CJLT RCAT
Canadian Journal of Learning and Technology
La Revue canadienne de l'apprentissage et de la technologie

Volume 42(3) Summer/été 2016

Playing and Learning: An iPad Game Development & Implementation Case Study

Jouer et apprendre : une étude de cas du développement et de la mise en œuvre d'un jeu sur iPad

***Jennifer Jenson**, York University*
***Suzanne de Castell**, Simon Fraser University*
***Rachel Muehrer**, Ryerson University*
***Erin McLaughlin-Jenkins**, York University*

Abstract

There is a great deal of enthusiasm for the use of games in formal educational contexts; however, there is a notable and problematic lack of studies that make use of replicable study designs to empirically link games to learning (Young, et al., 2012). This paper documents the iterative design and development of an educationally focused game, Compareware in Flash and for the iPad. We also report on a corresponding pilot study of 146 Grades 1 and 2 students playing the game, a paper and pencil related activity and completing a pre- and post-test. The paper outlines preliminary findings from the play testing, which included high levels of student engagement, an approaching statistical improvement from pre- to post-test, and a discussion of the improvements that needed to be made to the game following the pilot study.

Résumé

L'utilisation du jeu dans les contextes éducatifs officiels suscite beaucoup d'enthousiasme. Cependant, le manque d'études qui utilisent des modèles pouvant être répétés pour relier les jeux et l'apprentissage de manière empirique est remarquable et problématique (Young et coll., 2012). Cet article documente la conception et le développement itératifs d'un jeu aux accents éducatifs, Compareware, en Flash et pour l'iPad. Nous traitons également d'une étude pilote correspondante dans le cadre de laquelle 146 élèves de 1re et 2e année ont joué au jeu, réalisé une activité connexe à l'aide de crayons et de papier et passé des tests avant et après. L'article résume les conclusions préliminaires des essais du jeu, y compris des taux élevés d'engagement des élèves, l'amélioration statistique entre les tests avant et après le jeu, ainsi qu'une discussion des améliorations à faire au jeu après l'étude pilote.

Introduction

This paper documents the design, development, user testing, and pilot study of Compareware, an educational game designed for the iPad IOS operating system and for internet browsers in Flash. Compareware is playfully named after the popular WarioWare franchise and, like WarioWare, is a series of quick minigames that are played in succession. Using clear and intuitive visual design, Compareware asks its players to examine two pictures that are set side by side and choose vocabulary that indicates similarities and differences. For example, how are a tiger and a zebra different and how are they the same? Our intent in designing the game was to create an iPad experience that could be used in elementary classrooms; that was first and foremost intended for educational ends; and that supported a fundamental attribute associated with higher order 'metacognitive' thinking skills, namely, the ability to ascertain and articulate conceptual and semantic similarities and differences between objects.

Compareware grew out of a 3-year long multiliteracies study in which we noted anecdotally and through field notes and observations that participants often had difficulty in articulating how something was similar – either because they could not describe functional similarity (i.e., two very different pairs of shoes are similar because they protect the feet and/or are used for walking and/or are used to run) or because they had difficulty relating the degree or way in which two objects were similar (i.e., they are both for walking but one is for winter and the other for summer). While the degree of similarity seemed to be less difficult to articulate than the category or quality of the similarity (color, shape, size, form, function), those aged 5-8 still displayed difficulties mobilizing both the vocabulary and, so far as could be linguistically evidenced, the analytical skills necessary to articulate how two objects could be described as similar. Based on this preliminary work, Compareware was an attempt to see if we could design a game to scaffold learners who are less linguistically fluent to express—and to extend and develop—their understanding of artifact classification in terms of similarities and differences.

In the next section, we briefly review some of the recent literature on the use of games in education and on the categorization of artifacts. Our intention is to show how Compareware and the pilot study design fit broadly within current research initiatives. The sections that follow briefly describe the game's iterative design process and detail our study's methodology and some preliminary results.

Games and Learning: Forging Connections

Many have weighed in on the potential of games as sites of and for learning (Gee, 2003, 2005; Prensky, 2001; Squire, 2011); however, much of the early work is more polemical than empirical. In a recent critique of the literature on digital games and education, which includes examining studies of games that were built for educational purposes as well as commercial off-the-shelf games (COTS) mobilized for educational ends, Young, et al. (2012) quip: "After initial analyses, we determined that, to date, there is limited evidence to suggest how educational games can be used to solve the problems inherent in the structure of traditional K-12 schooling and academia. Indeed, if you are looking for data to support that argument, then we are sorry, but your princess is in another castle" (p. 62). Their argument is that educational research needs better methodologies for studying games, including the use of software to track player behavior in games and provide documentation of individual play styles and characteristics. In response,

Tobias and Fletcher (2012) argue that Young et al. had not examined "transfer" in games – that is, how a player might transfer a cognitive ability acquired in game to one outside the game (c.f. Anderson & Bavelier, 2011; Green & Bavelier, 2003). Tobias and Fletcher also reiterate an important consideration made in an earlier paper (Tobias, et al., 2011), that it is difficult to map the field when it is changing so rapidly. As a remedy, they suggest developing a taxonomy for games that will allow for increased clarity in analysis and discussion.

What these meta-reviews and others (e.g., Fletcher & Tobias, 2006; Ke, 2009; Sitzmann, 2011) point to is an ongoing problem in studies of game-based learning (GBL): The fact that despite theoretical claims, quite often it is not clear if games are pedagogically effective learning tools. Some studies have found very little in terms of learning from playing games (Ke, 2008; Papastergiou, 2009; Tsai, Yu, & Hsiao, 2012), while others suggest that games can be effective sites for learning (Barab, et al., 2009; Fletcher & Tobias, 2012; Hsu & Wang, 2010). For the purpose of this paper, we would like to emphasize the importance of acknowledging that the field is still emerging, as are its methods for evaluation and its salient research questions. We therefore situate this work as an educational game, designed in-house, and, as we detail in the next section, for a very particular purpose. This is in line with other GBL projects that are designed, developed, and tested with particular learning objectives in mind, including, for example, the development of a road safety game (All, et al., 2013), a game for health education (Liberman, 2001), a game about saving electricity (Tsai, Yu, & Hsiao, 2012), and a game to encourage empathy (Bachen, Hernández-Ramos, & Raphael, 2012).

While we play-tested the game in multiple schools with students aged 6-8, our questions did not focus on the benefits of using iPads as a mode of delivery for an educational game. Instead, we situated our questions for this paper within the GBL framework, asking 1) what, if anything, do students learn from playing Compareware; 2) what might be some effective means of measuring that; and 3) how might students' reading abilities affect their interaction with the vocabulary focus of the game? These questions were meant to inform the redesign process and help determine the appropriate grade levels for implementing the game. Before turning to the design of the game and the methods used in the pilot study, we also situate this work within the literature on artifact categorization (similarities and differences) as a pedagogical construct.

Artifact Categorization: A Brief Overview

Object categorization, as Bornstein and Arterberry (2010) argue, "conveys knowledge of other object properties as well as knowledge of properties of category members not yet encountered. In brief, categorizing is an essential cognitive and developmental achievement, but also presents a formidable cognitive and developmental challenge" (p. 351). The robust literature on artifact categorization in children and adults most typically divides that intellectual effort between a child's apprehension of physical similarities (shape, size, color) and its function, arguing that the latter is a kind of deeper understanding than the former (Bloom 1996; 2000). However, this research has been, for the most part, contradictory. For example, studies of children as young as 5 have shown that children attribute labels of physical similarities to objects at the expense of functional similarities (Graham, Williams, & Huber, 1999; Landau, Smith, & Jones, 1998; Merriman, Scott, & Marazita, 1993; Smith, Jones, & Landau, 1996), while other studies with children as young as 2 found the opposite: Functional similarity is prioritized over physical similarity (Deak, Ray, & Pick, 2002; Diesendruck, Markson, & Bloom, 2003). That

said, most research tends to show that preschool children are more likely to base their categorization on physical appearance rather than on function (Gentner & Rattermann, 1991; Woodward & Markman, 1998). In an overview of some of the methodological inconsistencies that may have produced these very different outcomes, Diesendruck, Hammer, and Catz (2003) claim that in their study "when functional and appearance information about artifacts are simultaneously available to children for the same length of time, through the same medium, and without adult direction, children weigh these two respects equally and highly" (p. 229). For our purposes, this is significant as the game we designed does not need adult direction, keeps players in the same medium, and the game contained images and text that supported both physical and functional artifact categorization.

What is clear is that there are a number of confounding factors that have yet to be resolved with respect to categorization. And, though the general consensus on whether young children are more likely to prioritize physical dimensions over an artifact's function is that "it depends," it is the case that more studies have concluded that the physical can have more weight than function (Kelmer Nelson, Frankenfield, Morris, & Blair, 2000). The Compareware study does not attempt to replicate the methods used in past studies of artifact categorization. Instead, it is interested in whether and how a game-like environment might support artifact categorization in young children without adult intervention.

Compareware: Design and Process

The title of the game plays on a title of the Nintendo DS game franchise WarioWare in which players create their own minigames through a series of visual programming choices made possible through the game's interface. Compareware invites players to compare two objects of increasing difficulty and in later levels under time constraints. The game takes place in an environment that is graphically very bright and is divided into six thematic areas: school, home, ocean, grocery, town and outdoors (see Figures 1-4 below). Players enter the game and are presented with two objects and asked "How are they the same?" in one instance and "How are they different?" in another. The images are randomly assigned and a set of six answers scrolls through the bottom of the screen, which the players must drag to the appropriate spot between the two images. The answers are in text and can be read out to players if they so choose, supporting those who might not yet read. There are also multiple levels in the game, with progress being marked by advancing to unlockable content as players win levels, a design feature that was chosen to make it more like a commercial game. Players also receive instant feedback on whether or not they have chosen the correct answer, and are only penalized by the game restarting if they appear to be randomly dragging and dropping answers.

Figure 1. Title screen of the game Compareware.

Figure 2. Home screen.

Figure 3. Corner Store vs. Diner, in level 1 of "Town."

Figure 4. Pie vs. Bread in level 1 of "Grocery."

Figure 5. Cruise ship vs. cargo ship in level 2 of "Ocean"

As players progress, their answers are recorded, and they are awarded one to five stars depending on the number of correct answers in a given series. Correct and incorrect answers are tracked in the game for each unique user, allowing us to track which set of images and which particular vocabulary are most often incorrectly chosen in each area of the game. We also attempted to include both physical and functional similarities as the literature on differences and similarities tends to track both; however, due to technical limitations we were unable to track whether and how players were more or less successful between the two categories. Players receive feedback from the game based on whether or not they select correct or incorrect answers. Correct answers, as indicated above, receive a star and a voice over which says “congratulations” and incorrect answers are indicated with the word chosen sliding back down to the bottom of the screen with a “bonk” noise to indicate that they are incorrect.

Compareware was designed in 4 months, with rapid prototyping of 3 playable levels that were designed and play-tested within the first 6 weeks of the project. Following the first round of play-testing, voice-over sound was added for all vocabulary present in the game; time constraints were removed in the early levels all together; and we created a way for users to turn both sound and time constraints on or off. Following initial play-testing and user feedback, we also altered the graphical interface for the drag and drop vocabulary in order to stylistically “match” the associated area of the game – e.g., in the ocean section, the drag and drop phrase or words are in a fish (see Figure 5) while in the grocery section they are conveyed in a shopping basket (see Figure 4).

Debugging the game was extensive and time consuming, taking 2 months post-development in its first iteration, then another 6 weeks after an initial play-testing session as a number of expected glitches were discovered when multiple users played the game. In addition, it became clear that we needed to rephrase some of the questions and answers: Some answers to the questions had to be adjusted so that the phrasing was consistent and some questions had to be

rewritten because their connection to the pictures was unclear. Additionally, some pictures had to be replaced so that they worked better with the concept. For example, the original picture for the bathroom was simply an open source image of a bathtub; this was changed to include a wider view of a recognizable bathroom.

As we have attempted to demonstrate in the discussion of the design of Compareware, we began with a theoretical framework that was developed into a concept for a game, which was then iteratively designed for a specific target audience, children aged 5-8. That iterative design was informed by the literature on similarities and differences, as well as design for game-based learning (Gros, 2007; Hirumi, Appelman, Rieber, & Van Eck, 2010; Papastergiou, 2009). In particular, we sought to create an environment that was both fun and engaging to play, that also potentially had a learning outcome that was measurable. In the next section, we shift the focus from the design process to the play-based study we conducted. Our primary question was how and if the game supports participants' learning related to the articulation of similarities and differences.

Methods

The purpose of this study was to document whether, how, and under what circumstances students learned to perform correct artifact categorization after playing Compareware in three different modalities: 1) in an iOS platform (iPad); 2) in Flash (on a PC in a computer lab); 3) through a paper and pencil activity that used images and text from the game. The first two were game-based and the third was a more traditional classroom activity. Students were also given a pre-test and a post-test that made use of the images from the game and asked them to categorize those images for similarities and differences. Every participant experienced each of the modalities, albeit in a different order due to constraints in booking time in computer labs. While our original intent was to examine participants' experience of the modalities separately, it was soon clear that each modality afforded its own strengths and limitations[1]. In total, 4 schools and 9 classrooms (5 Grade 1, 6 Grade 2) participated. Because of the variation in class size and those who opted out of the study, each classroom had between 18 and 25 participants aged 6–8, for a total of 146 participants. Students' reading abilities ranged from kindergarten to Grade 3 reading levels.

Using a mixed-methods approach, we collected qualitative data through audio-video recordings of students playing Compareware and through field notes in the classroom activity.

---

[1] The focus of our analysis for this paper is holistic as participants experienced each of the modalities, albeit in a different order. The iPad afforded two key strengths: 1) it allowed students to work individually, without adult support and 2) for those who chose to invoke the sound feature, they could listen without headphones. One limitation of using the iPads was that it was impossible to provide a unique login, making data retrieval nearly impossible. We had to pull all data from the iPads after each use. Playing Compareware in a computer lab was limited by the fact that 1) not all of the computers worked, which meant students had to sometimes share a machine, 2) to access the sound support students had to use headphones and not all computers headphone jacks worked, and 3) we were unable to retrieve scores as they were stored locally and we were unable to get permission from the school board to retrieve the local cache. The paper and pencil activity was enthusiastically completed by almost everyone, though it did mean that there was considerable adult (teacher and researcher) intervention to help with vocabulary.

Quantitative data was collected in three forms: Teachers provided us with a list of students' reading levels, and students filled out a questionnaire and responded to tests before and after playing the game. The questionnaire was on media and videogame experiences and habits. The pre-test and post-test asked students to write about similarities and differences based on images and vocabulary from the game. While they were identical in content, on the post-test we changed the order of the questions in order to try to control for students' remembering their answers from the pre-test.

Study participants were recruited by classroom. The project's principal investigators contacted school principals, who in turn found teachers at their schools who were willing to participate. Consent forms were sent out to each of the teachers' entire class. A few students in each of the classes opted out (n=7), but all interested students participated during regularly scheduled class time (n=146).

All participants completed the following tasks during 4-40 minute sessions: 1) time on the iPad to experiment with a pre-loaded application; 2) playing Compareware on the iPad; 3) completing a pen and paper activity based on the Compareware game; and 4) playing Compareware on the computer (in Flash). In the first session, each group took the pre-test and was assigned to one of the four activities. In the second, third, and fourth sessions the students completed each of the other three activities. On the final day, students also completed the post-test, which was identical to the pre-test. Because of limited computer lab availability Group 4, (Table 1) was only able to participate in three of the four activities. The order of activities for each group was as follows:

- Group 1: 1) Free time on iPad; 2) CW on iPad; 3) Pen and Paper 4) CW on Computer
- Group 2: 1) CW on Computer; 2) Free time on iPad; 3) CW on iPad; 4) Pen and Paper
- Group 3: 1) Pen and Paper; 2) CW on Computer; 3) Free time on iPad; 4) CW on iPad
- Group 4: 1) Free time on iPad; 2) CW on iPad; 3) Pen and Paper

Activities were ordered in this way for two reasons, one that was driven by a design question and the other that was simply expedient. In the first case, we were interested in how participants engaged with the game on the iPad versus the computer lab, and in the second, we simply needed time in between groups to save the player's games both on the iPad and in the computer lab so we could later analyze their questions.

Based on the current literature and general consensus regarding children's classification of objects, it was reasonable to hypothesize that students would have difficulty identifying similarities between objects before they began playing Compareware. We also hypothesized that students at all reading levels would improve their ability to identify both similarities and differences between objects after playing the game, and we hoped that weaker readers would use the feature in the game that read the words aloud to them. In the end, nearly all students had the sound on during the play periods and we noted that would repeat the vocabulary from the game as they played as a means of interacting with one another.

Findings

On the pre-test, students had almost the same scores naming similarities (a mean of 3.8 out of 8) and differences (mean of 3.7 of 8), an outcome consistent with the findings of Diesendruck, et al. (2003). Comparing pre- to post-test scores revealed that 55% of the students increased their scores after participating in all of the activities; however this finding was not statistically significant. That there was that degree of improvement is still rather surprising given that they played Compareware for, at the very most, 70 minutes over two days—a generous estimation given the time taken to begin and to conclude the activities. A large percentage of participants (36.7%) lowered their scores on the post-test, an effect that could have been caused by test fatigue given there were only 3 days between the tests. This effect could also have been produced by clearer instructions given to teachers on the post-test to allow students to answer what they could without coaching them to select the correct answer, something that we observed happening more frequently on the pre-test. There were no mean differences between groups.

Score Distribution

In terms of the whole sample, the post-test showed a good distribution in scores (see Table 1) ranging from 2 to 14 out of 16, with an average score of 8, which indicates that the Compareware tasks were set at an appropriate difficulty level for the participants.

Table 1

Frequency table of post-test total score distribution among participants

		Frequency	Percent	Cumulative Percent
Valid	2.00	4	3.6	3.6
	3.00	4	3.6	7.2
	4.00	6	5.4	12.6
	5.00	7	7.2	19.8
	6.00	7	7.2	27.0
	7.00	12	10.8	37.8
	8.00	19	17.1	55.0
	9.00	9	8.1	63.1
	10.00	20	18.0	81.1
	11.00	12	10.8	91.9
	12.00	3	2.7	94.6
	13.00	5	4.5	99.1
	14.00	1	.9	100.0
	Total	111	100	
Missing		38		
Total		149		

Pen and Paper Activity

Although, we were not able to collect in-game metrics in the pilot, as explained above, the pen and paper activity provided a detailed catalogue of the questions students answered correctly and incorrectly. By observing students fill out the worksheet, we were able to see where and why students misinterpreted the questions. Some questions were unclear either because the pictures that we presented for comparison did not sufficiently represent the target similarity/difference or there was some ambiguity in the way we phrased the question. At other times, misinterpretation was the result of students' reading difficulties. The pen and paper activity also proved valuable in the redesign of the game because participants could work collaboratively and at their own pace; students often vocalized their thought processes as they worked out the answers together. For example, one of the questions had a picture of a polar bear and a black bear. Students could indicate whether the characteristic "bear" was a similarity or a difference by circling their choice. One student reasoned that a polar bear and a black bear are the same because they are both bears. Another came to the opposite conclusion, circling bear as a difference because they are different kinds of bears. This process very quickly shed light on the way students experienced the game, and we were able to flag questions that might be confusing and needed revision for the final iteration.

Score Improvement by Reading Level

We compared students' reading levels with their pre- to post-test score improvement so that we might determine which readers benefitted the most from playing the game and participating in the pencil and paper activity. We first collected students' Developmental Reading Assessment Levels and Guided Reading Levels provided by their teachers. These reading levels ranged from C (Grade 1) to level O (Grade 3) with 13 levels in total (see Table 2). For purposes of analysis, we created four larger groups and labeled them as "Low" (C-F), "Medium" (G-J), "High" (K-L), and "Very High" (M-O). See Table 3 for the distribution of each of these groups.

Table 2

Participant's Reading Levels

		Frequency	Percent	Cumulative Percent
Valid	C	4	3.5	3.5
	D	1	.9	4.4
	E	6	5.3	9.7
	F	5	4.4	14.2
	G	2	1.8	15.9
	H	11	9.7	25.7
	I	11	9.7	35.4
	J	15	13.3	48.7
	K	12	10.6	59.3
	L	25	22.1	81.4
	M	17	15.0	96.5
	N	2	1.8	98.2
	O	2	1.8	100.0
	Total	113	100.0	
Missing		36		
Total		149		

Table 3

Reading Levels Regrouped for Analysis

		Frequency	Percent	Cumulative Percent
Valid	Low reading level	16	14.2	14.2
	Medium reading level	39	34.5	48.7
	High reading level	37	32.7	81.4
	Very high level	21	18.6	100.0
	Total	113	100.0	
Missing		36		
Total		149		

We ran a one-way ANOVA to compare the change in pre- to post-test scores between the 4 reading groups. Because of the small n values, the ANOVA came out non-significant, so we ran 3 independent sample t-tests to compare the "Very High" reading group with each of the other groups. The results of this test were the following: the comparison of the "Low" to "Very High" group mean change in score was not significant. However, the comparison of the "Medium" to "Very High" groups revealed a significantly higher mean change in score from pre- to post-test in the "Very High" group compared to the "Medium" group, with values of $t(39) = -$

2.09 and $p = .043$. Finally, participants in the "Very High" group had a significantly higher mean change in score from the pre- to post-test than the "High," with the values of $t(38) = -2.88$ and $p = .007$.

Discussion

The pilot study was invaluable in strengthening the study design and streamlining the game so that students were encouraged to continue to play at more challenging levels. After questions and in-game vocabulary were revised to minimize confusion, we realized that students needed more direction in order to navigate through the game. Initially students had been given a choice of a variety of topics from a home menu, but there had been no indication of where to start, how many levels were in each topic, or how many questions they had remaining. In order to give players a clearer idea of the structure of the game, we added a series of progress bars and screens with detailed directions, and we locked the hardest level so that players were required to successfully complete most of the game before they could move on to the most challenging questions. Finally, we found that students were simply performing the motions of play by dragging and dropping answers randomly rather than attempting to correctly answer the question; this occurred most often with the iPad. For example, we observed some, but not all students simply dragging the words as they scrolled along the bottom of the screen one at a time up to the answer area. While this is certainly a very good game strategy in that it meant that they were simply maximizing on the rules and mechanics of the game (drag and drop, no penalties) we wanted to encourage them to be more selective in their answers. Therefore, we added a feature that would insert a pop-up message encouraging players to try a new answer after a student had made three attempts to drag and submit the same wrong answer.

These findings do suggest that the comprehension and articulation of similarities and differences is linked to reading ability and that the game is most appropriate for and most beneficial to students with high Grade 2 to Grade 3 reading levels. Given the available data, it is difficult to judge why there was less impact at the lower reading levels; however, we speculate that there simply was not enough time spent on the activity for some of the participants, and that it remains difficult to demonstrate "transference" in game-based learning studies (Young et al., 2012).

Additionally, despite our instructions to the teachers not to coach student answers, in the pre-test especially, students were assisted to answer the questions. This happened mainly out of what we took as a desire on the part of the participating teachers to help their students complete the pre-test but also because some of the students simply could not yet read, and needed to have the questions read to them in order to answer them. While this certainly biased the pre-test, we argue that classrooms are not petri dishes or labs, and these kinds of under takings are fraught with these kind of often not reported on occurrences.

Conclusion

An ongoing challenge for this project was working within the daily ebb and flow of an elementary school. While administrators, teachers and parents were excited, supportive and welcoming, it was surprisingly difficult to schedule 5 days in a row in multiple classrooms in the same school (which was necessary to achieve the requisite sample size for this study). We often

lacked the required communication with administrators and teachers to achieve a schedule that allowed the students time for their regular programming as well as the study. Often, unforeseen circumstances meant that a class would arrive late to begin the study or need to leave early. On one occasion, a fire alarm disrupted the study and we had to schedule a make-up play session. If teachers were absent, often the supply teacher was unaware of the schedule, or the students were off-task more than usual and therefore not as focused on completing the study as they had been previously. As is often the case, school technology was unpredictable: The school computers did not always work, there were missing headsets and the internet firewalls had to be removed at the same school on more than one occasion to allow access to the game. That is all to say that keeping exact times for set-up, play time, paperwork and movement between classrooms for each group of participants was rarely possible, and so variation between students' experiences with the study is to be expected.

The purpose of this paper has been to detail the design and implementation of an educational game with a large play-testing group of 146 participants who completed tasks with the game and without it (paper and pencil activity). The study identified how and what students might have learned through Compareware's playful activities, including the paper and pencil activity, which features were effective in advancing its educational purposes, and which features need to be changed before a full study can be carried out. Other promising findings included the improvement shown on the post-test by over half of our participants after only two very short play sessions. Most important for us was that we saw improvement in students' abilities to correctly identify both similarities and differences after only a very short period of play, unassisted by adults, and that we have preliminary indications of some ways in which students' reading levels predict their success with digital as well as traditional pencil and paper literacies. Finally, user-testing the game enabled us to clearly identify necessary modifications to improve its affordances for both learning and for research. This work makes another contribution to research on games and education and on the use of games in classroom settings. While the length of this paper does not permit us to adequately detail the real enthusiasm exhibited by the students and teachers who participated in our project, we do want to underscore that playing games, as other studies have shown (See for example, Boyle, et al., 2012), is one very real way to foster student engagement. Compareware was not designed for hours and hours of play but to be played in short segments well-suited to the time constraints of schools, which very much appealed to and was understood by the young twenty-first century learners who participated in the study.

Acknowledgements:

We would like to thank the teachers and students who enthusiastically participated in this study, the Post-Doctoral Fellow who helped to carry out the study, Dr. Nick Taylor and the statistical analysis support that we had from Barry Dilouya. We also gratefully acknowledge the support for research and game development funding from the GRAND-NCE Network.

References

Anderson, A. F., & Bavelier, D. (2011). Action game play as a tool to enhance perception, attention, and cognition. In S. Tobias & J. D. Fletcher (eds.), *Computer games and instruction* (pp. 307-330). Charlotte, NC: Information Age.

All, A., Nunez Castellar, E. P., & Looy, J. V. (2013). An evaluation of the added value of co-design in the development of an educational game for road safety. *International Journal of Game-based Learning, 3*(1), 1-17. doi:10.4018/ijgbl.2013010101

Bachen, C. M., Hernández-Ramos, P. F., & Raphael, C. (2012). Simulating REAL LIVES: Promoting global empathy and interest in learning through simulation games. *Simulation & Gaming, 43*(4), 437-460. doi:10.1177/1046878111432108

Barab, S. A., Scott, B., Siyahhan, S., Goldstone, R., Ingram-Goble, A., Zuiker, S. J., & Warren, S. (2009). Transformational play as a curricular scaffold: Using videogames to support science education. *Journal of Science Education and Technology, 18*(4), 305–320. doi:10.1007/s10956-009-9171-5

Bornstein, M. H., & Arterberry, M. E. (2010). The development of object categorization in young children: Hierarchical inclusiveness, age, perceptual attribute, and group versus individual analyses. *Developmental Psychology, 46*(2), 350-365. doi:10.1037/a0018411

Bloom, P. (1996). Intention, history, and artifact concepts. *Cognition, 60*(1), 1-29. doi:10.1016/0010-0277(95)00699-0

Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.

Boyle, E. A., Connolly, T. M., Hainey, T., & Boyle, J. M. (2012). Engagement in digital entertainment games: A systematic review. *Computers in Human Behavior, 28*(3), 771-780. doi:10.1016/j.chb.2011.11.020

Deák, G. O., Ray, S. D., & Pick, A. D. (2002). Matching and naming objects by shape or function: Age and context effects in preschool children. *Developmental Psychology, 38*(4), 503-518. doi:10.1037/0012-1649.38.4.503

Dickens, H., & Churches, A. (2011). *Apps for learning: 40 best iPad/iPod Touch/iPhone apps for high school classrooms.* In *The 21st Century Fluency Series.* Thousand Oaks, CA: Corwin Press.

Diesendruck, G., Hammer, R., & Catz O. (2003). Mapping the similarity space of children and adults' artifact categories. *Cognitive Development, 18*(2), 217-231. doi:10.1016/S0885-2014(03)00021-2

Diesendruck, G., Markson, L., & Bloom, P. (2003). Children's reliance on creator's intent in extending names for artifacts. *Psychological Science, 14*(2), 164-168. doi:10.1111/1467-9280.t01-1-01436

Fletcher, J. D., & Tobias, S. (2006). Using computer games and simulations for instruction: A research review. In *The Proceedings of the Society for Applied Learning Technology Meeting* (pp. 1-14). Orlando, FL: New Learning Technologies. Retrieved from https://www.researchgate.net/publication/228511464_Using_computer_games_and_simulations_for_instruction_A_research_review

Gee, J. P. (2003). *What videogames have to teach us about learning and literacy.* New York, NY: Palgrave Macmillan.

Gee, J. P. (2005). Good video games and good learning. *Phi Kappa Phi Forum, 85*(2), 33-37.

Gentner, D., & Ratterman, M. J. (1991). Language and the career of similarity. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on thought and language: Interrelations in development* (pp. 225–277). London, UK: Cambridge University Press.

Graham, S. A., Williams, L. D., & Huber, J. F. (1999). Preschoolers' and adults' reliance on object shape and object function for lexical extension. *Journal of Experimental Child Psychology, 74*(2), 128–151. doi:10.1006/jecp.1999.2514

Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, 423, 534-537. Retrieved from http://www.nature.com/nature/journal/v423/n6939/full/nature01647.html

Gros, B. (2007). Digital games in education: The design of game-based learning environments. *Journal of research on technology in education*, *40*(1), 23-39.

Henderson, S., & Yeow, J. (2012, January). *iPad in Education; A case study of iPad adoption and use in primary school*. Paper presented at the 45th Hawaii Conference on System Sciences, Honolulu, HI.

Hirumi, A., Appleman, B., Rieber, L., & Van Eck, R. (2010). Preparing instructional designers for game-based learning: Part 2. *TechTrends, 54*(4), 19-27. doi:10.1007/s11528-010-0416-1

Hsu, H., & Wang, S. (2010). Using gaming literacies to cultivate new literacies. *Simulation & Gaming, 41*(3), 400-417. doi:10.1177/1046878109355361

Kagohara, D. M., Sigafoos, J., Achmadi, D., O'Reilly, M., & Lancioni, G. (2012). Teaching children with autism spectrum disorders to check the spelling of words. *Research in Autism Spectrum Disorders, 6*(1), 304-310. doi:10.1016/j.rasd.2011.05.012

Ke, F. (2008). A case study of computer gaming for math: Engaged learning for gameplay? *Computers and Education, 51*(4), 1609-1620. doi:10.1016/j.compedu.2008.03.003

Ke, F. (2009). A qualitative meta-analysis of computer games as learning tools. In R. E. Ferdig (Ed.). *Handbook of research on effective electronic gaming in education* (Vol. 1, pp.1-32). Hershey, PA: Information Science Reference.

Kelmer Nelson, D. G., Frankenfield, A., Morris, C., & Blair, E. (2000). Young children's use of functional information to categorize artifacts: Three factors that matter. *Cognition, 77*(2), 133-168. doi:10.1016/S0010-0277(00)00097-4

Landau, B., Smith, L. B., & Jones, S. S. (1998). Object shape, object function, and object name. *Journal of Memory and Language, 38*(1), 1–27.

Liberman, D. (2001). Interactive video games for health promotion: Effects on knowledge, self-efficacy, social support and health. In R. Street, W. Gold, & T. Manning (Eds.), *Health promotion and interactive technology: Theoretical applications and future directions* (pp. 103–120). Mahwah, NJ: Lawrence Erlbaum Associates.

McClanhan, B. (2012). A breakthrough for Josh: How use of an iPad facilitated reading improvement. *TechTrends, 56*(3), 20-28. doi:10.1007/s11528-012-0572-6

Merriman, W. E., Scott, P., & Marazita, J. (1993). An appearance-function shift in children's object naming. *Journal of Child Language, 20*(1), 101–118. doi:10.1017/S0305000900009144

Papastergiou, M. (2009). Exploring the potential of computer and video games for health and physical education: A literature review. *Computers & Education, 53*(3), 603-622. doi:10.1016/j.compedu.2009.04.001

Peluso, D. C. C. (2012). The fast-paced iPad revolution: Can educators stay up to date and relevant about these ubiquitous devices? *British Journal of Educational Technology, 43*(4), E125-E127. doi:10.1111/j.1467-8535.2012.01310.x

Preciado Babb, A. P. (2012). Incorporating the iPad2 in the mathematics classroom: Extending the mind into the collective. *International Journal of Engineering Education, 2*(2), 23 - 29. Retrieved from http://online-journals.org/index.php/i-jep/article/view/2084

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

Sitzmann, T. (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology, 64*(2), 489–528. doi:10.1111/j.1744-6570.2011.01190.x

Smith, L. B., Jones, S. S., & Landau, B. (1996). Naming in young children: A dumb attentional mechanism? *Cognition, 60*(2), 143–171. doi:10.1016/0010-0277(96)00709-3

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

Tobias, S., Fletcher, J. D., Dai, D. Y., & Wind, A. P. (2011). Review of research on computer games. In S. Tobias & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 127–222). Charlotte, NC: Information Age.

Tobias, S., & Fletcher, J. (2012). Reflections on "A review of trends in serious gaming." *Review of Educational Research 82*(2), 233-237. doi:10.3102/0034654312450190

Tsai, F. H., Yu, K. C., & Hsiao, H. S. (2012). Exploring the factors influencing learning effectiveness in digital game-based learning. *Educational Technology & Society, 15*(3), 240–250. Retrieved from http://www.ifets.info/journals/15_3/18.pdf

Woodward, A. M., & Markman, E. M. (1998). Early word learning. In W. Damon, D. Kuhn, & R. Siegler (Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (pp. 371-420). New York, NY: Wiley.

Young, M., Slota, S. Cutter, A., Jalette, G., Mullin, G., Lai, B., Simeoni, Z., Tran, M., & Yukhymenko, M. (2012). Our princess is in another castle: A review of trends in serious gaming for education. *Review of Educational Research, 82*(1), 61-89. doi:10.3102/0034654312436980

Authors

Dr. Jennifer Jenson is Professor in the Faculty of Education and Director of the Institute for Research on Digital Learning at York University. She has published on technology and education, games and learning and gender and digital gameplay. Email: JJenson@edu.yorku.ca

Dr. Suzanne de Castell is Professor Emerita in the Faculty of Education at Simon Fraser University. Her long academic career has included publications on philosophy and education, gender and technology and games and learning, among many others. Email: decaste@sfu.ca

Dr. Rachel Muehrer is a Research Associate in the Play: CES Lab run by Dr. Jennifer Jenson at York University. She has published on games and learning and music education. Email: rachel.muehrer@gmail.com

Dr. Erin McLaughlin-Jenkins is a Research Associate with the Institute for Research on Digital Learning at York University. Her research examines the historical intersections between science, technology, class, and philosophy. She has published on education, autodidactism, evolutionary theory, socialism, and information technology. Email: erink@yorku.ca