effect, but this study was intentionally conservative with respect to its design (randomized field trial using an intention-to-treat analysis, both considered "gold standard" methodologies for causal inference), its intervention (recreational gameplay of commercial games that were not designed to teach school content), its outcome (learning gains compared to control participants on traditional history content delivered via direct instruction), and its measurements (traditional multiple-choice tests whose items were taken from existing standardized tests rather than developed *ad hoc*). The open-ended-question results underscore the notion that more creative measures show stronger positive effects of recreational gameplay. They also demonstrate that different games will offer different types of experiences that prepare players preferentially for different topics of formal instruction and that these gameplay experiences can improve not only retention of facts presented by direct instruction but also students' choices about what to learn.

I propose two main conclusions from this study. From a theoretical perspective, there is a benefit to simply having demonstrated that the learning that occurs in naturalistic gameplay can be detected with the PFL framework. Showing that fruit can be plucked from this region of the digital-game-based-learning space (i.e., involving pre-instructional gameplay paired with a formal curriculum) strengthens the basic argument for digital game-based learning.

From a more pragmatic perspective, demonstrating that informal COTS gameplay can lead to learning gains in schoolish contexts suggests a specific policy prescription for educators: to consider the games their students are already playing not as competition for precious time that could be spent studying or doing homework but as rich source material for use in engaging curricula that could tie the compelling experiences found in the games with the powerful explanatory structures found in the standard curriculum. A concomitant policy prescription for commercial-game developers is that their games could contribute to efforts in digital game-based learning without having entire curricular units crammed into them—developers need only be thoughtful about how the experiences provided in their games might be tweaked this way or that to better serve as foundations upon which educators might build. This lowered bar for participation in the digital-game-based-learning space might encourage more commercial developers to lend their considerable strengths to the process of bringing classrooms into the 21st-century.

Endnotes

(1) The 16 participants who dropped out did so because either they were no longer enrolled in a relevant social-science class or because they had technical or time-management problems preventing them from completing the at-home gameplay. The 17th lost participant was removed for cheating on the post-lecture test.

(2) An all-possible-subsets selection procedure examines every model that could possibly be constructed using a set of predictor variables to find the "best" model according to some pre-defined criterion.

References

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. In A. Iran-Nejad & P. D. Pearson (Eds.), *Review of Research in Education* (Vol. 24, p. 61-100). Sage.

California Department of Education. (2009a). Released test questions: Grade 11 U.S. history - social science. Retrieved from: http://www.cde.ca.gov/ta/tg/sr/documents/cstrtqhssmar18.pdf on January 10, 2011.

California Department of Education. (2009b). Released test questions: World history - social science. Retrieved from: http://www.cde.ca.gov/ta/tg/sr/documents/cstrtqhssworld.pdf on January 10, 2011.

National Center for Education Statistics. (2011). NAEP questions tool. Retrieved January 10, 2011, from: http://nces.ed.gov/nationsreportcard/itmrlsx/search.aspx?subject=history.

SparkNotes Editors. (2005). Sparknote on World War II (1939-1945). Retrieved December 22, 2010, from: http://www.sparknotes.com/history/european/ww2/.

Acknowledgments

Financial support for this study was provided by a Stanford University School of Education Dissertation Support Grant and a Gerald J. Lieberman Fellowship. Theoretical support was provided by Dan Schwartz and Ed Haertel.