Computers in the Schools
Interdisciplinary Journal of Practice, Theory, and Applied Research

ISSN: 0738-0569 (Print) 1528-7033 (Online) Journal homepage: http://www.tandfonline.com/loi/wcis20

The Effects of Game Design on Learning Outcomes

Michael W. Martin & Yuzhong Shen

To cite this article: Michael W. Martin & Yuzhong Shen (2014) The Effects of Game Design on Learning Outcomes, Computers in the Schools, 31:1-2, 23-42, DOI: 10.1080/07380569.2014.879684

To link to this article: https://doi.org/10.1080/07380569.2014.879684

Published online: 18 Apr 2014.

Computers in the Schools, 31:23–42, 2014

ISSN: 0738-0569 print / 1528-7033 online
DOI: 10.1080/07380569.2014.879684

The Effects of Game Design on Learning Outcomes

MICHAEL W. MARTIN
United States Military Academy, West Point, New York, USA

YUZHONG SHEN
Old Dominion University, Norfolk, Virginia, USA

This article details the administration and results of an experiment conducted to assess the impact of three video game design concepts upon learning outcomes. The principles tested include game aesthetics, player choice, and player competition. The experiment participants were asked to play a serious game over the course of a week, and the learning outcomes were measured by comparing their pretest and posttest scores. The results of a one-tailed t *test indicated, with a* p *value of 0.043, that there was a statistically significant effect of the aesthetic presentation of the game upon the learning outcome. There was no indication of a significant effect by the player choice or player competition conditions, but the results from these experiment groups point to some potentially interesting interactions between the conditions and learning, as well as possible future lines of experimental inquiry.*

KEYWORDS serious games, game design principles, learning outcomes, learning

Video games are a rapidly growing part of modern culture. In the mid 2000s, the commercial entertainment software industry eclipsed Hollywood in terms of revenue (Yi, 2004), and in 2008, sales of video games surpassed the global sales of DVDs (Zelfden, 2009). The ubiquity of smart phones and the growing popularity of online communities, such as Facebook, have also served as vectors by which video games have infiltrated many aspects of our daily lives.

Address correspondence to Yuzhong Shen, Old Dominion University, Department of Electrical and Computer Engineering, Norfolk, VA 23529, USA. E-mail: yshen@odu.edu

For the past decade and a half, there has been a commensurate growth in efforts to apply these video games toward productive ends. This emerging field is known as *serious games*. Serious games have been characterized as the aspiring successor to the defunct field of *edutainment* (Zyda, 2005). Whereas edutainment may have been described as the worst of game development in the 1980s and 1990s, serious games are part of a revitalization effort intended to capitalize on the current explosive growth and cultural acceptance of video games.

There are several educational scholars and notable video game developers who have offered a wealth of insight into possible benefits of serious games. Marc Prensky described how video games can have the potential to teach children by providing inspirational educational experiences (Prensky, 2001). Prensky (2006) explored how these experiences can be used to teach the emergent skills required in our techno-centric, evolving world, as well as the traditional educational subjects that students might not commonly find inspiring (Prensky, 2001). Gee (2007) explored the teaching mechanisms that video games regularly employ, and how players benefit from them. Koster (2005) reversed the perspective on learning and games, describing how the most fun games are those in which the player is learning. Shelton and Wiley (2007) compiled a persuasive collection of essays on the role of games in learning, including Shaffer's (2007) essay claiming that video games allow a more authentic form of learning than traditional educational methods. Much of this work is founded upon insightful personal and professional anecdotes in video game development or educational fields. However, there is a deficit of empirical data to reinforce these qualitative assessments. There are compelling theories regarding the potential merits of the field, but there is not a large body of experimental data from which to verify that the merits exist, and that they do indeed result from serious game use. Further, though the theorized benefits or observed results may be well detailed, there does not appear to be as much discussion on how to create serious games and how to specifically engender the potential benefits. Our research sought to make a modest contribution to the body of knowledge by conducting quantitative experiments to empirically test the efficacy of some basic game design principles in eliciting learning outcomes.

This work builds upon the theoretical foundation previously presented (M. Martin & Shen, 2010a, 2010b; M. W. Martin & Shen, 2010). In these works, the authors proposed a definition of serious games that includes synthetic propositions stating that serious games should be self-encapsulated, intrinsically compelling, and should provide the user with choices. This definition is, in turn, drawn from an amalgam of different definitions of what a game or serious game is (Crawford, 2003; Koster, 2005; Michigan State University, 2010; Prensky, 2006; Salen & Zimmerman, 2003; Schell, 2008; Shaffer, 2007). The analysis and synthesis of these various definitions served as a means to identify potentially important features of serious games.

Based on these identified features, we developed a set of experiments to specifically test the effects of *being intrinsically compelling* and *providing users with choices* upon the learning outcomes derived from using a serious game. Through further analysis of available serious games and popular commercial entertainment, the quality of being intrinsically compelling was further broken down into *aesthetic presentation* and *player competition*. The feature of providing user choices was deemed suitably granular for implementation, without the need for further distinctions. These features are not intended to be an exhaustive list of potential serious game aspects, but instead, are simply a starting point to begin exploring the mechanisms that influence the effectiveness of serious games.

The experiment to test these three features was carried out using Element Solitaire©, a serious game. This game, developed by us, teaches the placement of chemical elements on the periodic table. The learning content in the game can be categorized as simple declarative knowledge (or rote memorization), and the game is intended for use by novice learners in the chemistry field. The game is conceptually based on a combination of the game of solitaire with the periodic table of elements.

DESIGN

Hypotheses

This study was designed to test the following three hypotheses:

> Hypothesis 1: A serious game with added aesthetic features, including enhanced graphics and music, will result in higher learning outcomes than an identical one without music and graphics.
>
> Hypothesis 2: A serious game with meaningful choices will result in higher learning outcomes than an identical game without meaningful choices.
>
> Hypothesis 3: A serious game with competition will result in higher learning outcomes than an identical one without competition.

These hypotheses are directly drawn from the serious game features identified previously.

Design

The three hypotheses, in turn, were directly translated into experimental conditions. The first condition was the presence of aesthetic effects within

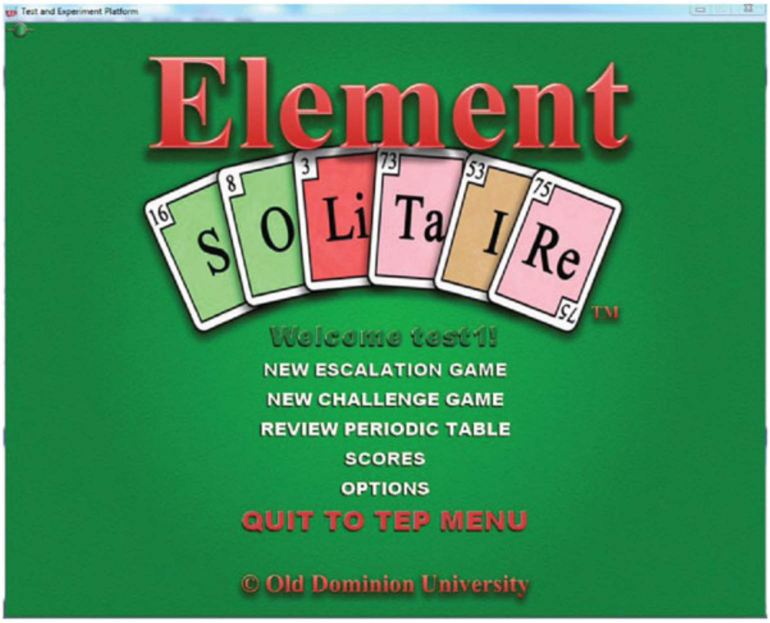

FIGURE 1 High aesthetic presentation title screen.
© Old Dominion University. Reproduced with permission of Old Dominion University. Permission to reuse must be obtained from the rightsholder. (Color figure available online.)

the game. For Element Solitaire, these artifacts included enhanced graphical presentation, as well as additional aesthetic effects such as sound effects, music, card movement animation, and "sparkle" particle effects. Figures 1–4 provide comparisons of the graphical differences in this condition. However, it is important to emphasize that the difference in aesthetic presentation went beyond static graphics and included not only animation, but also sound effect and music. Care was taken, however, to ensure that the learning content was consistently presented in all of the experiment conditions.

The second condition was the presence of specific choices within the game. These choices were made of two specific options available to the player: hint and skip. The players had a budget of 10 skips they could use to delay having to place an element that they might be unsure of. The hint gave the learners the option to "purchase" hints for a small score penalty. These hints gave the players information needed to correctly place an element, but diminished the score received when they ultimately place the element correctly.

The third condition was the presence of a network-enabled scoreboard that allowed the player to see what is commonly referred to in the entertainment software industry as a "leaderboard." This leaderboard listed the names and scores of the top 10 players. This is subtly different from a more direct scoreboard, as each player may only have one entry on the board, marking

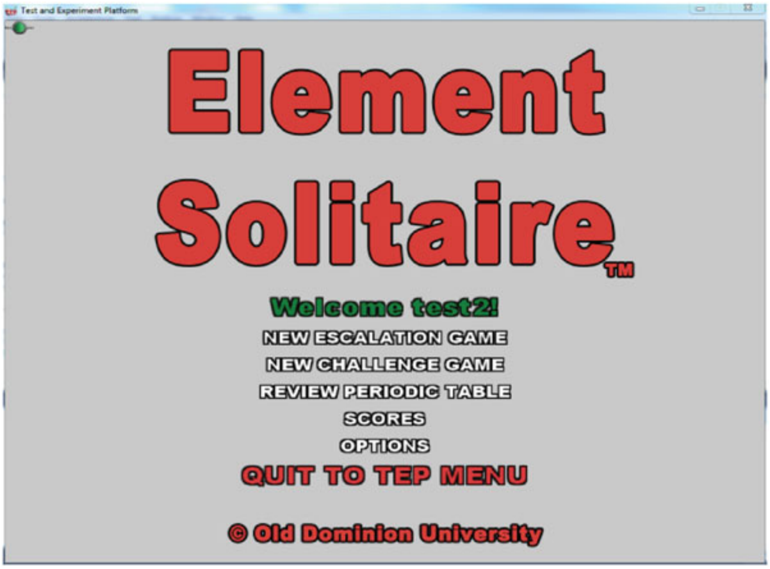

FIGURE 2 Low aesthetic presentation title screen.

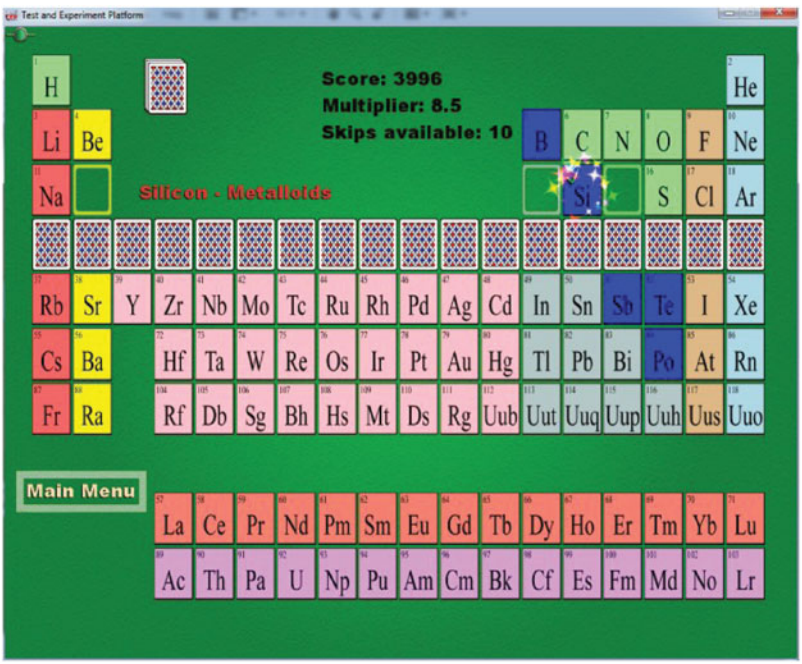

FIGURE 3 Game in play with the high aesthetic condition. The sparkle effect is visible around the newly placed silicon (Si) element.

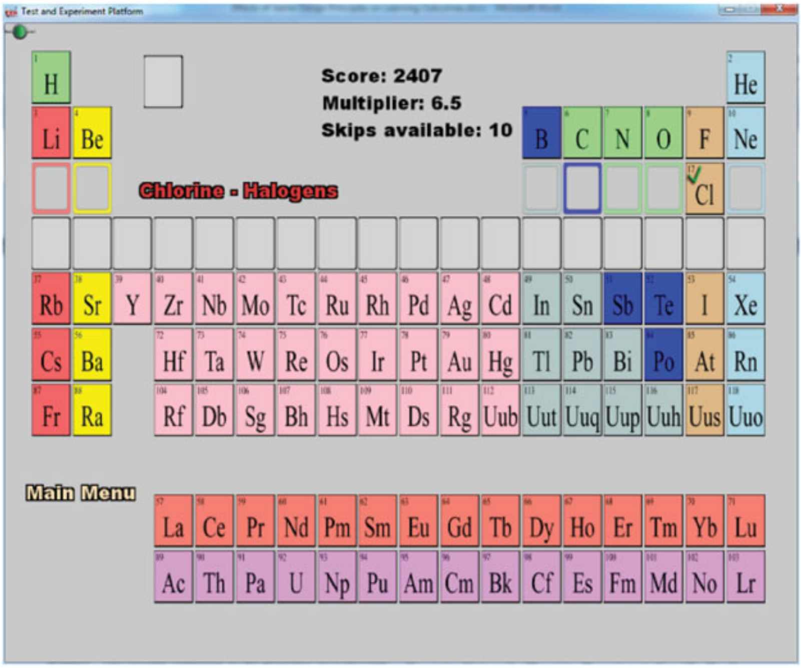

FIGURE 4 Game play with the low aesthetic condition.

his or her highest score. Players without this competition condition were only shown their own scores, in a traditional scoreboard fashion. Figures 5 and 6 show the difference in the scoreboards between the two condition levels. For privacy purposes, the usernames in Figures 5 and 6 have been blurred.

The first experimental treatment group was a baseline treatment, in which all three conditions were present. The game featured the full aesthetic presentation, the full breadth of available gameplay choices, and the networked leaderboard. This version of Element Solitaire was considered to be the standard version of the game, and this treatment group served as the baseline from which to compare the performance of the other experimental treatments.

The remaining experimental treatments were reductive iterations of this baseline treatment. The second treatment retained the choices and scoreboard conditions, and omitted the aesthetics condition. The third treatment retained the choices and aesthetics conditions, but omitted the scoreboard condition. The fourth and final treatment omitted the choices condition, while retaining the scoreboard and aesthetics conditions. The control group effectively omitted all conditions by not being exposed to the game at all.

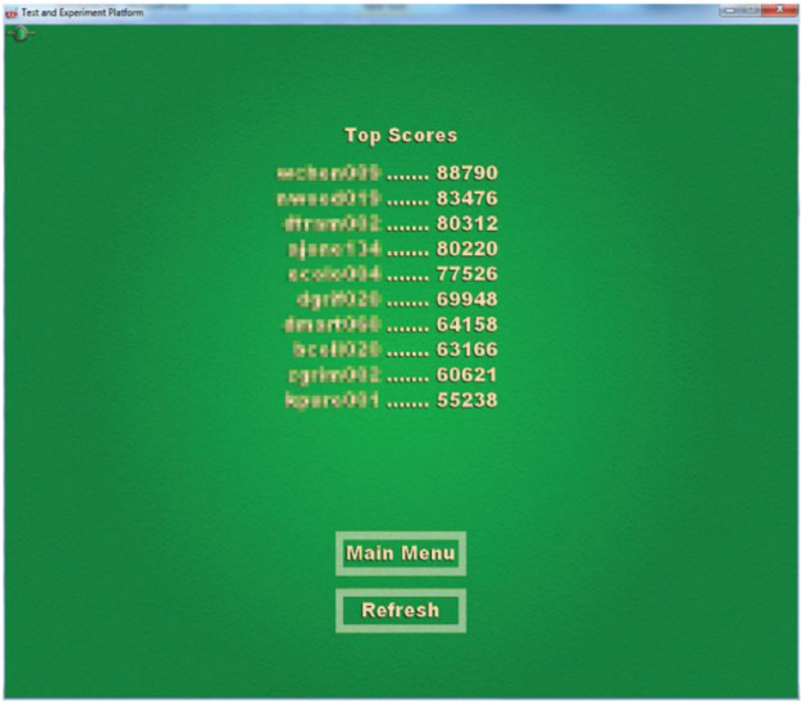

FIGURE 5 Scoreboard from a high competition condition participant. The scoreboard shows a leaderboard displaying the single top score of each of the 10 highest scoring players. Usernames are blurred to protect their identities.
 (Color figure available online.)

Control group participants were asked only to take the pre- and posttest. The group-to-condition interaction is shown in Table 1.

Measurements

Before playing Element Solitaire, all treatment groups took a pretest to create an initial assessment. Participants were then asked to play the game four times (with the exception of the control group). Once they completed the required number of games, they were then asked to take the posttest. The learning outcome was defined, for the purposes of this experiment, as the difference between the pre- and posttests scores. The posttest was administered one week after the pretest, and participants had that week in which to play the game. The dependent variable for the experiment was the difference in the participants' pre- and posttest scores.

The pretest and posttest were both embedded into the Element Solitaire program and administered to the users automatically at the appropriate times.

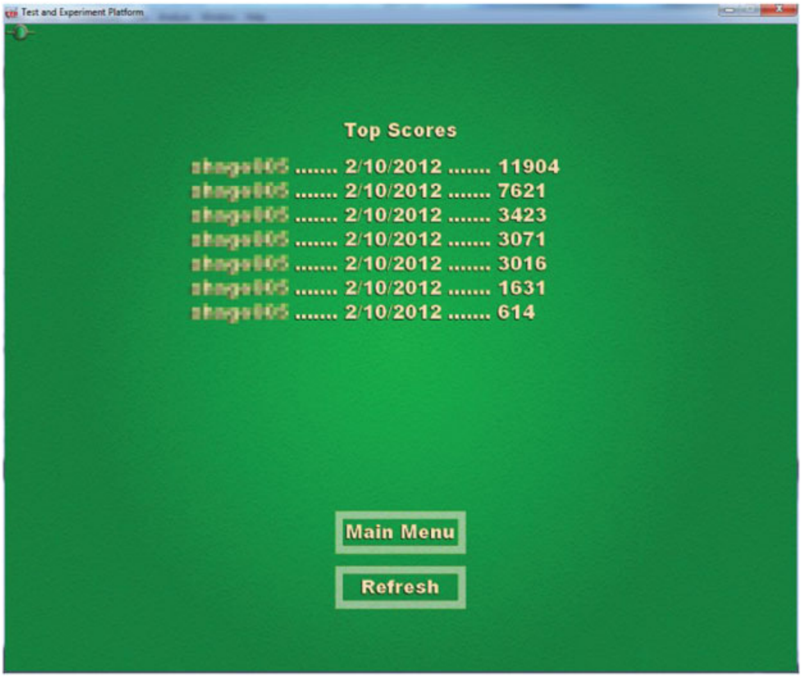

FIGURE 6 Scoreboard from a no-competition condition participant. Username is blurred to protect his or her identity.
 (Color figure available online.)

Figure 7 provides a screen capture of the tests, which were both identical. For each test, participants were given 20 minutes to fill in the element symbols in the provided periodic table. The symbols were entered via keyboard entry. As each symbol was entered, the corresponding symbol was automatically crossed out from the list provided at the bottom of the screen. In this manner, the test closely mimicked the learning content as presented through the game (where participants were given the cards with the symbols on them), as well

TABLE 1 Group-to-Condition Interactions Showing How the Three Hypothesis Conditions Were Translated Into Five Experimental Groups

Group\condition	Aesthetics	Choice	Competition (scoreboard)
Control group	—	—	—
Baseline	X	X	X
Diminished aesthetics	—	X	X
Diminished choices	X	—	X
No competition	X	X	—

Note: — = absence of condition; X = presence of condition

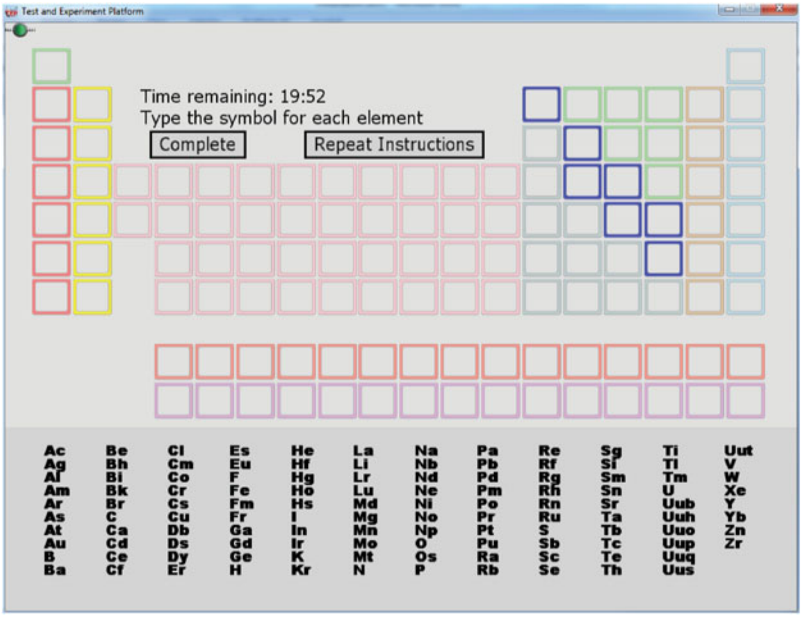

FIGURE 7 Pre- and posttest screen.
 (Color figure available online.)

as the testing artifacts of a pen-and-paper test (where the participants kept track of their entries by crossing out symbols).

There are 118 elements, and each correctly placed element is counted toward the participant's score. Incorrect or blank entries do not count. Therefore, the maximum possible score on an individual test is 118, and the minimum possible score is 0. Since the dependent variable is the difference between the pre- and posttest scores, the range of possible values for the dependent variable is −118 to 118. The score of −118 would represent a situation where a participant scored 118 on the pretest and 0 on the posttest, while 118 would represent a situation where a participant scored 0 on the pretest and 118 on the posttest.

METHODOLOGY

Participants

Research participants were drawn from a university chemistry department. Specifically, participants were all students in the CHEM 123, Foundations of Chemistry course. This course is described as a core requirement for science and engineering majors and is intended to prepare students for subsequent studies in molecular science (College of Sciences, 2012). Students taking this course are primarily freshmen or sophomores in their undergraduate studies and are typically between 18 and 23 years of age. This particular demographic was chosen to increase the probability that participants would be novices with regard to the learning content covered by the serious game interventions but would still have some level of interest in the subject matter.

The invitation to voluntarily participate in the program was given to two sections of this course, each taught by different instructors. The two classes had a total of 533 students. From these potential participants, 235 registered to participate in the experiment. One hundred seventy-two of the participants actually completed the experiment. It should be reiterated that the experiment was relatively intensive in terms of required effort. In order to complete the experiment, the participants had to take a pretest on the periodic table of elements, play the game a minimum of four times, and then take the posttest. The mean combined total time commitment was estimated at approximately one hour, over the course of one week.

The participation goal was to have 40 participants in each experimental group. The actual results fell slightly short of this goal. However, the number of samples in each treatment still provided a sufficient number of measures for statistical analysis. Table 2 shows the number of participants who completed the experiment, broken down by experimental group.

Procedures

For the execution of the experiment, the research participants were given a digital document (in PDF format) that provided a brief overview of the experiment and anticipated time requirements, and provided a link from which they could download the experiment program.

The experiment program, entitled "Testing and Experiment Platform (TEP)," was designed to encapsulate the experimental design and administration into a desktop application that participants could use at their own convenience. Providing this testing platform alleviated some of the practical barriers to student participation in a relatively time-consuming,

TABLE 2 Total Number of Participants Completing Experiment, Broken Down by Experimental Group

Completed samples per treatment	
Total	172
Control	38
Baseline	32
Degraded aesthetics	29
Degraded choices	36
No competition	37

voluntary experiment. The generic sounding name was also intended to obscure (primarily from the control group) the fact that some participants would be playing the Element Solitaire game while others would not.

When a participant first used the TEP program, he or she was prompted to enter a unique username. The participant was then given a pretest and asked to enter in the symbol names in the periodic table for as many of the 118 chemical elements as remembered. Once this was complete, the pretest score and username were stored on a network server.

After a participant was registered, the network server assigned him or her to a treatment group. Treatment assignments were incremented with each participant, ensuring that each treatment group never varied in initial population assignment by more than one.

Based on the treatment assigned by the server, the TEP client then provided participants with access to the variant of the Element Solitaire serious game intervention appropriate to their treatment group. The end result was intended to make the treatment assignments and differences transparent to participants. Whenever participants played the serious game, detailed records of their gameplay sessions were automatically reported to the server. These records included a date-time stamp for when a game was started, what kind of game was played, how long participants played the game, what score was received, whether they completed the game, what elements they missed, and how many hints or skips were used.

RESULTS

As described in the Methodology section, 172 of the original 235 participants completed the experiment. While the original apportionment of participants to experimental groups was tightly controlled, the number of participants who actually completed the experiment in each group varied, depending on individual participant motivation. Figure 8 shows the box plots of the differences between each participant's pretest and posttest score, as grouped by treatment. For these box plots, the darker box (red in online version of this article) represents the region of scores between the 25th and 50th percentile. The lighter box (green in online version of this article) represents the region between the 50th and 75th percentile, and the whiskers represent the respective 10th and 90th percentiles.

Initial inspection of the box plots indicated that the baseline treatment group had a relatively high positive difference in pretest and posttest scores in comparison to all other treatment groups, with the exception of the no-competition group. Even with the no-competition group, the median value in the baseline group was well below the median value for the baseline group, as evidenced in Table 3. However, the mean for the

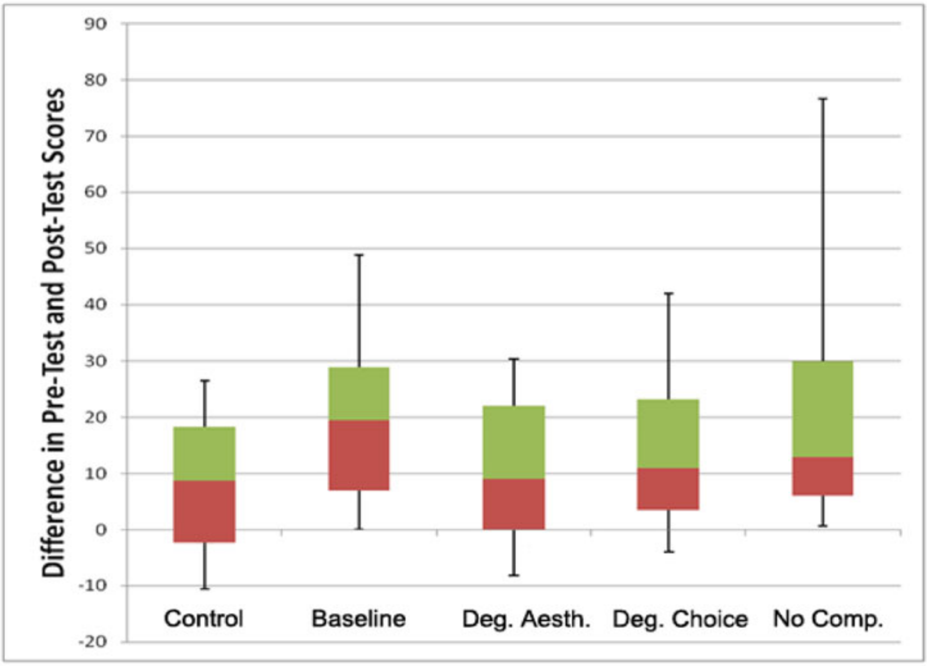

FIGURE 8 Box plot for the measures of learning outcome, grouped by treatment. (Color figure available online.)

no-competition group was actually higher than the mean for the baseline group.

To assess the statistical significance of the results, we selected a *t* test (Devore, 2004; Montgomery & Runger, 2003; Rumsey, 2003, 2009). The *t* test is more tolerant of smaller sample sizes than the more discriminating *z* test. One implication of this test selection is that the results are more conservative, in favor of the null hypothesis, which creates a greater burden of proof. Table 4 shows the *t*-test results for five comparisons of the treatment means. The means being compared are listed in the top row. The *p* values indicate the probability of finding the given observations if the null hypotheses

TABLE 3 Descriptive Statistics for the Differences in Pretest and Posttest for Each Experimental Group

	Control	Baseline	Degraded aesthetics	Degraded choice	No competition
Mean	7.92	20.09	8.90	16.67	23.08
Median	6.50	19.50	9	11	13
Standard deviation	23.94	21.40	26.96	25.33	30.14
Count	38	32	29	36	37

TABLE 4 *t* Test and *p* Value Results and Effect Sizes for Mean Comparisons Between Experimental Groups

	Baseline—deg. aesth.	Baseline—deg. choice	Baseline—no comp.	Baseline—control	Deg. aesth.—control
p value	0.0426	0.2749	0.3176	0.016	0.4394
Result	Significant reject H_0	Not significant cannot reject H_0	Not significant—cannot reject H_0	Significant—reject H_0	Not significant—cannot reject H_0
Cohen's *d*	0.674	0.217	−0.172	0.806	0.058
Effect size *R*	0.320	0.108	0.086	0.374	0.029

Note: H_0 = difference between treatments = 0; H_a = difference between treatments > 0. Critical α-value of 0.05.

are true. We selected a critical value (α) of 0.05 as the maximum *p* value acceptable for rejection of a null hypothesis.

The first hypothesis stated that the aesthetics of a serious game will have a positive effect upon the learning outcome derived from using a serious game. These tests indicated that there was, indeed, a statistically significant difference between the learning outcomes of the baseline and the degraded aesthetics experimental groups. In fact, the negative impact of the degraded aesthetics was so substantial that there was no statistical significance between the measured learning outcome of the degraded graphics experimental group and the control group, who did not play the game at all.

The second and third hypotheses stated, respectively, that the presence of meaningful choices and competition within a serious game will similarly have positive effects upon the learning outcome. The tests did not, however, indicate a statistically significant difference between the baseline group and the degraded choice experimental group or the no-competition experimental group. In the case of the no-competition group, cursory examination of the descriptive statistics for each group in Table 3, as well as their respective box plots in Figure 8, suggested that in some measures, the no-competition group might actually have performed better than the baseline group.

As a means of comparison, the test of the difference between the baseline group and the control group is included, and it does, indeed, indicate that the baseline group had a statistically significantly higher learning outcome than those who did not play the game.

As also evidenced by the box plots shown in Figure 8, there is a degree of overlap between each group. This is reflected in the Cohen's *d* and effect size *R* calculations. The effect sizes track fairly consistently with the *t*-test *p* values. The most significant changes in the learning outcomes had the greatest effect size and the least amount of overlap, with a Cohen's *d* value of 0.806 between the baseline and control groups.

DISCUSSION

Aesthetic Presentation

The first hypothesis was that the added aesthetic features, to include enhanced graphics and music, would result in higher learning outcomes than an identical game without music and graphics. The results of the *t* test comparing the baseline group and the degraded aesthetic experimental group did support this hypothesis. While we expected some indication of this outcome, the scope of the effect on the learning outcome, to the point where those playing with the degraded graphics did not perform statistically better than those who did not play the game, was surprising.

This surprise was amplified even more when we examined the number of average games played by each group. The initial supposition was that the high aesthetic treatment would have been more aesthetically pleasing and, hence, participants might have played the game more often. However, the data did not seem to support this notion. The mean for the number of games that the baseline group played and completed was 5.21 and 4.15, respectively (not all played games were completed). In comparison, the mean of played and completed games for the degraded aesthetic group was 6.23 and 4.98.

Caution should be taken when comparing these numbers as the experiment design created a significant confounding effect upon the number of games any participant played. Specifically, participants were told they had to play a minimum of four games. One could imagine scenarios in which this administrative imperative could either artificially inflate or depress the number of games that an individual participant might otherwise play. However, both experimental groups were subject to this same effect.

Bearing this caveat in mind, the difference still suggested that graphics not only improved the learning effect, but possibly made the learning more effective: Those with better graphics improved more with potentially fewer games. One possible explanation for this result might be a variation on the psychological "halo" effect (Nisbett & Wilson, 1977). The high-quality aesthetic presentation might have created a positive bias in the mind of users, prompting them to engage in a more effective and efficient manner with the learning material. In contrast, the low-quality aesthetic presentation may have caused users to dismiss the learning content out of hand.

Unfortunately for serious game developers, the practical implication of these experimental results is that enhanced graphics tend to cost more and take more time to develop. The entertainment software industry has been struggling with the burden of ballooning game development budgets, attributable, in large part, to the increased cost of creating high-quality assets (Develop, 2009; Graft, 2010; Takatsuki, 2007; Zelfden, 2009). The results of this research seem to indicate that serious game developers might have to

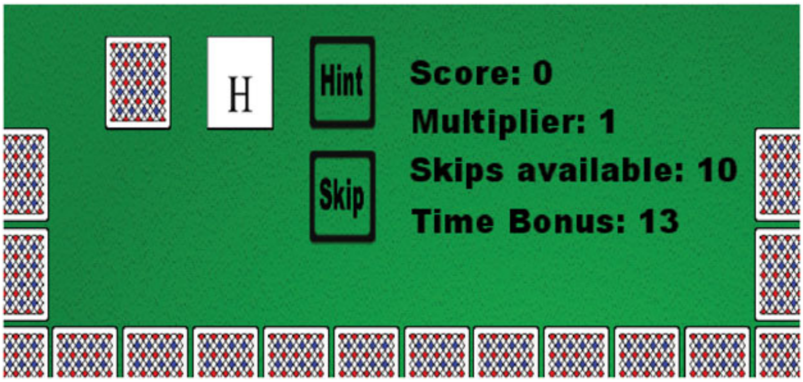

FIGURE 9 Placement of the hint and skip controls.
 (Color figure available online.)

be concerned with the same challenge of delivering high-quality aesthetic presentations to users as faced by the entertainment industry.

Choice

The second hypothesis was that meaningful choices in a serious game would result in higher learning outcomes than an identical game without meaningful choices. The statistical analysis of the degraded choice experimental group indicated that the treatment had no statistically significant effect upon the learning outcome and therefore the hypothesis was not supported. In trying to understand this outcome, we examined the records of the games played.

In the course of the experiment, 1,067 separate games were recorded; 789 of these games were played by participants who were not in the degraded choice treatment group. In those 789 games, only 3 total hints were used. In contrast, 349 skips were used.

Given the proximity and similarity of the hint user interface to the skip user interface, as shown in Figure 9, it appears unlikely that participants were unaware of the presence of the hint button. More likely, it seems that the hint button was never considered a viable or rewarding option to the player.

Damion Schubert (2008), a lead developer at Bioware Corporation, provided insight into the role of player choice in creating entertainment games, and provided guidelines for creating "meaningful choices." Shubert described how designers need to carefully evaluate the costs and benefits of the choices given to players, keeping in mind that the relative trade-offs depend heavily upon how players value resources associated with the game, to include their own time. In Element Solitaire, players suffer a point penalty for placing an element incorrectly on the table. The hint also deducts a point penalty, though this penalty is slightly smaller. However, the hint also takes more time, which the experiment participants may have valued more than points.

A possible reformulation of the current game might be to end the game when the player misplaces a specified number of elements. This version of player "death" could incentivize the use of the hint button, and induce players to consider more carefully their available options. It may be informative to compare this alternate game structure that places more emphasis on player choice with the existing, more forgiving game structure to further explore the role of choice in learning.

Competition

The third hypothesis was that a serious game with competition would result in higher learning outcomes than an identical one without competition. As with the second hypothesis, the analysis revealed no statistically significant difference between the no-competition and baseline experimental groups. Therefore, this hypothesis was also not supported by the experiment results. As previously discussed and demonstrated in Figure 7 and Table 3, however, the no-competition experiment group exhibited a much wider variance in performance than any of the other experimental groups, and its mean value is actually higher than the mean of the baseline group.

Possible explanations for this result might be found in educational research on the subject of competition. Educational literature has discussed proposed prerequisites for creating a learning environment conducive to constructive competition, as opposed to destructive competition (Burguillo, 2010; Sheridan & Williams, 2011; Williams & Sheridan, 2010). Constructive competition can enhance learner motivation and performance, while destructive competition can have the opposite effect. Without strict guidance and cues on the nature of the competition, the affective result upon the individual learner depends, in large part, upon how he or she internalizes the competition. Competitive personalities may be motivated by the competitive nature of the game, while noncompetitive personalities may become de-motivated. The two divergent responses to the same stimuli could create a potentially bi-modal response.

This may account for the wider variation seen in the no-competition experimental group measurements. While some participants might have perceived the leaderboard as a destructive competitive artifact, the no-competition experimental group was only shown a history of their own scores. In effect, they were only competing with themselves, which, in some cases, may have been perceived as more constructive.

Validity and Experiment Integrity

During administration of the experiment, it became apparent that efforts to create opaque walls between the treatment groups were not entirely effective. Within a few days of beginning the experiment, we received e-mails from concerned students asking why they did not get to "play the game." Quick examination of the server database revealed that all of these students were in the control group. We did not receive any e-mails from participants asking about the differences between experimental groups, but it would seem likely that control group participants were not unique in being exposed to the interventions assigned to other treatment groups.

The underlying intent for obscuring the different treatment groups from the participants was to maintain the integrity of the treatment groups in order to reduce the potential for treatment diffusion and compensatory rivalry effects between the respective subjects. While the effects of compensatory rivalry may be harder to judge, the high investment requirement in participating in the experiment, or even playing just one game, made the likelihood of treatment diffusion effects very low. Even if a participant were to find out about the other treatments, actually using these other treatments, due to the setup of the TEP program, would involve a significant amount of effort. It is possible that a participant might have played under someone else's profile, as well as his or her own, but the experiment records do not indicate these sorts of gameplay patterns. It seems unlikely that it occurred with sufficient frequency to have affected the experiment outcomes.

Lastly, there is also a concern that the convenience of the TEP program created the potential for participants to cheat while using the program by referencing some external aid. Looking at the pretest data, there are some anomalous results at the upper end of the spectrum, as shown in Figure 10. There is no evidence that these scores are the result of cheating, but regardless, they represent likely outliers in the distribution. It is presumable that someone who scores that highly on the pretest may not fit the desired participant target demographic descriptor of a "novice" in the target learning content.

In order to examine the effect that these potential outliers may have had, alternate t tests were conducted, excluding the data from those suspect participants. The results did not significantly change. The p value for the difference between the baseline and the degraded aesthetics group dropped slightly, from 0.043 to 0.016. The difference between the baseline and no-choice groups did drop from 0.28 to 0.11, but was still not below the critical value of 0.05. And lastly, the p value for the difference between the baseline and no-competition group increased slightly, from 0.32 to 0.35. None of the primary comparisons changed in their assessment of significance.

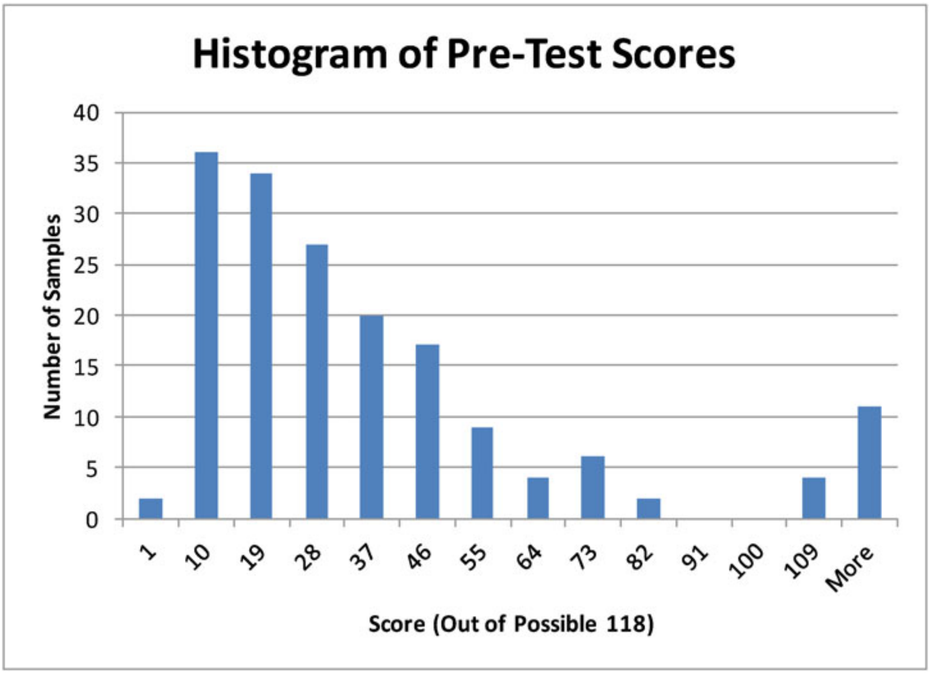

FIGURE 10 Histogram of pretest scores. Note the tail on the right side, consisting of 15 participants who scored higher than 100. (Color figure available online.)

CONCLUSIONS

The results of this experiment demonstrate that changes to the non-instructional aspects of a serious game can have significant effects upon the game's instructional value. This can even occur when the changes, from the strict perspective of learning content, appear to be subtle or even inconsequential. This may be simultaneously exciting and concerning.

It is exciting because it offers the potential to improve the delivery of instructional content. It offers the possibility that, through the use of techniques well honed in the commercial entertainment software industry, educators can craft effective serious games that can improve the quality of education for their students.

On the other hand, it is of concern because it adds an additional threshold of quality that must be achieved in order for the serious game to be effective. As previously discussed, the negative impact from a very simple degradation in aesthetic presentation made a startling impact upon learning outcomes. Furthermore, this additional hurdle to success is potentially independent of the quality of the learning content. For a serious game to be effective, it appears that the learning content has to be sound, as well as the aesthetic presentation of the game.

Upon further reflection, however, this should not be too much of a surprise to educators. There is a degree of subjective quality inherent in instructional delivery. It is not too much of a stretch to imagine how this might be true of serious games as well. One particularly applicable advantage of serious games may lay in the ability to quite literally codify that subjective quality and be able to perfectly reproduce it, on demand, to a potentially infinite number of learners.

REFERENCES

Burguillo, J. C. (2010). Using game theory and competition-based learning to stimulate student motivation and performance. *Computers & Education*, *55*(2), 566–575.

College of Sciences. (2012). *Old Dominion University Department of Chemistry and Biochemistry undergraduate course descriptions*. Retrieved from http://sci.odu.edu/chemistry/academics/under_courses.shtml

Crawford, C. (2003). *On game design*. Boston, MA: New Riders.

Develop. (2009). *Krome's Robert Walsh* [Interview]. Retrieved from http://www.develop-online.net/features/484/Interview-Kromes-Robert-Walsh

Devore, J. L. (2004). *Probability and statistics for engineering and the sciences* (6th ed.). Belmont, CA: Thomson-Brooks/Cole.

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

Graft, K. (2010). *EA: Budgets for games going in reverse direction*. Retrieved from http://www.gamasutra.com/view/news/30075/EA_Budgets_For_Games_Going_In_Reverse_Direction.php

Koster, R. (2005). *A theory of fun for game design*. Scottsdale, AZ: Paraglyph.

Martin, M., & Shen, Y. (2010a, October). *Defining and leveraging game characteristics for serious games*. Paper presented at the 2010 Virginia Modeling, Analysis and Simulation (MODSIM) Conference & Expo, Virginia Beach, VA.

Martin, M., & Shen, Y. (2010b, April). *Defining learning space in a serious game in terms of operative and resultant actions*. Paper presented at the Virginia Modeling, Analysis and Simulation Center 2011 Student Capstone Conference, Suffolk, VA.

Martin, M., & Shen, Y. (2010c). *Element Solitaire*. Norfolk, VA: Old Dominion University.

Martin, M. W., & Shen, Y. (2010, April). *Differentiating between serious games and computer aided instruction*. Paper presented at the Virginia Modeling, Analysis and Simulation Center 2010 Student Capstone Conference, Suffolk, VA. Retrieved from http://www.vmasc.odu.edu/downloads/Capstone_Papers/General_Science/Martin.pdf

Michigan State University. (2010, July 15). *MSU serious games design program*. Retrieved from http://seriousgames.msu.edu/

Montgomery, D. C., & Runger, G. C. (2003). *Applied statistics and probability for engineers* (3rd ed.). New York, NY: Wiley.

Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, *35*, 250–256. doi: 10.1037/0022-3514.35.4.250

Prensky, M. (2001). *Digital game-based learning*. New York, NY: McGraw-Hill.

Prensky, M. (2006). *"Don't bother me Mom, I'm learning!" : How computer and video games are preparing your kids for twenty-first century success and how you can help!* (1st ed.). St. Paul, MN: Paragon House.

Rumsey, D. J. (2003). *Statistics for dummies*. Hoboken, NJ: Wiley.

Rumsey, D. J. (2009). *Statistics II for dummies*. Hoboken, NJ: Wiley.

Salen, K., & Zimmerman, E. (2003). *Rules of play: Game design fundamentals*. Cambridge, MA: MIT Press.

Schell, J. (2008). *The art of game design: A book of lenses*. Boston, MA: Elsevier/Morgan Kaufmann.

Schubert, D. (2008, October). Designing choice. *Game Developer Magazine*, p. 2.

Shaffer, D. (2007). In praise of epistemology. In B. Shelton & D. A. Wiley (Eds.), *The design and use of simulation computer games in education* (pp. 7–27). Rotterdam, The Netherlands: Sense.

Shelton, B. E., & Wiley, D. A. (Eds.). (2007). *The design and use of simulation computer games in education* (Vol. 2). Rotterdam, The Netherlands: Sense.

Sheridan, S., & Williams, P. (2011). Developing individual goals, shared goals, and the goals of others: Dimensions of constructive competition in learning contexts. *Scandinavian Journal of Educational Research*, *55*, 145–164.

Takatsuki, Y. (2007). *Cost headache for game developers*. Retrieved from http://news.bbc.co.uk/2/hi/business/7151961.stm

Williams, P., & Sheridan, S. (2010). Conditions for collaborative learning and constructive competition in school. *Educational Research*, *52*, 335–350.

Yi, M. (2004). They got game. *San Francisco Chronicle*. Retrieved from http://www.sfgate.com/news/article/THEY-GOT-GAME-Stacks-of-new-releases-for-hungry-2663371.php

Zelfden, N. V. (2009). *What's killing the video-game business?* Retrieved from http://www.slate.com/articles/technology/gaming/2009/02/whats_killing_the_videogame_business.single.html

Zyda, M. (2005, September). From visual simulation to virtual reality to games. *Computer*, p. 8.