The Effect of Game Design on Game-Play Time and Learning Outcomes

Susan Gwee*[1], Ek Ming Tan*[2], Ahmed Hazyl Hilmy*[2]

*[1]English Language Institute of Singapore, Singapore,
*[2]Nanyang Technological University, Singapore

0299

The Asian Conference on Society, Education and Technology 2013

Official Conference Proceedings 2013

Abstract

In this paper, we investigate the effect of the design of a social studies mobile game on game-play time and learning outcomes of ninth-grade students. This social studies game-based curriculum involved 41 and 36 ninth-grade students in the first and second intervention studies using a mobile game played on Apple iPhones- *Statecraft X*-to engage in governorship practices in the game world and in the classroom. After the first intervention, modifications in game design were made to promote more collaborative learning, and to encourage students to use diplomatic strategies in the game. Survey data concerning social cohesion, diplomacy, power, agency, collaboration, and game-play time were collected from the intervention at the end of each intervention. Students from both intervention and comparison groups were also asked to write an essay that was subsequently assessed by the criteria of relevance, perspective, and voice. The results indicate that as a result of the change in game design, students spent more time playing the game. Students also showed positive shifts towards collaborative learning. There were significant effects of game design on social cohesion. There were no significant differences in the quality of student work between students in the first and second intervention studies. However, the differences in quality of student work between the second intervention and control groups were more significant than those between the first intervention and comparison groups. This paper will discuss the importance of adjusting the game design of serious games to enhance educational outcomes in the school curriculum.

INTRODUCTION

Scholars, educators, and media designers are interested to know how digital games might contribute to learning in the school context. Although digital games have been held up for their educative purposes (e.g., Gee, 2003; Prensky, 2001; Squire, 2006), there has been little empirical evidence that digital games result in better academic learning outcomes. This study is an attempt to fill the gap. The purpose of this paper is to examine the effect of the design of a digital game on gameplay time and learning outcomes in ninth graders.

Games and Learning

Games are used to help learning for main reasons: (a) motivation, (b) content mastery, as well as higher order thinking skills and (c) social skills (Jan, 2013). In schools, games are commonly used to address the challenge of motivating students to engage in learning. Indeed, prior research has supported the efficacy of games for enhancing and continuing student motivation (e.g., Malouf, 1988; McDonald and Hannafin, 2003).

However, it may be asked if students are learning while they are having fun. A prevalent concern is the extent to which game-playing enhances content mastery. In this regard, results are mixed. For instance, in a review of research published from 1984 to 1991 on the instructional effectiveness of the use of games in classrooms compared to instruction in conventional classroom, Randel, Morris, Wetzel, and Whitehill (1992) found that there were no differences in the post-tests between experimental and control groups. Similarly, McDonald and Hannafin (2003) found that there were no significant differences between third graders using a web-based computer game for reviewing social studies material compared to their counterparts using more traditional methods. However, Randel, Morris, Wetzel, and Whitehill (1992) found that their subjects were able to retain their material over a longer time period, with gains in learning favoring the experimental group at a delayed post-test.

Besides content mastery, researchers have identified critical skills for the twenty-first century such as problem solving, collaboration across networks and leading by influence, plan formulation and execution, adaptability, communication, and curiosity and imagination (Federation of American Scientists, 2006; Wagner, 2010). In addition to the acquisition of subject knowledge that digital games can support, parents appreciate the acquisition of other types of knowledge such as computer literacy, logical thinking, and hand-eye coordination, and perseverance (MacFarlane, Sparrowhawk, and Heald, 2002). They believe these are skills that their children will develop as a result of playing computer games.

The kinds of games that have gained greatest currency in schools are those that allow for students to undergo "drill and practice" exercises while having fun at the same time. Concomitantly, there is a significant body of research that has investigated the benefits in improving content mastery and motivation, as suggested above. This is to be expected, given that these are dominant concerns of educators in schools. However, using games to promote twenty-first century learning in school has been far more challenging.

In a review of 48 articles on the effectiveness of instructional games, Hays (2005)

found that an instructional game will only be effective if it is designed to meet specific instructional objectives and used as it was intended. Jan (2013) noted that games designed for twenty-first century learning are no longer the same games when they are used in the classroom. For instance, players revert to being students and are not allowed to play the game in their own playing styles. Moreover, instead of using information in the games designed for problem-solving, they tend to be encouraged to memorize them as they typically do as students. Echoing Hay's (2005) point for alignment between game design and use, Jan (2013) suggests that game design principles can be used to turn the classroom itself into a game for twenty-first century learning.

In this study, we investigate how changes in a mobile game designed for the learning of governance, situated in a game-like classroom and online environment where all participants, including the teacher engage in fictional role-playing, affect game-play time and the learning of twenty-first century skills. Specifically, we examine the effect of game design on participants' written argumentation, a key twenty-first century skill, in terms of their ability to advocate relevant solutions to the problems in ways that reflect multiple perspectives and their personal voice. We also investigate the effect of game design on participants' values and attitudes towards collaboration, social cohesion, agency, and power.

Statecraft X and Learning

The present study based on *Statecraft X,* a collaborative mobile game, supports a constructivist learning environment. Collaborative work can provide students the learning environment to leverage on their strengths, learning styles, skills, preferences, and perspective. In addition, working collaboratively also provides students the opportunity to participate in a community of practice, learning from each other through apprenticeship as a governor and education (Lave and Wenger, 1991).

Statecraft X also allows students to learn by experiencing for themselves and discovering their own meanings from their experiences. It allows them to play the game, reflect on the situation to form strategies, and then further investigate what effect their action has. It thus reflects Kolb's experiential learning cycle (Kolb, 1984). In addition, the learning materials supporting *Statecraft X* provide the link from the game world to the real world.

Hypotheses

We hypothesized that game design has an effect on game-play time and learning outcomes. We examined (a) game-play time, (b) quality of student work in terms of relevance, perspective, and personal voice, (c) values such as diplomacy, agency, social cohesion, power, and collaboration in ninth-grade students in order to better understand how game design influences these factors.

METHOD

Participants

Forty-one students (17 boys and 24 girls) participated in the first intervention study. One student was absent for the post-intervention survey administered to both intervention and comparison classes. Thirty-six students (14 boys and 22 girls) participated in the second intervention study. Two students from the second intervention study were absent in the collection of post-intervention survey and writing task data. The first group of students was in top class of the Express academic track while the second group of students belonged to the lower end of the Express academic track in the same school. On average, they were 14 to 15 years of age. Three social studies teachers participated in our study. One of them participated in both interventions. The school requested that one teacher was changed for the second intervention study so that more teachers in the school could experience the pedagogy of game-based learning.

For the first comparison group, there were 42 students (27 boys and 15 girls); for the second, 39 students (28 boys and 11 girls). However, one boy in the second comparison group completed the survey data but did not write the essay.

Materials

Apple iPhones with the installed *Statecraft X* game were loaned to all students who took part in the *Statecraft X* curriculum for the duration of the research intervention. *Statecraft X* was designed based on principles of governance in the ninth-grade social studies curriculum.

Teams competed against one another in this multiplayer strategy game to rule the fantasy kingdom of Velar populated by the four races of sentient beings. At the beginning of the game, the ruler of Velar passed away without leaving an heir, thus setting up the stage for different student governor-led political factions to compete for leadership of the kingdom.

The first game objective is that all the teams must collaborate to ensure that their kingdom, Velar, survive in the face of attackers from other kingdoms. Second, individual teams must consolidate their power and position by winning the trust of the people in their own towns and also the people in the towns of other teams. This game aims to allow students to think as governors and thus appreciate the complexity of the task of nation-building. To realise these two objectives, faction members must realise short-terms goals such as developing towns under their control, diffusing internal and external threats as well as maintaining diplomatic ties with factions within Velar and with neighboring kingdoms.

The game-play timings were set to accommodate the wishes of the school management. On weekdays, students could log in from 6 a.m. to 8 a.m., and from 2 p.m. to 10 p.m. On Saturdays, students could log in from 6 a.m. to 11 p.m.

In addition to the game world of Velar, students were also given materials from the fictional world of Bellalonia in which students were situated. The teacher, who

role-played the Grand Sage of Bellalonia, was provided with questions to elicit students' opinions about different aspects of governance. During the first lesson, students were given their final assignment where they, as fictional governors, had to solve problems in Bellalonia, a country formed a hundred years ago and populated by the ethnic group of Solians. Fifty years ago, the ethnic group, the Milous, immigrated from a neighbouring country in search of a better life. The Milous were hardworking and prospered in Bellalonia. The Solians, unhappy because they felt that they were entitled to the riches arising from the land, emigrated from Bellalonia. With the death of the old king, the subjects of Bellalonia thought that the monarchy could be changed to a democracy. The Grand Sage of Bellalonia had to choose governors in Bellalonia to form a council to help the young king who was not of age. The sage sent them to the game world of Velar where they faced challenges of governance. The following were the problems that the fictional governors had to solve: (1) high tax rate, (2) high unemployment rate, (3) high emigration, (4) low economic growth due to lack of resources and money, (5) political instability due to the death of the old king, (6) health epidemic of tuberculosis and malaria, (7) lack of education for new jobs, and (8) racial tensions between the majority and minority ethnic groups.

A web-portal was also set up to provide a space for students to be informed of events happening both in the game and fictional worlds. It was situated in Bellalonia so that students could role play as governors and offer opinions in a non-threatening context. Additional materials from the real world were provided to help students consider experiences from real world countries. These provided students with additional perspectives of governance beyond the game and the textbook.

The above materials were tied to the Play-between-Worlds curriculum model (see Figure 1). Students learn by "moving" from one world to another, and also by reflecting on their experience in the three worlds: game, fictional, and real.
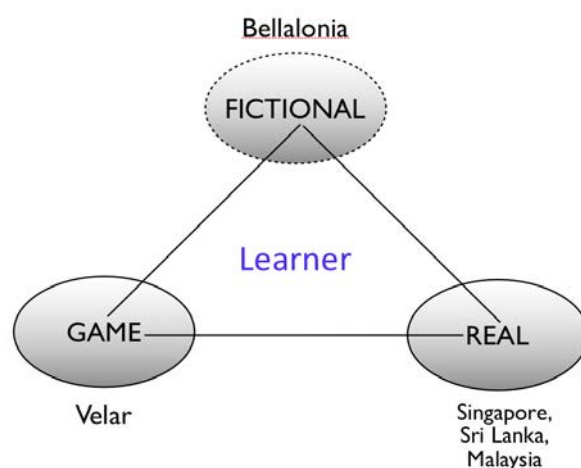

Figure 1. A Play-between-Worlds curriculum model.

The primary difference between the game design of the first and second interventions was the positioning of the game at the beginning of the lesson. In the first intervention, the winning team was positioned as the team that captured the capital city of Velar before the invasion of Salfreda. However, in the second intervention, the team that won was the team that had the highest composite score of average happiness of citizens in a faction, economic score (profit), and population levels in the

towns under their charge. The third author posted score charts on the *Statecraft X* website.

Another difference was that in the first intervention, the team members in each team did not share a pot of money given to the team; each team member managed their own funds, whereas in the second intervention, team members had to share a pot of money to encourage collaboration among team members as they had to discuss strategies in using the limited pool of money.

The third change in game design was the encouragement of the use of diplomacy instead of force to take over towns. Students were given the opportunity to organize a rally to win over the towns' citizens.

Instruments

We designed Survey 1 to examine the game-play time and collaboration perceptions of students in the two intervention classes. We designed the 18 survey items in Survey 2 for four constructs: (a) valuing social cohesion; (b) allowing diplomacy over unthinking use of force, (c) recognizing that government power is inextricably bound to responsibility, and (4) possessing agency, as a citizen, to influence a nation's present and future. We coded the students' responses with the numbers 1 to 6 corresponding to the responses 'Strongly Disagree', 'Disagree', 'Somewhat Disagree', 'Somewhat Agree', 'Agree', and 'Strongly Agree'. We designed a writing task for the intervention students who were graded on relevance of content, perspective, and personal voice.

Procedure

Prior to each research intervention, the teachers participated in a two-day professional development workshop designed to prepare them for the enactment of the *Statecraft X* curriculum. They were given the *Statecraft X* game to play and were shown the *Statecraft X* curriculum. They gave feedback on the lesson plans designed and worked with the research team to finalize the in-class and outside-classroom activities of the learning programme.

The class was divided into two groups for game-play as well as whole class discussions. Each teacher was in charge of one group. During the first session, the two groups of students were together in one computer laboratory. During the subsequent five sessions, the two groups were located in two separate computer laboratories.

All lessons were video-recorded. In the lead teacher's classroom, one video camera recorded the general classroom. One video camera was focused on three groups in the lead teacher's classroom. In the second teacher's classroom, one video camera recorded classroom activities.

Lesson 1

During the first of six lessons for each intervention, the research team administered a pre-test before the teacher explained to the students the Play-between-Worlds

curriculum model using the powerpoint slides provide by the research team. She also explained the history of Bellalonia and the current problems in Bellalonia. She gave out the final assignment of the *Statecraft X* programme which was a written speech that assesses the learning of principles of governance based on their in-class and outside-classroom experiences in the game, fictional, and real worlds. During the latter part of the first lesson, the game designer of *Statecraft X* game also presented the backstory of the game world of Velar and showed students the various actions that they could take in the game. At the end of the first lesson, iPhones were distributed to all participating students.

Lessons 2 to 5 in First Intervention

During the next four sessions of the first intervention, activities were conducted to complement the game. Teachers facilitated discussions about game-play so that students could draw lessons from their game-play. They encouraged students to draw upon the game and real world to propose solutions for the fictional world of Bellalonia.

Lessons 2 to 6 in Second Intervention

During the next four lessons for the second intervention, for the first thirty minutes of the lesson, students typed their individual responses to a series of questions related to Velar, Bellalonia, and the real world. In the last thirty minutes, the teacher gathered students in a circle at the front of the class and discussed their responses to these questions.

Lessons 6

During the final lesson, five students from each group presented their speeches in front of their groups. The teacher conducted a final discussion and the research team then administered a post-intervention survey (Survey 1) with questions on game-play time and collaboration.

Post Lesson 6

After the last lesson in each intervention, the first author administered a survey (Survey 2) and an essay in both interventions. Students were given thirty minutes to respond to the 18-item survey and write an essay.

Data Analysis

The data sources were (a) post-intervention Survey 1 data from intervention groups, (b) post-intervention Survey 2 data and essays from both intervention and comparison groups. For Survey 1, the survey questions on collaboration, and game-play time were compared between first and second intervention groups using *ANOVA*.

For Survey 2 data and the essays, the *ANOVA* was used to compare results between the first intervention and comparison groups, and between the second intervention and comparison groups. Table 1 summarizes the data sources, dependent measures, and analysis strategies used to answer the two research questions of this paper.

Table 1
Alignment of Research Questions, Data Sources, Dependent Measures, and Analysis Strategies

Research question	Data source	Dependent measure	Analysis strategy
Does game design have an effect on game-play time?	Survey 1	Time spent per week in game-play	*ANOVA*
Does game design have an effect on learning outcomes?	Written speech	Relevance, perspective, and voice scores	*ANOVA*
	Survey 2 on values	Scores on the constructs of social cohesion, power, diplomacy, and agency	*ANOVA*
	Survey 1	Collaboration	*ANOVA*

Essay Scores

The first and second authors assessed (a) relevance, (b) perspective, and (c) personal voice (see rubric in Appendix 1). Relevance refers to how relevant the policies proposed by a student are to the social and economic needs of the different segments of the country's population and whether this student has given examples from both traditional and non-traditional sources to support his or her proposed policies. Perspective refers to whether a student could give multiple perspectives to the proposed policies and integrate them or whether he could only give the textbook perspective. Personal voice refers to the voice used by a student and whether it matched the situation, how authentic the voice was, whether opinions were well-defined and detailed, whether she communicated strong feelings and honest statements, and whether she showed that she cared about the topic.

The first and second authors assessed each essay separately and awarded a mark for each criterion. They both hold graduate degrees and had at least eight years of teaching experience in Singapore schools, and were part of the *Statecraft X* research team. The first author also had three years of experience in an improving teachers' assessment literacy research project where she trained teachers to assess student work based on a scoring guide and exemplars of student work during assessment workshops. She also acted as an adjucator during score resolution sessions if two teams of teachers gave different scores to the same student work. After having assessed all essays separately, they came together to moderate the marks for each criterion in each essay. When there was a discrepancy between the mark given by the first and third authors, they compared the features of the essay with the benchmark performance given in the scoring guide and discussed why the student should be

awarded a certain score. They considered any evidence that challenged the original scores and achieved a consensus score. They then assigned this consensus score for each criterion.

The results of Johnson et al.'s (2005) study suggest that when scores differ between two raters, discussion as a core resolution method is the best method compared to the averaging of two scores. Johnson et al. (2005) reported that for the use of an analytic rubric for grading essays, the scores arrived at after discussion between two raters were closer to expert-criterion scores than averaged scores between the two raters. To calculate exact and adjacent agreement between the first and second authors, we transformed the students' scores to the level scores as indicated in the analytic rubric, i.e., scores of 1-5, 6-10, 11-15, and 16-20 were transformed to level scores 1, 2, 3, and 4 respectively.

RESULTS

Game-Play Time

In the first intervention, the mean game-play time was 12.09 hours per week and the mean game-play time was highest at home (see Table 2). In the second intervention, students reported spending significantly more time playing the game in all the spaces than in the first intervention. The assumption of homogeneity of variance is violated; therefore, the Welch F-ratio is reported. There was a significant main effect of the type of intervention on total game-play time $F(1,56.37) = 8.30$, $p = .006$; and game-play time in the spaces of home, $F(1,74) = 4.23$, $p = .043$; school, $F(1,42.09) = 7.07$, $p = .048$; mall, $F(1,40.37) = 10.41$, $p = .002$; and walking $F(1,56.15) = 4.07$, $p = .048$.

Table 2

Mean Hours Spent per Week across Spaces Students in First and Second Interventions

	Intervention			
	First *(n=40)*		Second *(n=36)*	
Space	*M*	*SD*	*M*	*SD*
Home	7.64	8.34	12.46	11.92
School	1.74	2.19	4.78	6.52
Bus	1.14	2.19	1.73	2.20
Walking	0.59	1.18	1.35	1.96
Car	0.47	0.90	1.08	2.38
Eating out	0.37	0.70	.94	1.23
Mall	0.14	0.41	.92	1.40
Total	12.09	12.32	23.37	20.37

Quality of Student Work

Table 3 shows the distribution of scores given to students in the first and second intervention groups by the two raters. The exact and adjacent agreement rates between Raters A and B for the first intervention study were 93%, 100%, and 78% for relevance, perspective, and voice, respectively, while the exact agreement rates were 59%, 54%, and 37%. For the second intervention study, the exact and adjacent agreement rates between Raters A and B were 85%, 97%, and 94% for relevance, perspective and voice, respectively whereas the exact agreement rates were 47%, 65%, and 41%, for relevance, perspective, and voice, respectively.

Table 4 summarizes the results of the study for the variables relevance, perspective, and voice in the first intervention and comparison groups Levene's test of homogeneity of variance showed that the variances were not homogeneous for relevance of content and perspective. Thus, the Welch test was used for these two dependent variables. There was no significant main effect of type of intervention on relevance of content.

Table 3
Comparison of Frequency of Scores of Raters A and B across Intervention

Score	Intervention 1		Intervention 2	
	Rater A	Rater B	Rater A	Rater B
Relevance				
Level 1	7	5	1	3
Level 2	14	12	2	9
Level 3	14	18	23	22
Level 4	6	6	8	0
Perspective				
Level 1	4	6	0	0
Level 2	11	8	6	5
Level 3	22	19	21	28
Level 4	4	8	7	1
Voice				
Level 1	2	2	0	0
Level 2	11	7	4	6
Level 3	14	19	12	21
Level 4	14	13	18	7

Table 4
Summary of Means, Standard Deviations, 95% Confidence Intervals for Scores on
Relevance, Perspective, and Voice in First Intervention and Comparison Groups

	Intervention 1 (n = 41)				Comparison 1 (n = 42)			
			95%CI				95%CI	
Variable	M	SD	LL	UP	M	SD	LL	UP
Relevance	11.07	4.20	9.75	12.40	5.55	2.73	4.70	6.40
Perspective	11.46	3.89	10.23	12.69	6.02	2.91	5.12	6.93
Voice	14.44	3.14	13.45	15.43	8.40	3.29	7.38	9.43

Note. M = mean; SD = standard deviation; CI = confidence interval; LL = lower limit; UL = upper limit.

Table 6 summarizes the results of the study for the variables relevance, perspective, and voice for the second intervention and comparison groups. The means of all the variables are at least two times higher in the second intervention group compared to the second comparison group.

A further analysis of the data revealed that the mean differences between the second intervention and comparison groups in relevance, perspective, and voice were highly significant at $p < 0.001$ (see Table 7). Thus, the quality of students' essays in the second intervention group was significantly higher with respect to each criterion of assessment: relevance, perspective, and voice. The effect sizes of the variables of relevance, perspective, and voice were also very large at 0.598, 0.661, and 0.431, respectively. That is, the differences between students in the second intervention and comparison groups were significantly large.

Table 5

Summary of the *ANOVA* Analysis of Quality of Student Work between First Intervention and Comparison Groups

Variable		Sum of Squares	df	Mean Square	F	p	η^2
Relevance	Between Groups	633.44	1	633.44	50.74	< .001*	.385
	Within Groups	1011.19	81	12.48			
	Total	1644.63	82				
Perspective	Between Groups	613.89	1	613.89	52.17	< .001*	.392
	Within Groups	953.17	81	11.77			
	Total	1567.06	82				
Voice	Between Groups	755.45	1	755.45	73.00	< .001*	.474
	Within Groups	838.22	81	10.35			
	Total	1593.66	82				

Note. df = degree of freedom; η^2 = eta squared or effect size.

As can be seen from Tables 5 and 7, the effect sizes for two of the dependent writing measures, relevance of content and perspective of the second intervention and comparison groups were bigger than those found in the first intervention and comparison groups.

Table 6

Summary of Means, Standard Deviations, 95% Confidence Intervals for Scores on Relevance, Perspective, and Voice in Second Intervention and Comparison Groups

Variable	Intervention 2 (n = 34)		95%CI		Comparison 2 (n =38)		95%CI	
	M	SD	LL	UP	M	SD	LL	UP
Relevance	12.09	2.82	11.00	13.07	4.38	3.38	3.29	5.48
Perspective	12.32	2.20	11.56	13.09	4.44	3.38	3.34	5.53
Voice	14.00	2.98	12.96	15.04	6.46	4.94	4.86	8.06

Note. M = mean; SD = standard deviation; CI = confidence interval; LL = lower limit; UL = upper limit.

Table 7
Summary of the *ANOVA* Analysis of Quality of Student Work between Second Intervention and Comparison Groups

Variable		Sum of Squares	df	Mean Square	F	p	η^2
Relevance	Between Groups	1077.98	1	1077.98	101.12	< 0.001*	0.598
	Within Groups	695.97	71	9.80			
	Total	1773.95	72				
Perspective	Between Groups	1130.09	1	1130.09	132.78	< 0.001*	0.661
	Within Groups	593.03	71	8.35			
	Total	1723.12	72				
Voice	Between Groups	1032.25	1	1032.25	51.43	< 0.001*	0.431
	Within Groups	1219.69	1	17.18			
	Total	2251.95	72				

Note. *df* = degree of freedom; η^2 = eta squared or effect size.

Collaborative Learning

Table 8 shows the response of students to the post-intervention survey questions regarding collaborative learning. They responded to the likert scale of 1 to 6 (1 = Strongly Disagree, 6 = Strongly Agree). The single significant main effect of intervention was the survey question "Teamwork helps me think better" although the mean of all responses showed positive shifts towards collaborative learning (see Table 9).

Table 8
Summary of Means, Standard Deviations, 95% Confidence Intervals for Survey Questions on Collaborative Learning in First and Second Intervention Groups

Survey question	First Intervention (n = 41)		95%CI		Second Intervention (n =36)		95%CI	
	M	*SD*	*LL*	*UP*	*M*	*SD*	*LL*	*UP*
CL1: I enjoy learning more, on my own than in a team.	3.83	1.20	3.44	4.21	3.64	1.33	3.19	4.09
CL2: I know how to work with others.	5.05	.55	4.87	5.23	5.08	.77	4.82	5.34
CL3: Teamwork helps to make learning more interesting.	5.05	.71	4.82	5.28	5.22	.64	5.01	5.44
CL4: Teamwork makes me think better.	4.70	.94	4.40	5.00	5.11	.67	4.89	5.34
CL5: Collaboration is good for learning.	4.98	.62	4.78	5.17	5.17	.56	4.98	5.36
CL6: Discussions with my team mates in the *Statecraft* curriculum is useful for my learning.	4.83	.81	4.57	5.09	4.92	.69	4.68	5.15
CL7: Group work is enjoyable in the *Statecraft X* curriculum.	5.12	.72	4.89	5.36	5.13	.72	5.02	5.30
CL8: I learn more playing Statecraft game with friends than	4.93	.66	4.72	5.13	5.00	.63	4.79	5.21

playing the game alone.

Note. M = mean; *SD* = standard deviation; *CI* = confidence interval; *LL* = lower limit; *UL* = upper limit.

Table 9
Summary of the Welch Analysis between First and Second Intervention Groups for Attitudes towards Collaborative Learning

Survey question	Welch Statistic	df1	df2	p
CL1	.407	1	70.727	.526
CL2	.046	1	62.835	.831
CL3	1.234	1	73.996	.270
CL4	4.914	1	70.330	.030*
CL5	2.004	1	73.997	.161
CL6	.282	1	73.782	.597
CL7	.282	1	73.782	.597
CL8	.257	1	73.634	.614

Note. df = degree of freedom.

Social Cohesion

The results of Survey 2 showed no statistical differences for all dependent measures of social cohesion, agency, diplomacy, and power for the first intervention. In the second intervention, the results (see Table 10) showed significant differences between the intervention and comparison groups for two dependent measures: social cohesion ($F = 19.28, p < 0.001, \eta2 = 0.214$) and agency ($F = 5.45, p < 0.022, \eta2 = 0.071$).

Table 10
Comparison of Results of Specific Survey Questions on Social Cohesion and Agency between the Second Intervention and Comparison Groups

Survey questions on social cohesion and agency	F	p	η^2
SC1. I feel that trust between races is important.	20.14	<0.001	0.221
SC2. I think that it is important for people of different races to live together happily.	10.99	0.001	0.134
SC3. I want to work well with people from different races.	10.88	0.002	0.133
SC4. I find it easy to work with people from different races.	10.47	0.002	0.129
SC5. A multicultural society is preferable to one with a single dominant culture.	5.14	0.026	0.068
A1. The government will listen to citizens who keep complaining	5.59	.02	.073

DISCUSSION

In the second intervention, changes in game design led to more game-play time for both boys and girls. This shows that students in the second intervention were more

engaged in game-based learning than students in the first intervention. The significant increase in game time game indicates that students showed much more interest in game-based learning in the second intervention compared to the first.

Although students in the second intervention also achieved slightly better outcomes in relevance of content and perspective, they achieved slightly lower outcomes in voice. It could be that the increase in game-play time allowed them to experience governorship for a longer time and that might have contributed to better learning outcomes.

The change in game design requiring collaboration among team members had a positive effect on attitudes towards collaboration. The students in the second intervention felt that teamwork helped them to think better.

CONCLUSION

We have shown that game design has an effect on the learning of twenty-first competencies. We have argued that a better-designed game resulted in students being more engaged, resulting in better outcomes in terms of argumentation, collaboration and social cohesion. Specifically, designing games which require more collaboration and strategizing and prioritizing, all twenty-first century competencies as they involve higher-order thinking skills and social skills, help bring about the development of such skills among students.

REFERENCES

Gee JP (2003) *What Video Games Have to Teach Us about Learning and Literacy*. New York, NY: Palgrave Macmillan.

Hays RT (2005) *The effectiveness of instructional games: A literature review and discussion*. Technical Report No. 2005-004, Naval Warfare Center Training Division, Orlanda, FL.

Federation of American Scientists (2006) *Harnessing the Power of Video Games for Learning*. Available at: http://agamerseducation.wordpress. com/2005/10/25/harnessing-the-power-of-video- games-for-learning.

Jan M (2013) A literature review of game-based learning. *Singteach, 45*. Available at: http://singteach.nie.edu.sg.

Johnson RL, Penny J, Gordon B, Shumate SR and Fisher SP (2005) Resolving score differences in the rating of writing samples: Does discussion improve the accuracy of scores? *Language Assessment Quarterly* 2(2): 117-146.

MacFarlane A, Sparrowhead A and Heald Y (2002) Report on the educational use of games: An exploration by TEEM of the contribution video games can make to the education process. Available at: http://www.teem.org.uk/publications/teem_gamesin_full.pdf.

Malouf DB (1988) The effect of instructional computer games on continuing student motivation. *The Journal of Special Education* 21(4): 27-38.

McDonald KK and Hannafin RD (2003) Using web-based computer games to meet the demands of today's high-stakes testing: A mixed method inquiry. *Journal of Research on Technology in Education* 35(4): 459-472.

Prensky M (2001) *Digital Game-Base Learning*. New York, NY: McGraw-Hill.

Randel JM, Morris BA, Wetzel CD and Whitehill BV (1992) The effectiveness of games for educational purposes: A review of recent research. *Simulation and Gaming* 23(3): 261-76.

Squire K (2006) From content to context: Videogames as designed experience. *Educational Researcher* 35(8): 19-29.

Wagner T (2010) *Change Leadership: A Practical Guide to Transforming Our Schools*. New York, NY: Basic Books.

Appendix 1
Rubric for Scoring Relevance, Perspective, and Voice in Students' Essays

	Level 1	Level 2	Level 3	Level 4
Criteria	1-5 marks	6-10 marks	11-15 marks	16-20 marks
Relevance	Policies proposed are not relevant to the social and economic needs of the country. Most examples given are simplistic. Do not diverge from the textbook.	Examples given may include a non-textbook source or an innovative interpretation. Examples given may include a non-textbook source or an innovative interpretation.	Policies proposed meet the social and economic needs of the majority of the population. Examples given are appropriate and include non-traditional sources.	Policies proposed meet all the social and economic needs of this country. Examples given effectively support all the policies proposed. Examines examples given for its relevance.
Perspective	Offers only the textbook perspective.	Limited discussion of perspectives other than the textbook perspective. Alternatives are not integrated. Treats other viewpoints superficially.	Offers multiple perspectives, but they are integrated in a limited way Attempts to investigate viewpoints.	Integrates diverse multiple relevant perspectives. Multiple viewpoints are thoroughly discussed, explained and qualified.
Personal voice	Is indifferent towards the topic. Does not communicate feelings. Does not offer any opinion. Writing is phony, stilted or awkward. The reader is clueless about the personality of the writer. Voice used is inappropriate for the situation.	Cares about the topic in a limited way. Communicates feelings as an afterthought. Opinions are emergent in nature. Major inconsistencies cast doubt on the authenticity of the piece. The reader has to examine the piece carefully for an indication of the writer's personality. Voice used matches the situation at times.	Cares about the topic. Communicates feelings. Opinions are outlined. A few inconsistencies in the piece. The reader gets a glimpse of the writer's personality. Voice used largely matches the situation.	Cares deeply about the topic. Communicates strong feelings and honest statements. Only the writer could have written it. Opinions are well-defined and detailed. Writing is authentic. The writer's voice is consistent throughout the essay. The writing sounds real. Displays a well-developed personality. The reader has the impression that he is getting to know the writer very well. Voice used matches the situation very well.

Appendix 2
Sample of Student Essay in Second Intervention

A Sample of High Quality Work

Imagine that you are running for an election to be a member of parliament and that you have to formulate policies to convince the citizens of your country that you are the best candidate. Justify your proposed policies by using examples from what you have learnt, what you have read, and your personal experiences.

I think that I am the best candidate because I always listen to the citizens' needs. I will do the best that I can to keep them happy and will also care about their well-being. This is the most important part of being a member of parliament. Without happiness, citizens will most likely rebel against the government and this may lead to work strikes. This will indirectly lead to low economic growth and thus falling into poverty. Without a good medical care service, citizens may be vulnerable to serious illnesses such as bird flu or SARS. As the saying goes, 'Prevention is better than cure'. By living a healthy lifestyle, very few citizens will be sick. I can play my part in the subsidies of medicine and hospital fees so that all will get good treatments from doctors and nurses with having to pay just 50% of the price.

I have learnt that by keeping the country safe from any external attacks such as war, we need to have strong defence. By having a strong defence, no country would want to or will think twice before declaring war. This could also make the citizens happy which is more important because they would feel safe by the defence that the country provides. I will try my best to make a strong defence and also commercialise this so that the citizens will know that the country would be safe from any attacks either by all-out war or by terrorist attacks.

I think that these 3 factors could make the country progress and a prosperous country is the outcome if these factors are being carried out efficiently.