Computers & Education 57 (2011) 2178–2195

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Exploring Newtonian mechanics in a conceptually-integrated digital game: Comparison of learning and affective outcomes for students in Taiwan and the United States

Douglas B. Clark [a,*], Brian C. Nelson [b], Hsin-Yi Chang [c], Mario Martinez-Garza [a], Kent Slack [b], Cynthia M. D'Angelo [d]

[a] Vanderbilt University, Tennessee, USA
[b] Arizona State University, Tempe, Arizona, USA
[c] National Kaohsiung Normal University, Kaohsiung, Taiwan
[d] University of Wisconsin, Wisconsin, United States

A R T I C L E   I N F O

Article history:
Received 14 October 2010
Received in revised form
9 May 2011
Accepted 10 May 2011

Keywords:
Games
Applications in subject areas
Cross-cultural projects
Interactive learning environments
Pedagogical issues
Secondary education

A B S T R A C T

This study investigates the potential of a digital game that overlays popular game-play mechanics with formal physics representations and terminology to support explicit learning and exploration of Newtonian mechanics. The analysis compares test data, survey data, and observational data collected during implementations in Taiwan and the United States with students in grades 7–9. Results demonstrate learning on some core disciplinary measures and high levels of learner engagement, indicating the potential benefits of this genre of conceptually-integrated games, but also suggesting that further research and development will be needed to more fully harness this potential. Encouragingly, striking similarities were observed across the two countries in terms of learning and engagement, suggesting that this genre of learning games may prove suitable for engaging students in active exploration of core science concepts across multiple countries.

1. Introduction

International and national science standards call for inquiry learning and a focus on depth of understanding, but science instruction in many countries often focuses more on breadth than depth and involves minimal opportunity for inquiry-related learning (e.g., Mullis, Martin, & Foy, 2008; Newton, Driver, & Osborne, 1999; Roth et al., 2006). Furthermore, school-based science curricula tend to center on explicit formalized knowledge structures, seldom connecting this knowledge with students' tacit intuitive understandings. Digital games potentially provide a medium for addressing these challenges in an engaging inquiry-based manner for students (Gee, 2003, 2004, 2007; Klopfer, Osterweil, & Salen, 2009). A growing body of research and scholarship on games and cognition emphasizes cycles of prediction, observation, and refinement as core mechanics of game play processes (e.g., de Freitas & Neumann, 2009; Salen & Zimmerman, 2003; Squire et al., 2003; Wright, 2006). Prediction and observation are focal inquiry processes in science education (e.g., Champagne, Klopfer, & Gunstone, 1982; Tao & Gunstone, 1999), and thus digital games might support students in exploring challenging core science concepts in a context that also emphasizes key inquiry processes rather than the decontextualized rote learning central to much traditional science instruction. Prior studies have suggested, however, that learning in digital games is often more tacit than explicit (e.g., Squire, Barnett, Grant, Higginbotham, 2004). Furthermore, there may be differences between students in different countries that would preclude implementing a generalized approach to learning within a digital game across multiple contexts (e.g., Lee & Luykx, 2007). The current study explores these two challenges. More specifically, the current study explores the two following questions:

* Corresponding author. Vanderbilt University, Peabody #230, 230 Appleton Place, Nashville, Tennessee 37203-5721, United States. Tel.: +1 615 322 5865; fax: +1 615 322 8999.
E-mail address: doug.clark@vanderbilt.edu (D.B. Clark).

0360-1315/$ – see front matter
doi:10.1016/j.compedu.2011.05.007

1. Can a conceptually-integrated digital game that overlays popular game-play mechanics with formal physics representations and terminology support explicit learning about Newtonian mechanics as demonstrated through post-test measures based on assessment items from the formal discipline (i.e., the Force Concept Inventory)?
2. How similar or different are the learning and affective experiences of students playing the game in two different countries (i.e., Taiwan and the United States)?

2. Background

The idea that games might provide affordances for science learning and inquiry is not idiosyncratic. In 2006, the Federation of American Scientists issued a widely publicized report stating their belief that games offer a powerful new tool to support education and encouraging governmental and private organizational support for expanded funded research into the application of complex gaming environments for learning (FAS, 2006). In 2009, a special issue of *Science* (Hines, Jasny, & Merris, 2009) echoed and expanded this call. Many studies provide evidence for the potential of digital games to support science proficiency in terms of conceptual understanding and process skills to operate on that understanding (e.g., Annetta, Minogue, Holmes, & Cheng, 2009; Barab et al., 2007; Clark, Nelson, Sengupta, & D'Angelo, 2009; Coller & Scott, 2009; D'Angelo, Clark, Nelson, Slack, & Menekse, 2009; Dieterle, 2009; Hickey, Ingram-Goble, & Jameson, 2009; Holbert, 2009; Kafai, Quintero, & Feldon, 2010; Ketelhut, Dede, Clarke, & Nelson, 2006; Klopfer, Scheintaub, Huang, Wendal & Roque, 2009; Moreno & Mayer, 2000, 2004; Nelson, 2007; Nelson, Ketelhut, Clarke, Bowman, & Dede, 2005; Steinkuehler & Duncan, 2008). Studies also show that games can support students' epistemological understanding of nature and development of science knowledge (e.g., Barab, Sadler, Heiselt, Hickey & Zuiker, 2007; Clarke & Dede, 2005; Neulight, Kafai, Kao, Foley, & Galas, 2007; Squire & Jan, 2007; Squire & Klopfer, 2007) and students' attitudes, identity, and habits of mind in terms of their willingness to engage and participate productively in scientific practices and discourse (e.g., Anderson & Barnett, in press; Annetta et al., 2009; Barab, Arici & Jackson, 2005; Barab et al., 2009; Dede & Ketelhut, 2003; Galas, 2006; McQuiggan, Rowe, & Lester, 2008).

2.1. Leveraging popular game-play mechanics to teach physics

Many popular commercial games offer interesting pedagogical opportunities for physics education with their focus on physics-based problem solving that involves careful manipulation force and motion. Specific titles of note in this genre have included, for example, *Portal, Marble Madness, Marble Blast, Orbz*, *Tiger Woods PGA, Switchball*, and *Mario Galaxy* (see examples in Fig. 1).

While these commercial physics games provide students with a strong intuitive 'feel' for physics concepts, they don't appear to (and were not designed to) help students make the leap from tacit understanding to more formalized knowledge. For example, Masson, Bub, and Lalonde (2011) found that students playing a commercially available physics-based game (*Enigmo*) "improved their ability to generate realistic trajectories" (p. 1). However, the game did not help them learn more from a direct instruction "tutorial" when compared to a control group. The tutorial focused explicitly on formalized concepts and "explain[ing] the forces acting on moving objects and objects at rest", but the game itself gives the students an idea about what trajectories look like and some information about angles of incidence vs. reflection. This mismatch between the content and experience in the game and the expected learning could be one reason for the limited gains.

Game-based experiences thus appear to require scaffolding in order for students to make the connections between the game and the more formalized knowledge required in a school-based context. These findings suggest that simply having players engage with physics-based games is not sufficient to help them learn physics. This result is not overly surprising; few people would suggest that playing soccer, for example, will teach people physics even though soccer is clearly a physics-based game in many ways. *Enigmo* was developed as a commercial recreational game rather than as a learning experience, and thus unsurprisingly follows along this path. Soccer, *Enigmo*, and other games, however, could potentially be re-envisioned or redesigned in a manner that would support explicit articulation and exploration of the core physics implicit in their game experience.

2.2. Designing physics games for learning

While most research on games for science learning has focused on 3D virtual worlds that players explore to collect and analyze data to form and test hypotheses (e.g., Barab et al., 2009; Ketelhut et al., 2006; Ketelhut & Schifter, 2010), some research has focused on building the learning goals into the actual movement and mechanics of the fabric of the world using popular game-play mechanics from the genre of physics games discussed above (Clark, & Martinez-Garza, in press; Clark, Nelson, D'Angelo, Slack, & Martinez-Garza, 2010; Clark, Nelson, D'Angelo, Slack, Menekse, et al., 2010; Nelson, Erlandson, & Denham, 2010). Clark and Martinez-Garza label these latter games as

Fig. 1. Tiger Woods PGA (2006), Orbz (2004), and Mario Galaxy (2007) all involve physics puzzles as their core game mechanic.

“conceptually-integrated” and the former games as “conceptually-embedded” to distinguish between the approaches. Essentially, conceptually-integrated games have the potential advantage of engaging the player with the science ideas targeted in the game a higher percentage of the play time (potentially the vast majority of play time) whereas conceptually-embedded games involve other interactions while moving and exploring the world between specific inquiry locations and activities, or as backdrop for those activities. The disadvantage of integrating the science learning goals directly within the motion and mechanics of the game world, however, is that while the players may spend the vast majority of game-play time interacting with the core ideas as a means of navigating through the world, making the core ideas and relationships explicit rather than tacit is a much bigger challenge than in the conceptually-embedded approach. The challenge for conceptually-integrated games therefore focuses on helping players articulate connections between the tacit ideas they develop with the explicit formal concepts and representations.

Examples of conceptually-integrated games include *Supercharged* (Anderson and Barnett, in press; Barnett, Squire, Higginbotham, & Grant, 2004; Jenkins, Squire, & Tan, 2004; Squire, Barnett, Grant, Higginbotham, 2004), *SURGE* (Clark, Nelson, D’Angelo, Slack, & Martinez-Garza, 2010; Clark, Nelson, D’Angelo, Slack, Menekse, et al., 2010; Clark et al., 2009; D’Angelo, 2010; D’Angelo et al.,2009; Nelson et al., 2010), and *FormulaT Racing* (Holbert & Wilensky, 2010). *Supercharged,* for example, is a 3D game in which players utilize and explore the properties of charged particles and field lines to navigate their ship through space. The player’s spaceship is moved through the game world by taking advantage of the properties of charged particles in the space. Three middle school classes participated in a mixed methods pilot study comparing learning outcomes of students playing *Supercharged* ($n = 35$) and those using a guided inquiry in-class curriculum ($n = 61$). Average post-test scores were significantly higher ($p < .05$) for the students who played Supercharged. The test included 12 questions on electromagnetism combined with pre-post interviews of a random sub-set of the students that were then transformed into additional quantitative data (Squire, Barnett, Grant, Higginbotham, 2004). This learning was only possible, however, when the teacher collaborating in the research created activity structures outside of the game to engage students in predicting and explaining what was happening in the game and reflecting on connections of the tacit intuitive knowledge that the students were building through game play to the representations and concepts of the formal discipline.

Holbert (2009) further documented the intuitive knowledge used and developed through commercial games with game mechanics compatible with the conceptually-integrated approach. He coded observational data of talk and gestures collected during ethnographic observations of and individual clinical interviews with children playing popular video games (*Mario Kart Wii* and *Burnout Paradise*). Holbert identified that children’s intuitive schema of velocity, acceleration, and momentum were at play while they were playing these games. These schemas have been previously identified as registrations (Roschelle, 1991) and phenomenological primitives (p-prims) (diSessa, 1993), and have been shown to play productive roles in the development of understanding of physics.

How can we leverage these intuitive understandings within conceptually-integrated games? Research on the success of certain simulations suggests that conceptually-integrated games focusing on force and have potential for learning (e.g., Klopfer & Purushotma, in press). These conceptually-integrated physics games involve close structural relationships with simulations. These games utilize accurate physics engines and engage students in exploring underlying concepts by having them vary parameters and observe the resulting outcomes of those choices in terms of changes to the trajectories of objects modeled by the engine. Research on simulations has shown that they can provide leverage in terms of harnessing a user’s spatial learning and perceptual systems in ways that text and verbal interactions do not (Lindgren & Schwartz, 2009). Strong evidence suggests that various types of simulations (and thus potentially digital games) used in conjunction with appropriate curricula and instruction can foster aspects of scientific expertise such as model-based reasoning, systems-thinking, construction of scientific explanations, and other conceptual skills and understanding (e.g., Edelson, Gordin, & Pea, 1999; Edelson, Salierno, Matese, Pitts, & Sherin, 2002; Harel & Papert, 1991; Papert, 1980; Raghavan & Glaser, 1995; Roschelle & Teasley, 1995; White, 1993; White & Frederiksen, 1998; Wieman, Adams, & Perkins, 2008). Just as in games, however, research on simulations suggests that students may use trial-and-error as opposed to mindful strategies (Chang, Quintana, & Krajcik, 2010) or focus on heuristics rather than deep learning (Turkle, 1997).

While this research suggests that conceptually-integrated physics games could potentially support learning while engaging players in important inquiry processes, the research also suggests that creating a design that does not result in shallow learning while also not destroying the flow and “game-ness” of the game may prove challenging. More specifically, how might we bridge and scaffold the rich tacit understandings developed during game play with explicit formalized understandings? In *Thought and Language*, Vygotsky (1986) discusses the potential for leveraging intuitive understandings from everyday experience (“spontaneous concepts”) with instructed scientific concepts to build robust understandings. The question remains whether or not the intuitive spontaneous concepts developed in games can actually be successfully leveraged into robust instructed concepts in the format and terminology of academic assessment and across domains recognized as central by the scientific disciplines. The current study explores the potential of overlaying popular game-play mechanics with formal representations and terminology in an attempt to bridge the tacit understandings from game play with the formalized explicit academic understandings.

Research Question 1. Can a conceptually-integrated digital game that overlays popular game-play mechanics with formal physics representations and terminology support explicit learning about Newtonian mechanics as demonstrated through post-test measures based on assessment items from the formal discipline (i.e., the Force Concept Inventory)?

2.3. Game-based learning environments across countries

In addition to the challenges of helping students articulate the tacit intuitive understandings they develop through game play in terms of the explicit concepts and representations of the formal discipline, there are also important questions to ask in terms of the degree to which the resulting learning environments will be viable across different countries. In other words, will a game-based learning environment developed in one country be productive in other countries?

Hofstede (2001, 2008) has carefully chronicled how differences across countries can dramatically impact teaching, learning, and other phenomena. In recent years, researchers in the field of science education have adopted theoretical perspectives that view learning as mediated by linguistic, cultural, and social factors to better understand the schooling experiences (e.g., Aikenhead & Jegede, 1999; Lee, 2005; Warren, Ballenger, Ogonowski, Rosebery, & Hudicourt-Barnes, 2001). In a synthesis of research literature, Lee (2005) includes theoretical perspectives which suggest that learning is possible when it occurs in contexts that are culturally, linguistically, and cognitively meaningful and relevant to students. Science teaching tends to vary across countries (Aikenhead & Otsuji, 2000) and learning is productively considered a cultural process even within a single classroom (Nasir, Rosebery, Warren, & Lee, 2006). Building on these ideas, a number of researchers have explored connections between students' understandings from "everyday" and culturally-specific contexts and their learning of formal disciplinary concepts and idea (George & Glasgow, 1988; Rosebery, Warren, Ballenger, & Ogonowski, 2005; Snively & Corsiglia, 2001; Warren & Rosebery, 2008). To better understand the potential of digital games more globally, we therefore explore similarities and differences between how students from two countries (and two different cultures) relate to, and learn from, the conceptually-integrated physics game at the heart of the current study. The purpose of this study focuses specifically on identifying potential similarities and differences between countries. This will then provide a foundation and starting point for future studies exploring the complex cultural factors in the two countries that might contribute to these similarities and differences.

Research Question 2. How similar or different are the learning and affective experiences of students playing the game in two different countries (i.e., Taiwan and the United States)?

3. Methods

The current study presents ongoing research with a conceptually-integrated physics game that we have developed called SURGE. The research is funded by the U.S. National Science Foundation through a grant with the same name (Scaffolding Understanding by Redesigning Games for Education). The current study analyzes and discusses results from research with SURGE in schools in Taiwan and the United States. The following sections describe the participants, data collection, secondary research conditions, and the game context.

3.1. Participants

3.1.1. Taiwan

Data in Taiwan was collected for 71 eighth grade students and 137 ninth grade students in seven classes taught by two teachers at two public middle schools. The students were in standard science classes, 58.9% female, and represented the full range of students at the school because science is not a "tracked" subject at the school (i.e., a subject in which students are "tracked" into different classes based on achievement).

3.1.2. United States

Data was collected in the U.S. for 72 students in five seventh grade classes in a diverse but predominantly African-American urban school in the southeastern United States. The students were in the standard 7th-grade science classes, 58.3% female, and represented the full range of students at the school because science is not a "tracked" subject at the school.

3.2. Data collection

The study was conducted during the school day over the course of approximately three class periods for each group of students in each country. The students completed a pre-test, post-test, and short written survey. Data collected and analyzed in the studies included:

- Pre-tests and post-tests focusing on players' understanding of formal "instructed concepts" through an instrument including 12 multiple-choice items based on the Force Concept Inventory (FCI) (Hestenes & Halloun, 1995; Hestenes, Wells, & Swackhamer, 1992; Jackson, 2007). Four questions focused on the application of impulses to objects, four focused on interpreting kinematics (i.e., position, velocity, and acceleration) in dot trace representations, and four focused on the application of constant acceleration to objects. Nine of the twelve items were based directly on a simplified version of the FCI (Jackson, 2007), including three of the impulse questions, two of the dot trace kinematics questions, and all four of the constant acceleration questions. We created one impulse question and two dot trace kinematics questions that mirrored the FCI questions so that we could have four questions in each category. This study focuses on these FCI questions and does not analyze four vector concept items (Nguyen & Meltzer, 2003; Shaffer & McDermott, 2005) that were part of a separate study (D'Angelo, 2010). The FCI, recognized in the international physics education community as one of the best measures of academic conceptual understanding in physics, measures formalized conceptual understanding of Newtonian mechanics and kinematics. In the Taiwan study, these tests were conducted using a separate survey website (surveymonkey.com), and in the USA studies, they were done within the SURGE software environment.
- Our studies included a survey, taken by students after completing the post-test. It was composed of (a) a multiple-choice section, asking students to rate their affective response to SURGE on a Likert-type scale, whether they believed that SURGE would be more appealing to girls or boys, queried students about their gender, and how much they themselves played video games normally; (b) a free-response section in which we asked students which pre-test questions did playing SURGE help them to answer, and what added features or improvements they would like to see in future versions of the game.
- Observations were made by researchers of students playing the game during each trial in terms of students' actions, reactions, and attitudes playing the game. These observations were collected primarily as field notes to inform revisions to the design of SURGE. The research group then discussed these observations and synthesized them to provide additional insight into the learning and affective findings from the tests and surveys.
- In addition to the data sources discussed above, which were common to Taiwan and the U.S. implementations, the U.S. implementation also collected students' scores during game play on individual trials of each level. We analyze the data on game-play scores and numbers of trials attempted to check for interactions between gender, learning outcomes, how much students reported liking the game, how many times the students attempted levels (which means how many times they replayed levels rather than simply choosing to move on to the next level), and how successful students were in playing the game (as represented by higher scores on levels during game play).

3.3. Secondary research conditions

In addition to the main comparison between countries, students in each country were also randomly assigned within their classrooms to a secondary condition as pilots for future development. These secondary conditions did not affect core game play. These comparisons were conducted to check potential subtleties in game design on learning.

3.3.1. Taiwan

Students were randomly assigned to one of two conditions focusing on the degree of storyline included in the level introduction and level summary "cut scene" pages. In the control condition, cut scene pages included more minimal storyline in the cut scenes. The experimental condition added some additional detail to the storyline. Both conditions included the same instructions and science content.

3.3.2. United States

Students were randomly assigned to one of two conditions in terms of the inclusion of the velocity cross representation, which is one of the formal representations included in the SURGE. One condition included this representation in the game levels and the other did not.

Ultimately, neither of these secondary conditions made a significant impact on learning outcomes, and thus they are not discussed in these analyses, but they were part of the data collection process and they are therefore described here.

3.4. SURGE game context

We utilized our SURGE physics game environment as the platform for research. We built SURGE within the Unity 3D game engine (unity3d.com). SURGE is a conceptually-integrated game, as defined earlier, rather than a conceptually-embedded game for learning (i.e., the science to be learned is integrated directly into the mechanics of navigating through the game world rather than being embedded as an activity that is visited in the game world, as is typically the structure in many virtual worlds designed for science learning). The SURGE platform is intended to investigate design principles for connecting students' intuitive "spontaneous concepts" about kinematics and Newtonian mechanics into formalized "instructed concepts" by overlaying mechanics of popular commercial video games with "marble" mechanics such as *Mario Galaxy* and *Switchball* with formal representations and connections to formal concepts of Newtonian mechanics. SURGE incorporates the game play designs of these popular "marble" games in the context of a space-based adventure (Figs. 2 and 3).

Students play the game as the character Surge, a smart and brave female alien, who is being called upon to save the adorable Fuzzies from the evil Emperor Hooke. Another character, named Lerpz, gives Surge advice throughout the game and helps scaffold the actions and physics concepts that the students learn while playing. The Fuzzies also help Surge along the way, providing information about the special Motion Map Regions in the game and encouraging Surge as she travels through the levels. Students use the arrow keys on their keyboards to navigate around barriers and through corridors, trying to find the Fuzzies that they need to save. The levels are designed so that physics concepts build upon one another and gradually introduce the student to new ideas and ways of interacting with the game world. The two studies in this paper focus on thirteen levels of the SURGE game, broken up into two modules. The first module uses an impulse control system, where every time the student pushes an arrow key a fixed impulse is applied to Surge's ship (represented as a white ball in the game). The second module uses a constant force control system, where students can hold down an arrow key to apply a constant force in that direction. Overlaid on the screen are different read-outs of information for the student, including their ship's current speed, the number of impulses they've used, the number of collisions with the walls, and their elapsed time on a given game level. Students are told to minimize their collisions, level completion time, and number of impulses in order to get a high score. There are also on-screen buttons used to reset or pause the level and to stabilize Surge's ship if it starts moving out of control. A vector representation of students' velocity is also on the screen, showing their current speed and direction. Some levels include a Motion Map Region, where students must maintain a constant velocity, increase their speed, or decrease their speed (a Fuzzy tells them which one to do) in order to continue in the level.

Our driving design principles and initial goals for SURGE are summarized below.

- *Overlay popular game-play mechanics with key formal physics representations including vector representations and dot traces.* We wanted to build these formal representations into the game such that using the representations would be fundamentally useful and advantageous to players. Our vector display, for example, included both composite and component vectors so that players could more easily determine how many impulses or acceleration would be required in a given direction to achieve their goals. Similarly, the dot

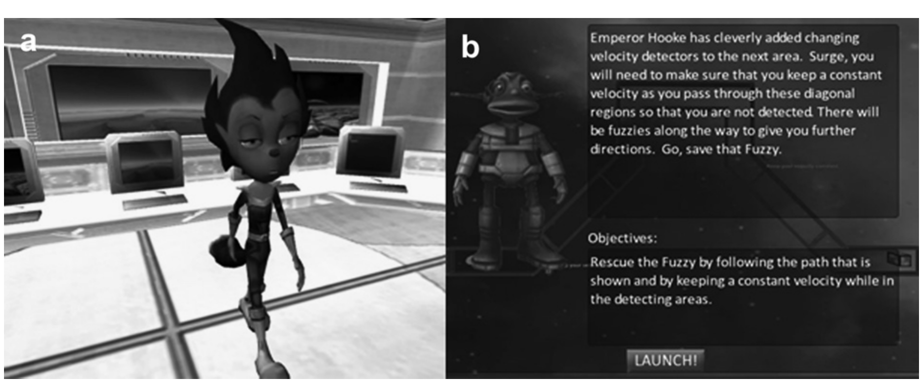


Fig. 2. a–b. Surge and Lerpz in different versions of the cut scenes between levels.

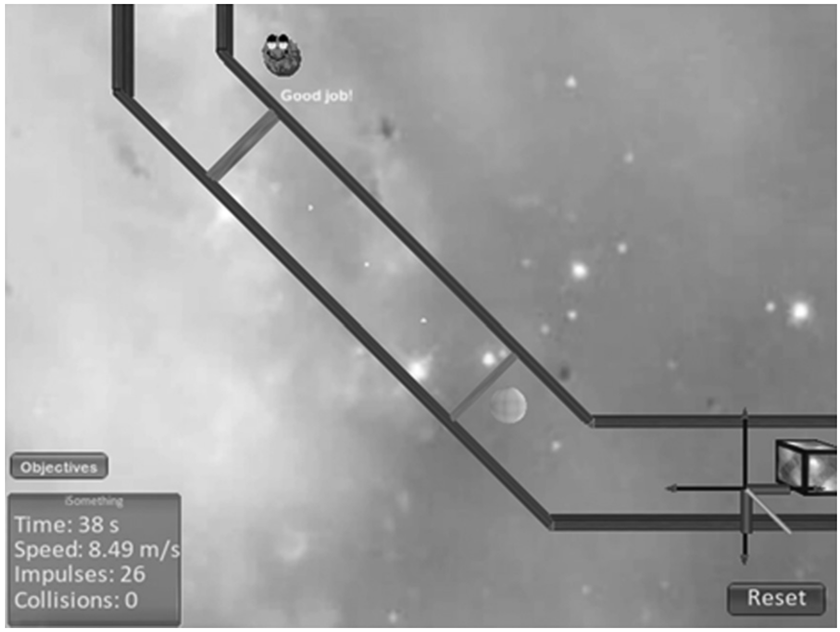


Fig. 3. Screenshot from an impulse level in SURGE. Students must guide Surge in her spherical spaceship through maze-like prisons to rescue Fuzzies.

trace representations were designed to help the player visualize constant and changing velocities as part of helping them master these ideas and techniques for puzzles within the game involving "velocity detection zones" that served as keys the players used to unlock passageways.

- *Each level involves specific challenges directly linked to physics concepts. To complete these challenges, students need to learn and apply many principles related to mechanics (e.g., impulse, inertia, vector addition, elastic collision, gravity, velocity, acceleration, free-fall, mass, force, projectile motion).* In levels with the "velocity detection zones", for example, the puzzle involved figuring out what constant velocity entailed in order to maneuver safely through the level and unlock passageways. In order to navigate through the level, a player needed to understand characteristics of constant and changing velocities. Other levels created their challenges or puzzles around combining vector components. All of these challenges, however, were enacted through the player's navigation through the game world. The challenges were thus integrated into the mechanics of world rather being challenges simply embedded into the world. This is, in fact, what distinguishes conceptually-integrated games from conceptually-embedded games.[1]
- *Each level highlights one or two topics, and levels allow students to connect the concepts together and to see the relations that exist among the topics.* For example, in the multiple dimensional motion levels, students learn and apply the concept of applying impulses at right angles to produce motion in two dimensions. This builds on students' knowledge of additive and canceling impulses and motion in one dimension, and extends that knowledge to include the resultant motion of impulses at right angles. This approach allows students to gain a firm grasp of a concept before new concepts are introduced.
- *Protect students from failure and seek to minimize frustration for novice players.* We did not want to create a game that would be productive for players with extensive gaming experience but that was frustrating or less productive for less experienced players. We therefore intend to minimize frustration and scaffold success for less experienced players. As an example, the early levels do not include fixed failure triggers forcing the level to reset (e.g., our initial versions of SURGE do not include the possibility of Surge's ship exploding after a set number of collisions). If a player can complete a level they earn at least a bronze medal for that completion. Silver medals are intended to be fairly challenging to attain, and the gold medals are meant to be very challenging. This is a design principle that we discovered required modification (as discussed later in this article) to be balanced with the idea that players also cannot be allowed to progress to subsequent levels without first attaining a certain baseline of mastery. There is therefore a tension between protecting students from failure while also requiring certain levels of mastery for advancement. We discuss this in further detail later in this article.
- *Integrate physics ideas and terminology into pre-level and post-level story and feedback screens and within the levels of the game itself.* Several levels, for example, include "detection" corridors where the player needs to maintain a constant velocity, increase their velocity, or decrease their velocity in order to open a gate. This specific terminology is explicitly delivered as instructions by a friendly non-player character (a "Fuzzy") in the level and is central to succeeding in the level. As discussed later in this article, we continue to consider this design principle as critical to helping players connect intuitive and formal understandings, but we found through this research that we need to (a) focus further on the "just in time" aspects of this scaffolding within the play of the level itself and (b) make

[1] Note that while the focus of SURGE and this article is on the structure and affordances of conceptually-integrated games, the authors also acknowledge that conceptually-embedded games for learning (such as games set in virtual worlds where players takes on the roles of scientists) also have valuable affordances, but conceptually-embedded games are more often studies and are not the purpose of this article (although the authors do conduct research and development on these games as parts of other projects and think highly of their specific affordances). Thus, this article should not be read as claiming that conceptually-integrated games are superior to conceptually-embedded games, but instead should be read to think specifically about the affordances of conceptually-integrated games and their design for learning.

the pre-level and post-level story elements more interactive and game-like themselves to increase the players' interaction and engagement with the ideas in these pre-level and post-level elements (otherwise players often skip right over the pre-level and post-level story elements).

- *Incorporate foundational multimedia principles into the design to the interfaces (e.g., Clark, Nguyen, & Sweller, 2006; Mayer, 2009) to reduce unnecessary cognitive load for the player to process the ideas and information in the game.* Games tend to involve contexts that are visually much richer than those found in other multimedia formats for learning (such as simulations). This can easily lead to players not understanding which aspects or details of the screen are salient and which are just environmental detail. We have observed in our research, for example, that players sometimes don't even realize that they have the key representations unless they are centered on their ship (the sphere they are moving through the game). If the representation is placed in the lower corner of the screen, players sometimes don't notice it at all. In a recent study where players in each classroom were randomly assigned to versions of the game that included the vector representation either in the corner of the screen or centered on their sphere, a player who had the representation in the corner pointed to the screen of his neighbor (whose vector representation was centered on the sphere) and asked why he didn't get a "speed representation" – When we pointed our his representation, he said, "Wow! I can't believe I didn't see it!" Thus, we believe that careful application of multimedia principles to signal and cue attention may be even more important in game design than in the design of other multimedia formats for learning.

3.5. Relationship of SURGE to the normal curriculum

The implementations of SURGE for this study preceded the normal units that the teachers in Taiwan and the U.S. planned to teach. SURGE was introduced to the students in this context. As a teacher in Taiwan explained to the students, "SURGE relates to the concepts of force and motion, which you will learn later in the textbook". The students were therefore given some framing for the game in terms of curricular goals, but they had not studied the topics so as to provide any in-depth background for the game in terms of the curriculum for the current study. Thus, SURGE preceded the normal curriculum, but the teachers provided some connections for the students between SURGE and what was to follow in the curriculum.

Later when the Taiwanese and U.S. teachers taught the force and motion unit (text-book based in Taiwan and a mix of textbook and other materials in the U.S.), they asked students to recall SURGE and used the examples from SURGE, such as how a spaceship can change its direction in outer space? Some students in both countries expressed that they hoped to play SURGE again when the teacher mentioned examples from SURGE. All three teachers thought it would have been beneficial if they opened SURGE and explained concepts in the curriculum in terms of the students' prior experience with SURGE, but due to the limits of time, the teachers only verbally mentioned SURGE instead of bringing students to the computer lab to use SURGE again. Thus, while the teachers did not later use SURGE again during the normal unit, they did refer to it to provide examples during their regular unit's textbook-based teaching.

4. Results

The following sections outline the (a) learning outcomes on the tests, (b) affective outcomes from the surveys and interviews, and (c) observations by the researchers in each country. We then discuss and compare and discuss similarities and differences between and within countries.

4.1. Learning outcomes

4.1.1. Overall

A matched-pairs *t*-test analysis was conducted on the pre and post-test scores achieved by students. This test revealed significant learning gains across the two countries for the 12-item pre-test and post-test when considering total scores, *M* (pre-test) = 3.79, *SD* = 2.20; *M* (post-test) = 4.16, SD = 2.49; t(250) = 2.70, *p* (one-tailed) = 0.004. An item-by-item analysis using McNemar's χ^2 test showed significant gains were on item 4, χ^2 (1251) = 14.02, $p < .001$, Cohen's $d = 0.1541$ and item 11, χ^2 (1251) = 4.40, $p = .036$, Cohen's $d = 0.344$ (Fig. 4a and b present these items).

4.1.2. Taiwan

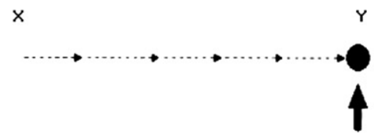
A matched-pairs *t*-test analysis was conducted on the pre and post-test scores achieved by students on 12-item test. This test revealed significant learning gains when considering total scores, *M* (pre-test) = 4.35, *SD* = 2.18; *M* (post-test) = 4.73, SD = 2.54; t(179) = 2.431, *p* (one-tailed) = 0.008, Cohen's $d = 0.1619$. An item-by-item analysis using McNemar's χ^2 test revealed that students improved their performance in some specific parts of the assessment more than in others: significant gains were observed on items 4, χ^2 (1180) = 20.10, $p < .001$; item 11, χ^2(1180) = 10.41, $p = .001$; item 1, χ^2 (1180) = 5.14, $p = .023$; and almost significant gains on item 6, χ^2 (1180) = 3.75, $p = .053$ (Fig. 4a and b present these items). Although the net performance gain for these items was modest (0.365 SD max, 0.178 SD min), these gains were consistent across gender and levels of gaming experience.

4.1.3. United States

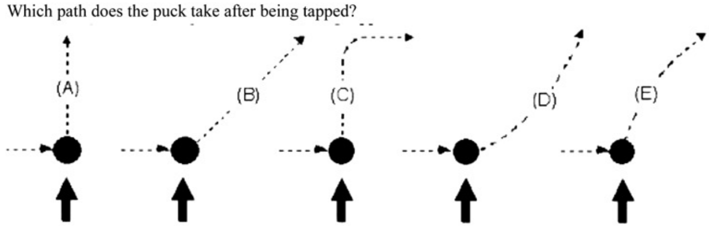
A matched-pairs *t*-test was conducted on the test scores achieved by students before and after a single play session of SURGE, approximately 45 min long. Students did not show a significant improvement in test total scores, *M* (pre-test) = 2.38, *SD* = 2.32; *M* (post-test) = 2.62, SD = 2.67; $t(70) = 1.211$, *p* (one-tailed) = 0.115. A item-wise analysis using McNemar's test for χ^2, similar to the one used in the Taiwan study, was used to find significant gains in specific items, but did not show conclusive results, partly because the much smaller sample size ($N = 71$ vs. 180 in the previous study) was detrimental to the statistical power of our assessment, and also potentially because a number of students raced through the post-test so that they could continue playing the game (see discussion of researcher observations in the Results section below). That said, three of the items with the largest gains matched three of the four items from the Taiwan implementation with the largest gains. These items included items 1, 6, and 11.

a

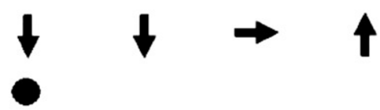
Item 1. A hockey puck slides on very smooth ice in a rink at a constant speed (imagine that's there's no friction) in a straight line from location **X** to location **Y**. In the figure, you're looking down at the puck. When the puck reaches **b**, a player **taps** it from the direction of the heavy print arrow.



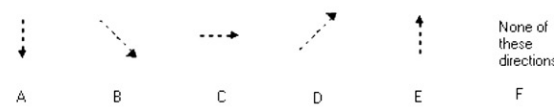
Which path does the puck take after being tapped?



Item 4. A different hockey puck is sitting still on the ice. A player hits it lightly in different directions. Down, Down, Right, Up. Each hit is the same strength.

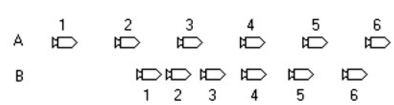


What direction does the puck end up travelling after the four quick hits?



b

Item 6. Two spacesphips are travelling. A scientist takes measurements. Later he makes this drawing. The little spaceship figures show where both of the spaceships are (their positions) at every second of time. The spaceships are both travelling to the right.



Based on their positions in the diagram above, describe the forces on the two spaceships

(A) Spaceship A has no force on it, Spaceship B has a force towards the right on it.

(B) Spaceship A has no force on it, Spaceship B has a force towards the left on it.

(C) Spaceship A and Spaceship B have no forces acting on them

(D) Spaceship A has a force towards the right on it, Spaceship B has no force on it.

(E) There is not enough information to answer the question.

Item 11. Imagine that you're a space traveler far in the future; you're traveling to another star system. Your spaceship drifts sideways in outer space from location **x** to location **y**. No forces act on the ship during this time. At **y**, the captain turns on the ship's engine, producing a force (called a **thrust**) on the ship at a right angle to the line **xy** (toward the top of this page). The thrust stays constant until the ship reaches some location **z**.

At location z the captain turns off the spaceship's engine, so the thrust from the engine drops to zero. Which path does the ship follow beyond location **z**?

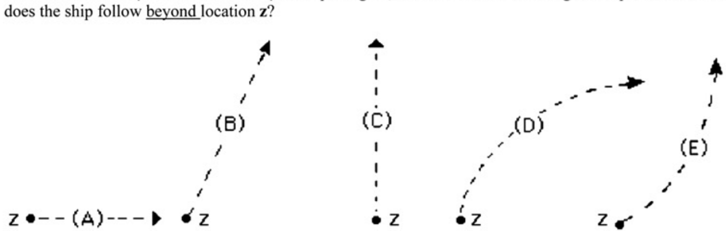


Fig. 4. a. Items 1 and 4 from the pre/post-test based on items from the simplified FCI (Hestenes & Halloun, 1995; Hestenes et al., 1992; Jackson, 2007). b. Items 6 and 11 from the pre/post-test based on items from the simplified FCI* (Hestenes & Halloun, 1995; Hestenes et al., 1992; Jackson, 2007). (* Question 6 was based directly on the Simplified FCI test, which included the scientist's gender as male. As part of our goal of supporting gender equity through SURGE, future versions of the assessment will change the scientist's gender to female.)

4.2. Affective findings

4.2.1. Taiwan

The majority of the students liked their experiences using SURGE. When asked "how much did you like playing SURGE?" in the post-survey, over half (62.4%) of the 208 students liked or really liked playing SURGE. About 31% thought it was okay, and only 6.4% did not like it (Fig. 5). This result is consistent with the classroom observations that many students showed high motivation and engagement as they used SURGE. Moreover, 69.8% thought that SURGE is fun for both male and female students. The majority of the students believe that SURGE is fun regardless of gender.

The students had used computers in the laboratory for their computer courses but not for their science courses. Their experiences in playing computer games varied: 42.6% of the students usually spend 1–2 h per week and 33.2% spend 3–6 h per week playing computer games; 10% spend more than 12 h and another 10% do not play games at all. In general, more males in this group play video games than females, with 20 males vs. 6 females reporting intensive video game experience (i.e., 12 or more hours or play per week), and 18 females vs. 3 males reporting no video game use at all (Fig. 6).

When asked "What did the game help you learn on the test?", many students directly specified the item numbers of the test on which playing SURGE helped the learning. Of the responses, 41.1% indicated between one and four specific items, and 46.9% indicated more than five items of the test that the learning experience with SURGE helped. Generally speaking, the majority of the students thought that SURGE helped them on the test, and 22.4% indicated that SURGE helped a lot, but 12% of the students indicated that SURGE did not help.

When asked to make suggestions to help improve SURGE, many students (44.2%) focused on the representational aspects. For example, students indicated that if the Chinese characters would be clearer the learning process would even be better facilitated, which is perfectly reasonable as we discuss later in the implementation observations section below. About 17% of the students made suggestions on the content. Some of these students suggested that we should increase the complexity of the tasks while some other students suggested decreasing the complexity. Some other students made suggestions on how to make the game more fun, such as including monsters, music and sound effects, and so forth. These are not directly related to the learning content but may seem essential to the students to fulfill their expectation of what constitutes a computer game.

SURGE elicited generally positive affective responses from participants in this study ($M = 2.21$, where 1 is most positive response and 5 least positive, $SD = 0.97$), with more positive response among males ($M = 1.90$, $SD = 0.78$) than females ($M = 2.41$, $SD = 1.01$). The test group as a whole expressed the opinion that SURGE would be more interesting to females than to males (45 responses vs. 8, of those who expressed an opinion).

4.2.2. United States

Students in the United States had a more positive response to SURGE than did the Taiwan students ($M = 1.72$, where 1 is the most positive response and 5 the most negative, $SD = 0.89$), although interest in the game did not correlate with improved learning outcome. SURGE was slightly better received by females ($M = 1.59$, $SD = 0.83$) than by males ($M = 1.90$, $SD = 0.96$), although paradoxically, students as a whole expressed the opinion that the game would be more interesting to males (16 responses vs. 6). No students reported a strongly negative response to SURGE (Fig. 5).

With regards to their gaming experience, 34% of the students reported they normally play 1–2 h per week, 33% responded 3–6 h per week, and 10% (6 females, 1 male) said they played no video games at all. Fifteen percent of students reported playing games extensively (12 h per week or more). Males in this group are more likely to be video game players than females; 14 of the 30 males in this sample play 7 h per week or more, while only 4 of the 42 females say they play that much (Fig. 6).

When asked for suggestions for improving SURGE, 24 students (33%) provided substantive responses; the most popular suggestions addressed issues of game design or game metaphor ($n = 14$), requests for more varied content ($n = 6$), and increased difficulty ($n = 4$). The number and quality of suggestions we received revealed to us the extent and quality of the participants' gaming templates and vocabularies, which remain central to our view of how the well the game is received by students.

After the post-test, students were asked to report on what they had learned from playing SURGE. Out of 71 respondents, 35 mentioned a better understanding of velocity and/or speed (49%), 6 mentioned acceleration (8%), while 7 felt they had learned nothing at all (10%). Since

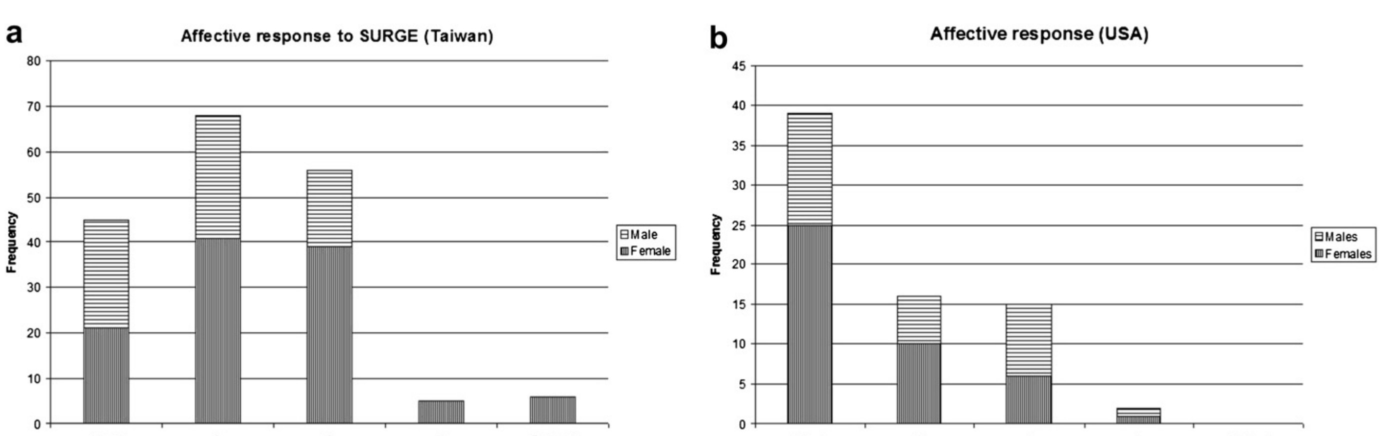


Fig. 5. Distribution of affective response to SURGE from (a) Taiwan group and (b) USA group on a Likert-type scale. Students in the USA group responded more favorably to SURGE.

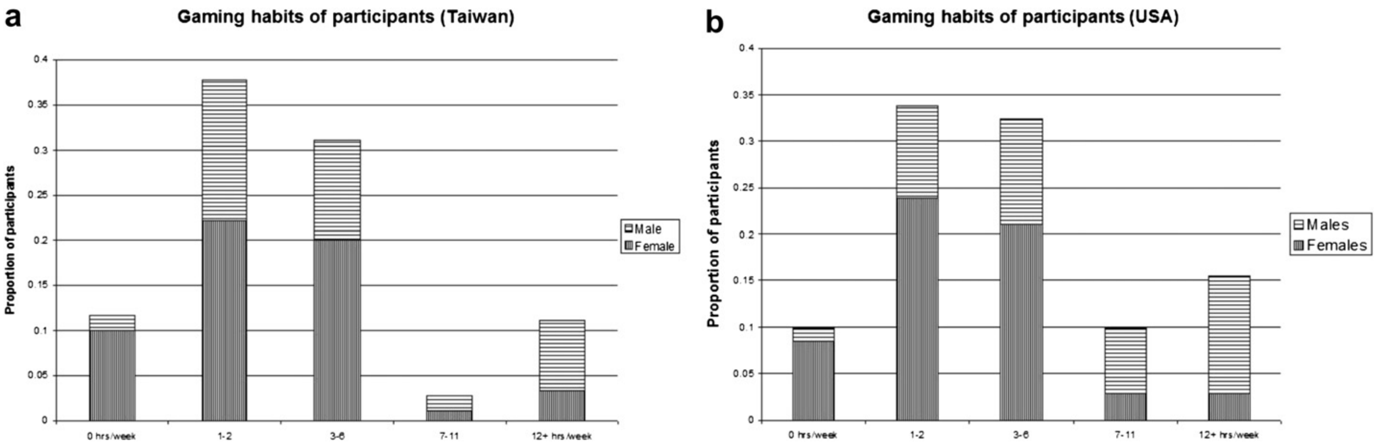


Fig. 6. Survey responses of the (a) Taiwan pilot study and (b) the USA pilot study concerning average time spent gaming per week.

velocity is addressed in the game in the form of detection zones focusing on formal physics dot trace representations, we conclude that this particular feature is effective at focusing students' conceptual thinking. Although students tended to recall the concept of velocity in the surveys, whether or not the detection zone feature supports learning outcomes is unclear, as none of the FCI assessment items that were used highlight the concept of velocity-as-vector, which is distinct from the more intuitive notion of speed.

4.3. Relationship of game-play scores to learning, affective reaction, and gender

In addition to the data sources discussed above, which were common to Taiwan and the U.S. implementations, the U.S. implementation also included the collection of students' scores during game play for each individual trial (attempt/play-through) of each level. We wanted to check for any possible interactions between gender, how frequently a student replayed levels, achievement in game play (as represented by a student's best score on each level), gains on the post-test, and how much students expressed liking the game (as measured by their rating of how much they liked the game on the survey after playing the game). We did not observe any interactions between greater achievement in game play, as measured in z-normed scores per level based on the each student's best score on each level, and learning gains or expressed affective reaction to SURGE. Similarly, number of times players replayed levels they had completed (either for fun or to attempt to get a higher score) does not correlate with learning gains, students' affective response to SURGE or increased gaming habit/experience. In comparing test gains with affective reaction on the survey, we found that test gains did not correlate with affective reaction, suggesting that students do not necessarily have to enjoy SURGE to benefit from play.

In terms of gender, we found that boys did not achieve better learning outcomes than girls (as reported earlier), regardless of boys' greater self-reported gaming experience (1–6 h per week was the most frequent response for girls, 7 h or more was more frequent for boys). However, there was a different play dynamic among boys and girls. We found that boys had more completed trials (attempts) than girls on levels in the game, M (boys) = 20.66 vs. M (girls) = 17.38, $t(70) = 2.32$, $p = .012$. Since the version of SURGE they played has a total of 13 levels, these numbers indicate that both genders were replaying levels, not merely advancing through levels without regard for the scores they had achieved. Boys on average more replayed levels than girls, M (boys) = 1.76 plays/level attempted, M (girls) = 1.59, $t(70) = 1.85$, $p = .033$.We also found that boys had higher "best scores" on levels in the game. When considering only the highest score achieved for all replays of each level (the player's "best score" on a level), boys averaged 0.115 SD above the mean for all students on each level, while the girls scored 0.017 SD above the mean for their "best score", although the difference is not statistically significant ($p = .25$). Furthermore, the variances in "best scores" within both genders are much larger within than between genders (within boys = 0.43 SD and within girls = 0.30 SD). If there were a difference here in terms of "best scores" between genders, it would likely stem from the earlier reported finding that boys replayed each level more than girls (1.76 replays/level vs. 1.59), so the increased average "best score" would perhaps be the result of repeated practice.

In summary, as reported, boys and girls did not demonstrate differences in learning outcomes, nor were learning outcomes correlated with reported gaming habit. This suggests that SURGE provides a relatively novel experience for this age group that does not favor more experienced gamers nor does it favor boys overall, although boys appear to replay levels somewhat more frequently.

4.4. Implementation observations

As outlined in the Methods section, field notes were collected by researchers as students played the game during each trial in terms of students' actions, reactions, and attitudes playing the game. The research group then discussed these observations and synthesized them to provide further insight into the test and survey data.

4.4.1. Taiwan

In general the teachers and students had positive reactions to SURGE. The teachers were satisfied when they observed that the students were all engaged in SURGE. All the students showed high interest and engagement in SURGE at the beginning of the game, and this engagement remained strong for most students throughout the program. The students concentrated on the SURGE interface with little off-task behavior. They asked technical and conceptual questions, indicating that they were mindfully engaged. When questions arose students sought help from the teacher and sometimes from their peers. The students were heard making discovery sounds such as "a-ha" or laughing with enjoyment, indicating that they were positively motivated.

Student engagement was high in spite of technical challenges related to displaying Chinese characters clearly on the older computers used by students in the Taiwan implementation. While Unity (the program that the game was developed in) could display English characters perfectly on older PC computers, we discovered that Unity did not handle Chinese characters as well on older PC computers, with the result that the Chinese characters in many cases became too blurry to for the students to read. To attempt to compensate, we created accompanying paper document materials keyed to the game levels showing Chinese text so that students could read it. While this worked, it was sub-optimal, and we believe that this approach diminished the learning opportunities for students by adding confusion to the task and increasing learner extraneous cognitive load related to switching attention between the game and the document. Most students read the printed documents only at the beginning of the game and then went through the levels without looking at the documents again.

We observed that some students completed all the levels in SURGE much more quickly than others. Some students showed quick proficiency at controlling SURGE game elements, resulting in finishing the game faster. This might relate to the diversity of students' previous experiences playing computer games. Students also learned efficient strategies for completing the game while playing. For example, many students did not succeed in passing the detection zones in the first several levels and needed to go back. However, most students learned to control the spaceship to successfully pass the detection zones after four or five levels. Students were generally observed to be purposeful and mindful in their interactions with the game independent of the pace at which they progressed through the game, with the exception of their interactions with the printed documents. The teachers thought that the different amounts of time that students needed to finish could pose classroom-management challenges for teachers. Another challenge for teachers involved deciding how and when to guide and remind students to slow down and think to improve learning with SURGE. The students might have thought they were supposed to complete the tasks in SURGE as quickly as possible without paying attention to the other goals that SURGE highlighted.

We observed three design elements in SURGE that seemed to impact students' performances, each in a different way: the velocity cross representation, the "stabilize" button, and the virtual medal award system. The velocity cross representation helped students visualize the current status of the spaceship and plan their subsequent actions. The "stabilize" button helped students regain control if the spaceship collided with walls too much or involved too many forces. Over-dependence on the stabilize button was observed in the USA implementation. In contrast, only a few students overused the stabilize button in the Taiwan implementation. The medal award given at the end of each level motivated students to seek strategies to better reach the goals (and thereby earn 'higher' medals). We observed that the students cared about and compared with peers what kind of medal they received at the end of each level for the purpose of having fun (as opposed to taking this seriously as the only goal to achieve). They hoped to earn a gold medal but seemed comfortable when getting other medals. The students cared much less about the score they received at the end of each level than the color of their medal.

Overall, the Taiwanese teachers stated their belief that SURGE brought novel learning experiences to the students, which related to the high motivation and engagement that the students showed. Although the students were quite proficient at using computers in their everyday experiences, the majority of their prior science learning experiences involved learning knowledge from their teacher and textbooks through lectures. This was the students' first time to use a game-like program in their science lessons. The context, functions and features of SURGE also kept the motivation and engagement high throughout.

4.4.2. United States

As in Taiwan, the teachers and students had positive reactions to SURGE. The students were very excited to see the laptops in their classroom when they entered. As one student exclaimed, "Alright! This is more like it! This is what science is supposed to be!" The students were very excited about playing the game and there was animated talk between players across the room. An assistant principal at the school came to observe one of the class periods as part of his standard annual observations of new teachers at the school. At the end of the class period, he was visibly very excited about SURGE. He said that he had "never seen these kids so engaged and on-task". When the lead author returned to the school the following week to judge science fair projects, several students asked when they could play SURGE again, and the mother of another student came up to the lead author to tell him how much her daughter had liked SURGE. These types of comments in conjunction with the survey responses and our own observations of student engagement lead us to believe that SURGE was indeed engaging and interesting for the students. While engagement was high, though, we also noticed a few issues during this implementation meriting consideration.

As seen in Taiwan, students were not interested in reading the introductory and summary cut scenes for the levels. Unless we talked to the students and explained that students in earlier classes had gotten lost when they didn't read the introduction screens, most students spent little or no time reading the introduction screens and instead clicked right through to the level itself. Similarly, they tended to check the summary screens only for the color of the medal they had earned (bronze, silver, or gold).

Also as seen in Taiwan, students completed the game at different rates. While some students were interested in the color of medal they earned, other students were less concerned about the points and medals than they were in simply advancing to the next level as quickly as possible. As one student said after crashing and bludgeoning SURGE to the end of each level and seldom earning higher than a bronze medal, "Ha! I beat your game!" Many students mirrored this general sentiment of having great interest in the achievement in getting to the end of the game, with less interest in the achievement of getting gold medals on each level. After finishing the last level, some students went back to try and earn gold medals on each level, but a number of students also went back and created new unintended challenges for themselves, such as seeing how many collisions they could rack up or how high a velocity they could achieve. Essentially, most of the students pursued challenging goals, but their goals didn't always parallel the performance goals that we had aligned with developing understanding of the targeted core concepts.

We also noticed that some students positioned the mouse pointer over the stabilize button and used that feature each time they wanted to come to a standstill. We had added the stabilize button to help players who were having a hard time regaining control. However, because the players largely were not interested in points or were not aware of the points and how the scoring worked, they used this stabilize button as their primary means of stopping. They did not view this strategy as circumventing the purpose or goals of the game in any manner. From their perspective, they appeared to view this as the intended, logical, and obvious approach within the game, which it could well be considered to be from a game mechanics perspective, even though this perspective competed negatively with our pedagogical goals of having students apply impulses to moderate their velocity (and thus come to understand the independence of the *x* and *y* axis velocity components and the need to account for both when changing directions in the maze).

5. Discussion and comparisons

In summary, there were important trends across the implementations.

5.1. Embodied and intuitive learning

We see learning gains on the assessment items most directly related to the embodied experience of navigating the SURGE ship through the game space, particularly for items where key terms and ideas were integrated into the game play in a manner that player's success was made contingent on consciously understanding the terms and ideas. As an example, players tend to make connections between their experiences maneuvering their ship in the game and questions regarding how the application of constant force or impulse will affect the trajectories on the tests. Similarly, while there were not specific questions on the pre/post-test about the meaning of constant velocity, observations and surveys suggest that students clarified their understanding of velocity as including both speed and direction, as well as ideas of increasing velocity, decreasing velocity, and constant velocity, through the inclusion of those concepts directly into the challenge of unlocking gates to allow passage through the levels. As we have described, out of 71 students in the U.S. implementation, 35 specifically answered 'velocity' when asked what the game helped them learn. The term velocity was integrated into the game itself, appearing frequently on pre and post-game level summary screens as part of the goals and feedback information provided to players. Seemingly, use of the formalized physics term in conjunction with game play making use of velocity as a core element bolstered student awareness of the term. It is less clear, from the data collected thus far, whether familiarity with the term correlates with stronger understanding of the concept. In our future work, we can make use of the data collected around formalized language use associated with game play in SURGE to help teachers scaffold follow-up classroom-based discussions around the terms most frequently appearing in game log files. These scaffolded discussions can help tease out the level to which integrated formalized language in the game bolsters not just vocabulary acquisition but also corresponding conceptual understanding, in addition to echoes of conceptual learning found via FCI questions.

Interestingly, the FCI questions tend to focus more on intuitive understanding rather than measuring explicit understanding. We chose the FCI because it is the most well known test of conceptual understanding of force and motion. We wanted to use a recognized and accepted benchmark of physics understanding rather than a test we created (which could be viewed as less strong evidence for learning from a physicist's perspective). In our future studies, though, we will develop more of our own items for the test that include a focus on the more formal aspects of understanding. Increasing the sensitivity of the assessment items to the connections between formal and intuitive understandings will likely require further scaffolding of explicit formal ideas and their connection to the intuitive understanding that students' are currently developing through game play. This further emphasis on connections between intuitive and formal understandings will continue to be a core focus of SURGE research and development.

This will involve rethinking and redesigning how we handle the introductions and summaries of levels for players, and we will need to increase our emphasis on supports for connecting intuitive and formal understandings within the levels themselves. We observed similarities across the implementations with respect to the tendency of players to skim or skip the text in the introductory and summary screens. This underscores the importance of scaffolding concepts, terminology, and feedback "just in time" as discussed by Gee (2003, 2007). We also noted that players tend to streamline play goals and obstacles such that very careful alignment between play goals and obstacles and learning goals is critical to the success of these games from a learning perspective.

5.2. Comparison of learning gains across countries

As reported, the matched-pairs *t*-test of gains across the full test was significant for the students as group, $t(250) = 2.0792$, p (one-tailed) $= 0.019$. Effect size and power are modest at 0.1066 and 0.211, respectively. In addition, we found it interesting to note that, for a specific sub-set of items, students from both groups showed some increase in scores. Fig. 7 provides a comparison of gains across the three items that evidenced substantial gains for both countries.

Using software, we simulated this phenomenon using a hypergeometric distribution, considering several sets of parameters to account for differing assumptions, the discussion of which is not central to the topic of this paper. Depending on which set of parameters is used, the probability of these three assessment items emerging randomly from the two studies is $0.03 < p < .11$. Although it is possible that this is a random effect, an examination of the assessment items in question indicates why this might not be the case (Fig. 4a and b). These items ask students to perform a similar task: to predict the motion of an object when acted upon by a new force. This task of internal simulation is remarkably similar to the process that a player must perform in order to effectively play SURGE, i.e., to guide the SURGE ship through a maze by applying a series of forces. An additional feature shared by these assessment items is the presence of elements reminiscent of the control scheme used in SURGE, or a clear analog thereof.

For these reasons, we believe that increased student performance on these particular items may be a function of how the items themselves address the physics concepts in a more embodied way than other items which are conceptually similar but presented differently; under this framework, students who play SURGE would develop an internal physics model that is that is more contextually and conceptually embodied (see Gee, 2003, c.f. Wilson, 2002, for an overview on embodied cognition and its theoretical implications for learning) than what general instruments such as the FCI tend to detect. Part of our current effort is now devoted to developing an instrument that places greater emphasis on the embodied cognitive elements of SURGE, those that students seem to access according to these results.

Considering the differences across countries, and in curricular knowledge of physics between the groups, the existence of a stable sub-set of assessment items in which we see improved scores is very encouraging; it indicates a fundamental match between the way the students engage and reflect on the principles of physics presented in SURGE and the manner in which these principles are assessed by our FCI-based instrument, without an overriding dependence on frameworks that are dependent on cultural context. This idea is further reinforced by the

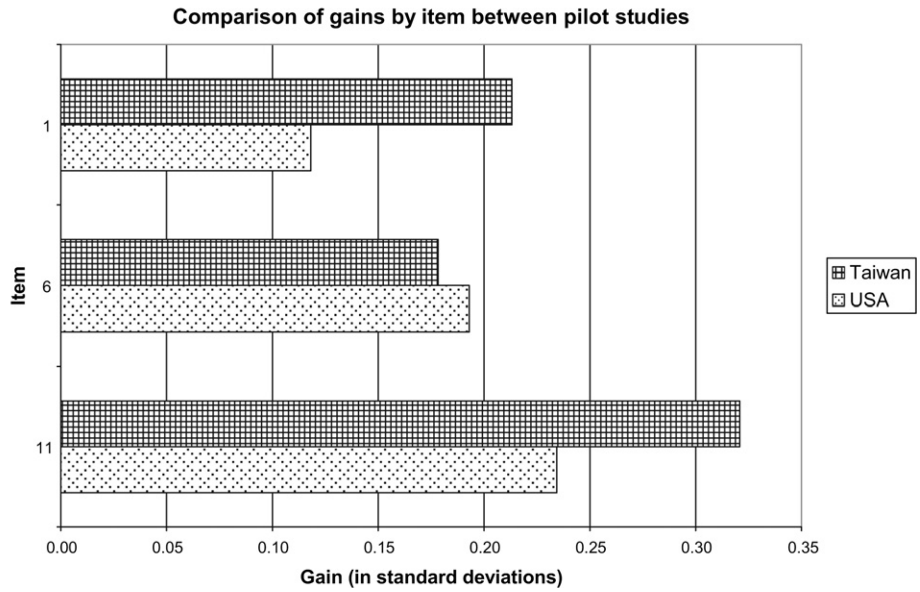


Fig. 7. Comparison of gains in a sub-set of assessment items between the Taiwan and USA studies.

observed trend of both groups to generally disregard textual (i.e., language-based) content, and that, due to technical issues, some students in the Taiwan group had no access to the text of the game at all. Our contention is not, however, that students were adopting a culturally-neutral stance while playing SURGE (see Sections 5.3 and 5.4 for our discussion of these stances), but rather, that their learning was not contingent on specific cultural cues.

5.3. Comparison of game play experience/habits across countries

The survey questions used in the SURGE assessment for the Taiwan and U.S. implementations gave us an enlightening view into the differences and similarities of these two groups. One interesting point of comparison is that both these groups present almost identical "gaming profiles": the proportions of students who play video games for a certain amount of time per week is very similar across groups, both aggregate and by gender (Fig. 6). These profiles are typical for American adolescents (Cummings & Vandewater, 2007), and while we did not preselect students on any gaming experience criteria, it is helpful for our research that these findings are not predicated on whether students have highly-evolved or developing gaming templates and vocabularies.

The distribution of genders is also worth noting. Both groups show a gender difference at the extreme ends of the gaming spectrum: in both groups, respondents who say they almost never play video games are most likely to be girls, whereas the most avid video game players are likely to be boys. One of our initial concerns regarded whether or not gender would be an important factor for learning when students played SURGE, both by itself and as a predictor of a more evolved gaming vocabulary. There is a perception that the video game medium as a whole has a bias toward males (Cassell & Jenkins, 1998; Kafai, Heeter, Denner, & Sun, 2008), so we would not have been surprised to see a gender difference in learning gains or attitudes toward SURGE. However, in neither implementation was gender a significant factor affecting performance in the assessment.

5.4. Comparison of perceptions and reactions toward SURGE across countries

In terms of important affective similarities, the majority of players across ages, genders, levels of game playing experience and cultures in both implementations found the SURGE enjoyable and engaging (Table 1). The high levels of engagement appeared to go beyond a "novelty effect" in that players' engagement remained high from beginning to end of the SURGE game. This is a notable finding because one might be justifiably concerned that embedding science learning curricula in digital games might be enjoyable or engaging only to subsets of the students, which would not be acceptable for digital games intended to be adopted as part of a standard science curricula.

At the same time, however, we saw common themes across implementations that demonstrate some of the challenges that come with embedding science learning in casual computer games. First, students in both implementations demonstrated varied levels of what we label "mindful play". While some students, particularly those in the Taiwan implementation, carefully read through the objectives, storyline, and ongoing feedback messages designed to support learning, larger numbers of players appeared to ignore these elements in their enthusiasm to play the levels. Similarly, there was a shared lack of interest in the scoring systems we designed into the game among players in both implementations, although some players in both implementations focused intensively on the medal award system, trying to earn higher (i.e., silver and gold) medals. Each of these findings may relate to the overall high level of engagement seen in game play, and demonstrate both the promise and challenges inherent in using casual computer games as a platform for science learning. The high level of motivation seen by all students reinforces the idea that games can engage a broad spectrum of learners from multiple cultural backgrounds and interests. However, the observed engagement centered primarily on the real-time game play in SURGE, to the detriment of various visual scaffolds, feedback messages, and reward systems aimed at moving players beyond tacit understanding of the physics concepts underlying

Table 1
Perceptions and reactions toward SURGE in Taiwan and U.S. Implementations.

Observation	Taiwan	US
High Engagement	• High engagement for all students • Novelty effect promoted engagement • But engagement lived beyond "novelty effect": students stayed engaged throughout • Design features seem to contribute to continued engagement • Student-led competition was a motivating factor	• Students visibly and audibly excited during game play • Student-led competition was a motivating factor
Varied Completion rates	• Wide range: some finished very quickly, others took much longer • Computer game familiarity: some students took more time to master the controls of the game • "Mindful" play slows completion time • Teacher class management impacts completion rates	• Wide range: some finished very quickly, others took much longer • Computer game familiarity: some students took more time to master the controls of the game • "Mindful" play slows completion time – most frequently by female students but also by male students with less experience playing digital games. • Teacher class management impacts completion rates
SURGE design impacting performance	• Velocity cross: positive for learning • Stabilize ship function: mechanics • Medal reward system: motivation • Point system: not valued by players • Blurry Chinese text hindered learning	• Medal reward system: motivation for some • Point system: not valued by players • Stabilize ship function: overuse? • Gate challenges: confusing to some students
Varied levels of "Mindful Play"	• Less 'mindful' play results in quicker completion of the game • "Mindful" players read screen info and went more slowly • Blurry Chinese text reduced opportunity for mindful play	• Many students don't read on-screen instructions, narrative, or feedback • Main goal for some was to get through quickly • Self-created game goals unrelated to intended goals

the game. The continuing challenge in SURGE, and in similar game-based learning projects, is to tightly integrate central game play goals with the need for and use of game mechanics and design elements that directly impact learning.

Another common theme seen across implementations that presents a challenge to the use of computer games as platforms for science instruction is that there was a wide range of "time to completion" among students. Some students finished all the levels in SURGE very quickly, relative to the rest of the students. From our observations, it appeared that these players fell into two groups: those who largely ignored the narrative, feedback, and scoring elements of the game, and/or those who seemed to be highly proficient "gamers". Students who took longer to finish the game levels were those who spent more time reading the on-screen information and/or those who needed more time to master the mechanics of the game. In any instructional intervention (not just ones involving games), there will be variation in completion time. In SURGE, we are working to include design elements that will help reduce 'extraneous' variation in completion rates related to computer gaming experience and less 'mindful' play. To better support students with less gaming experience, we can include a more heavily structured introduction level to the game that offers a kind of pre-training for students who need it on the basic mechanics of the game. To encourage more mindful play, we are working, as described previously, on more tightly integrating the narrative, feedback messages, and rewards system into the central elements of game play to make them relevant and central to success in the game.

With regard to students' affective response to SURGE, the question of gender is more nuanced. Students in both countries appeared to misjudge the intended audience of the game: boys in Taiwan liked the game better than girls, but the Taiwan group as a whole believed the game was better suited for girls; whereas girls in the US study were more positive toward SURGE than the boys, and the US group was of the opinion that the game was more suitable for boys. It is as if each subgroup (the Taiwan girls and the U.S. boys) expressed a stronger enjoyment of the game precisely because it is not the type of game the larger group believes they (that subgroup) would enjoy (see Pelletier, 2008, for an analysis on gender construction through expressed gaming preferences). Also noteworthy is that it is not a single gender (boys or girls) across both groups that are expressing what they perceive as a counter-cultural preference, but rather boys *and* girls within different groups. This suggests that while students may be truly perceiving the game as gender-neutral (as evidenced by their survey responses), they are accessing some differing cues from their respective cultures that are informing their beliefs about what kinds of games they "should" enjoy, and their opinions express whether or not they agree with that assessment. A full analysis of the country-level cultural differences between Taiwan and the US that may impact perceptions of, and success with, the SURGE game is beyond the scope of the current study. However, we believe this is a rich area for future research around educational game use in formalized learning environments.

5.5. Differences in gaming experience, templates, and vocabularies within countries

From our observations, players differ substantially in terms of the game templates and game "vocabularies" that players bring with them into the game. In some cases, this prior experience provides affordances for learning and in others it provides obstacles. Essentially, the experiences, templates, and vocabularies that players bring from prior games can be used to orient and motivate them in SURGE (or other games designed for academic learning) or can produce complications and confusion when our goals or mechanics deviate from these experiences, templates, and vocabularies. Differences in these prior experiences seem to overshadow differences in gender or country. In terms of affordances, there are many game design elements and goals that we can incorporate that the vast majority of students seem to intuitively understand immediately. When our designs can mirror those templates, our path is made much easier in orienting the player in how to proceed. Similarly, students quickly grasp the ideas and motivation of engaging in a game.

When our pedagogical goals do not parallel standard templates and design vocabularies, however, we need to help students operate in ways that diverge from their experiences, and we encounter greater challenges than if the players had not played games at all. Our impulse controls, for example, diverge from the control schemes used in most games. Most games involve holding down a movement key to produce a constant velocity in that direction. A few games involve holding down a movement key to produce a constant acceleration in that direction. Our impulse controls involve discrete pushes of a movement key to apply discrete acceleration bursts in that direction. Most players initially hold down the keys in the face of no effect because they are accustomed to the ideas of holding down keys to move. At a higher level, players are accustomed in many games for the real goal to involve quickly getting to the end of the levels, and for this goal to overshadow other goals of performance along the way.

Complicating this issue is the fact that some players bring more extensive game backgrounds with more robust sets of templates and vocabularies for thinking about games. These players are more confident as they approach learning in SURGE, with more robust trial and error strategies for solving obstacles they encounter. They also are less prone to encountering frustration, both because of their greater prior experience facilitating progress in the game as well as their self-perception of their competence in games, which buoys them through challenges that may arise. These same players, however, also tend to view SURGE as a game and seem less likely to reflect on what they are learning. Less experienced game players, particularly girls who have less experience playing digital games, often seem more likely to approach the challenges in SURGE more methodically and consciously than these more experienced players, but also seem more likely to withdraw from the experience or to be intimidated by the experience because they don't label themselves or self-identify as "gamers" and thus can feel out of their element. While our quantitative findings show no differences in learning outcomes based on reported gaming experience, we will continue to watch for more subtle relationships between players' prior gaming experience and their learning experiences. We are only just beginning to study these issues in SURGE, but they may prove the most important lessons to learn for our own project and the field.

6. Conclusions

In conclusion, we now synthesize the findings in light of our two driving questions. We first discuss the degree to which learning in SURGE reflects academically desirable learning. We then discuss the relative similarities and differences observed in relation to the degree to which SURGE seems portable between the two countries. We then close with some thoughts about the implications of our findings in terms of the potential of conceptually-integrated digital games for science learning.

6.1. Can a conceptually-integrated digital game that overlays popular game-play mechanics with formal physics representations and terminology support explicit learning about Newtonian mechanics as demonstrated through post-test measures based on assessment items from the formal discipline (i.e., the force concept inventory)?

To what degree does the learning in SURGE reflect academically desirable learning? At the most concrete level, early versions of SURGE seem to support learning on some core items of the Force Concept Inventory, an assessment instrument developed by university physicists as measuring important conceptual understanding of Newtonian mechanics. Thus players are making progress on some challenging core concepts identified by disciplinary experts. Progress thus far has not been as extensive as we might have hoped, but considering that entire university level calculus-based courses do not necessarily help students increase their scores on the FCI as dramatically as instructors might hope (e.g., Hestenes & Halloun, 1995), we feel that any measurable increases for middle school students after roughly an hour playing SURGE are fairly impressive. Essentially, the Force Concept Inventory was developed by physicists for undergraduate physics students. Undergraduate students who have completed entire traditional physics courses at the undergraduate level often don't perform that much better on the post-tests than on the pre-tests (e.g., Hestenes & Halloun, 1995). To assume that we could change students' performance dramatically on FCI items, particularly middle school students' performance, in an hour of game play was probably a bit over-ambitious in hindsight, but we are encouraged by the progress we have made, particularly in light of the early stages of development of SURGE.

Following this line of reasoning, we think it would be beneficial for future work with SURGE and similar games to develop measures more sensitive to incremental changes in student understanding. For example, the FCI questions related to interpreting positional dot traces require a student to (a) interpret the velocity at various time points in one dot trace, (b) interpret the velocity at various points of time on a second dot trace, and (c) make relative comparisons or syntheses of those sets of information. Rather than having all of the assessment items represent this level of complexity and cognitive load, it would seem beneficial to include a range of item difficulties to increase the value of the assessment in measuring a range of competencies and understandings. In the case above, we need complex assessment items like the FCI item described above, but we also need simpler items, such asking students to interpret a single positional dot trace. We thus are currently developing assessment items similar to those on the FCI but that represent a more gradual range of item difficulties while still focusing on the core concepts measured by the FCI (e.g., some items that focus on interpretation of a single dot trace representation). Creating a range of item difficulties in this manner would allow more sensitive measurement of incremental progress in students' understanding of disciplinary concepts in light of the shorter curricular time frames of most digital games.

Similarly, we would like to create items that are more sensitive to students' ability to explicitly connect intuitive understanding and formal understanding in their answers. As discussed earlier, the FCI items upon which we based our assessment tend to focus more on intuitive understandings. The FCI was designed purposefully in this manner as a conceptual test to avoid the chance of testing rote knowledge, but we would like to explore the potential for assessments that explicitly measure students' ability to explicitly connect intuitive understanding and formal understanding in their answers as part of our work.

In addition to creating more sensitive assessments, the results of this study also suggest that SURGE needs to provide more extensive supports for students to help them articulate their intuitive understandings from game play with the explicit formal concepts and representations of the discipline. Squire et al. (2004) found that their learning gains were dependent on the teacher developing curriculum outside of the game to focus students on these connections. The versions of SURGE in the implementations described in this paper focused on overlaying the game play mechanics with formal representations and terminology, and resulted in some gains, but more structured supports to focus students on the phenomena of interest and their relationship to the formal representations and concepts appears warranted. Current development in SURGE is focusing on integrating just-in-time supports (Gee, 2003, 2007), visual signaling (Mautone & Mayer, 2001; Nguyen & Meltzer, 2003), and metacognitive supports for prediction and explanation (Champagne et al., 1982; Chi & VanLehn, 1991; Tao & Gunstone, 1999) into the fabric of game play without destroying the "game-ness" of the game play. The next rounds of implementations will provide insights on the efficacy of these additional design strategies.

Beyond conceptual learning, SURGE also engages students in learning in ways that parallel calls for increased use of scaffolded exploration and inquiry in science. While digital learning games like SURGE do not represent typical inquiry learning environments, they do engage students in learning about core science concepts (in this case Newtonian mechanics) through exploration, prediction, and observation rather than through rote learning. A growing body of research and scholarship on games and cognition emphasizes cycles of prediction, observation, and refinement as core mechanics of game play processes (e.g., Salen & Zimmerman, 2003; Squire et al., 2003; Wright, 2006). Conceptually-integrated games like SURGE offer opportunities to further accentuate these inquiry-related aspects of game play, which is a driving consideration in the design of the metacognitive supports for prediction and explanation alluded to in the paragraph above, and which builds on a line of research on prediction, observations, and explanation from the science education and psychology literatures (e.g., Champagne et al., 1982; Chi & VanLehn, 1991; Tao & Gunstone, 1999). Although SURGE is still in its infancy, we feel that the results and lessons learned thus far bode well for the potential for games like SURGE to support learning of core science ideas through inquiry-related processes of exploration.

6.2. How similar or different are the learning and affective experiences of students playing the game in the two countries (i.e., Taiwan and the United States)?

With respect to our second question, we see remarkable levels of similarity between student reactions, outcomes, and gaming habits in each country. Across both implementations, students demonstrated high levels of motivation and engagement in playing SURGE, throughout the implementations. This is a positive finding that echoes results seen in may other education games studies (e.g., Nelson et al., 2005; Barab et al., 2007, 2009; Galas, 2006). On a more challenging note, we observed that students within both countries showed wide variation in the amount of time it took them to complete the game, with some students finishing very quickly and others taking much longer. Also, we witnessed varying degrees of 'mindful' play across implementations, with some students focusing on the text and visual scaffolds included in the game and many others paying very little attention to these details. One design detail that did seem to draw the attention for large numbers of students in both implementation sites was the medal reward system. This very similar extrinsic reward tool was popular with many participants. Our future game revisions will include an expanded role for the medal system, centered on designing the medal levels to more closely match specific components of physics knowledge and application of that knowledge in the game.

That these commonalities and differences in approaches to game play did not result in strong differences in learning is notable. The two groups are different enough in cultural terms that such differences could be expected (see, for example, the diametrically opposed gendered response between the groups about whether SURGE is more of a game for boys or girls), but our data regarding learning gains does not support this view. It may be that, while students play SURGE from perspectives that are largely culturally-specific, their learning is based more on the context that SURGE provides, which is common for both groups. In other words, the students regarded SURGE through divergent cultural lenses, but these lenses did not significantly modify their experience of play of the game, nor did it preempt their learning on the specific concepts that SURGE highlights.

Overall, the findings of this study suggest that conceptually-integrated games like SURGE should be fairly portable across multiple countries, which makes the development time and cost more attractive in light of the potentially larger audiences of students who might be served through these digital learning games. Future studies and work by our group will explore potential cultural explanations for these similarities and differences.

6.3. Final thoughts

The short take-home message is that games can be designed to teach science concepts by integrating the science concepts within the fabric of the game mechanics (i.e., conceptually-integrated games). This requires very careful design, however, as suggested by the findings of this study and others such as the *Enigmo* study (Masson et al., 2011) and work on *Supercharged* (Squire et al., 2004). Most games may support students in developing an intuitive understanding of the physics involved, for example, in order to eventually be able to "beat" the game, but without specific scaffolds the players won't necessarily learn anything consciously or explicitly about the physics in a formal sense. This isn't surprising considering that people don't learn formal physics simply through playing soccer or similar physics-based games in the "real" world. If those same players could be scaffolded in connecting formal concepts and representations to their experiences as they played soccer in a manner that helped them articulate those connections, however, they might well learn about formal physics by playing soccer in that context. The challenge in physical settings, however, is in overlaying and connecting those representations and ideas. This is where digital games like SURGE may provide significant affordances in terms of providing scaffolding, signaling, and overlays to make these connections explicit. Our future work and the future work of other groups will continue to explore these possibilities, particularly in light of their potential to support students across multiple countries engaging in exploration of challenging core concepts like Newtonian mechanics.

Acknowledgement

This research was funded by the United States National Science Foundation DR-K12 Award #0822370 (Scaffolding Understanding by Redesigning Games for Education).

References

Aikenhead, G. S., & Jegede, O. J. (1999). Cross-cultural science education: a cognitive explanation of a cultural phenomenon. *Journal of Research in Science Teaching, 36*, 269–287.

Aikenhead, G. S., & Otsuji, H. (2000). Japanese and Canadian science teachers' views on science and culture. *Journal of Science Teacher Education, 11*(4), 277–299.

Anderson, J., & Barnett, G. M. (2010). Using video games to support pre-service elementary teachers learning of basic physics principles. *Journal of Science Education and Technology*, Online First: http://dx.doi.org/10.1007/s10956-010-9257-0.

Annetta, L. A., Minogue, J., Holmes, S. Y., & Cheng, M. (2009). Investigating the impact of video games on high school students' engagement and learning about genetics. *Computers & Education, 53*(1), 74–85.

Barab, S. A., Arici, A., & Jackson, C. (2005). Eat your vegetables and do your homework: a design based investigation of enjoyment and meaning in learning. *Educational Technology, 45*(1), 15–20.

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

Barab, S. A., Zuiker, S., Warren, S., Hickey, D., Ingram-Goble, A., Kwon, E.-J., et al. (2007). Situationally embodied curriculum: relating formalisms and contexts. *Science Education, 91*(5), 750–782.

Barnett, M., Squire, K., Higginbotham, T., & Grant, J. (2004). Electromagnetism supercharged!. In *Proceedings of the 2004. International conference of the learning sciences* Los Angeles: UCAL Press.

Cassell, J., & Jenkins, H. (1998). *From Barbie to Mortal Kombat: Gender and computer games*. illustrated edition. The MIT Press.

Champagne, A. B., Klopfer, L. E., & Gunstone, R. F. (1982). Cognitive research and the design of science instruction. *Educational Psychologist, 17*(1), 31.

Chang, H.-Y., Quintana, C., & Krajcik, J. (2010). The impact of designing and evaluating molecular animations on how well middle school students understand the particulate nature of matter. *Science Education, 94*(1), 73–94.

Chi, M. T. H., & VanLehn, K. A. (1991). The content of physics self-explanations. *Journal of the Learning Sciences, 1*(1), 69–105.

Clark, D.B., & Martinez-Garza, M. (in press). Prediction and explanation as design mechanics in conceptually-integrated digital games to help players articulate the tacit understandings they build through gameplay. C. Steinkuhler, K. Squire, & S. Barab (Eds.), Games, learning, and society: Learning and meaning in the digital age. Cambridge: Cambridge University Press.

Clark, D.B., Nelson, B. C, D'Angelo, C.M., Slack, K. & Martinez-Garza, M., (2010). SURGE: Integrating Vygotsky's Spontaneous and Instructed Concepts in a Digital Game. *Proceedings of the ninth international conference of the learning sciences*, 384–385. Chicago, IL.

Clark, D.B., Nelson, B.C., D'Angelo, C.M., Slack, K., & Menekse, M., Martinez-Garza, M. (2010). Comparing the impact of overlaying physics-based video games with formal physics representations in Taiwan and the United States. Proceedings of the national association of research in science teaching (NARST) 2010 meeting. Philadelphia, Pennsylvania.

Clark, D.B., Nelson, B. C, Sengupta, P., D'Angelo, C. M. (2009). Rethinking science learning through digital games and simulations: genres, examples, and evidence. Paper commissioned for the national research council workshop on games and simulations. Washington, D.C. http://www7.nationalacademies.org/bose/Gaming_Sims_Commissioned_Papers.html

Clark, R., Nguyen, F., & Sweller, J. (2006). *Efficiency in learning evidence-based guidelines to manage cognitive load*. San Francisco: Pfeiffer.

Clarke, J., & Dede, C. (2005). Making learning meaningful: an exploratory study of using multi-user environments (MUVEs) in middle school science. *Paper presented at the American Educational Research Association Conference, Montreal, Canada*.

Coller, B., & Scott, M. (2009). Effectiveness of using a video game to teach a course in mechanical engineering. *Computers & Education, 53*(3), 900–912.

Cummings, H. M., & Vandewater, E. A. (2007). Relation of adolescent video game play to time spent in other activities. *Arch Pediatr Adolesc Med, 161*(7), 684–689.

D'Angelo, C.M. (2010). Scaffolding vector representations for student learning inside a physics game. Unpublished doctoral dissertation. Arizona State University.

D'Angelo, C.M., Clark, D.B., Nelson, B.C., Slack, K., & Menekse, M. (2009). The effect of vector representations on students' understanding of motion. Poster presented at the Physics Education Research Conference (PERC)/American Association of Physics Teachers (AAPT) 2009 meeting. Ann Arbor, Michigan.

Dede, C., & Ketelhut, D. J. (2003). Designing for motivation and usability in a museum-based multi-user virtual environment. *Paper presented at the American Educational Research Association Conference, Chicago, IL*.

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

Edelson, D. C., Gordin, D. N., & Pea, R. D. (1999). Addressing the challenges of inquiry-based learning through technology and curriculum design. *Journal of the Learning Sciences, 8*(3/4), 391–450.

Edelson, D. C., Salierno, C., Matese, G., Pitts, V., & Sherin, B. (2002). Learning-for-use in earth science: Kids as climate modelers. *Paper presented at the Annual Meeting of the National Association for Research in Science Teaching, New Orleans, LA, April 2002*.

Federation of American Scientists. (2006). *Harnessing the power of video game for learning*. Retrieved September 1, 2010 from. http://fas.org/gamesummit/.

de Freitas, S. D., & Neumann, T. (2009). The use of 'exploratory learning' for supporting immersive learning in virtual environments. *Computers & Education, 52*(2), 343–352.

Galas, C. (2006). Why Whyville? *Learning and Leading with Technology, 34*(6), 30–33.

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

Gee, J. P. (2004). *Situated language and learning: A critique of traditional schooling*. London: Routledge.

Gee, J. P. (2007). *Good video games and good learning: Collected essays on video games, learning, and literacy*. New York: Peter Lang.

George, J., & Glasgow, J. (1988). Street science and conventional science in the West Indies. *Studies in Science Education, 15*, 109–118.

Harel, I., & Papert, S. (1991). Software design as a learning environment. In I. Harel, & S. Papert (Eds.), *Constructionism*. Norwood, NJ: Ablex.

Hestenes, D., & Halloun, I. (1995). Interpreting the FCI. *The Physics Teacher, 33*, 502–506.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher, 30*, 141–158.

Hickey, D., Ingram-Goble, & Jameson, E. (2009). Designing assessments and assessing designs in virtual educational environments. *Journal of Science Education and Technology, 18*(2), 187–208.

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

Hofstede, G. (2001). *Cultures consequences: Comparing values, behaviors, institutions and organizations across nations* (2nd ed.). Thousand Oaks, CA: Sage.

Hofstede, G. (2008). Cultural differences in teaching and learning. *Paper presented at FUHU conference on Education and Training in the Multicultural Classroom*. Copenhagen. (http://fuhu.dk/filer/FBE/Arrangementer/Denmark%20Unlimited%20080508/FBE_geert_hofstede_teaching_learning.pdf).

Holbert, N. (2009). Learning Newton while crashing cars. *Poster presented at Games, Learning and Society 2009, Madison, WI*, June, 10–12.

Holbert, N. R., & Wilensky, U. (2010). FormulaT Racing: combining gaming culture and intuitive sense of mechanism for video game design. In K. Gomez, L. Lyons, & J. Radinsky (Eds.), *Learning in the disciplines: Proceedings of the 9th international conference of the learning sciences*. *Short papers, Symposia, and Selected Abstracts, vol. 2* (pp. 268–269). Chicago, IL: International Society of the Learning Sciences.

Jackson, J. (2007). Simplified force concept inventory. Unpublished assessment. Arizona State University.

Jenkins, H., Squire, K., & Tan, P. (2004). You can't bring that game to school! Designing supercharged! In B. Laurel (Ed.), *Design research* Cambridge, Massachusetts: MIT Press.

Kafai, Y. B., Heeter, C., Denner, J., & Sun, J. Y. (2008). *Beyond Barbie® and Mortal Kombat: New perspectives on gender and gaming*. The MIT Press.

Kafai, Y. B., Quintero, M., & Feldon, D. (2010). Investigating the 'why' in Whypox: casual and systematic explorations of a virtual epidemic. *Games and Culture, 5*(1), 116–135.

Ketelhut, D.J., Dede, C., Clarke J., & Nelson, B. (2006). A multi-user virtual environment for building higher order inquiry skills in science. Paper presented at the 2006 AERA Annual Meeting, San Francisco, CA, 7 to 11 April 2006; Available at http://muve.gse.harvard.edu/rivercityproject/documents/rivercitysympinq1.pdf.

Ketelhut, D. J., & Schifter, C. C. (2010). Teachers and game-based learning: improving our understanding of how to increase efficacy of adoption. *Computers & Education*.

Klopfer, E., Osterweil, & Salen. (2009). *Moving learning games forward. The education arcade*. Retrieved on September 01.09.09 from. Massachusetts Instiute of Technology. http://www.educationarcade.org/.

Klopfer, E., & Purushotma, R. Using Simulations as a Starting Point for Constructing Meaningful Learning Games. J. Fromme, & A. Unger (Eds). *Computer games/players/game cultures: A handbook on the state and perspectives of digital game Studies*. Springer, in press.

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

Lee, O. (2005). Science education and student diversity: summary of synthesis and research agenda. *Journal of Education for Students Placed at Risk, 10*(4), 431–440.

Lee, O., & Luykx, A. (2007). Science education and student diversity: race/ethnicity, language, culture, and socioeconomic status. In S. K. Abell, & N. G. Lederman (Eds.), *Handbook of research on science education*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Lindgren, R., & Schwartz, D. L. (2009). Spatial learning and computer simulations in science. *International Journal of Science Education, 31*(3), 419–438.

Masson, M. E. J., Bub, D. N., & Lalonde, C. E. (2011). Video-game training and naive reasoning about object motion. *Applied Cognitive Psychology*, Advance online publication.

Mautone, P., & Mayer, R. (2001). Signaling as a cognitive guide in multimedia learning. *Journal of Educational Psychology, 93*(2), 377–389.

Mayer, R. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press.

McQuiggan, S., Rowe, J., & Lester, J. (2008). The effects of empathetic virtual characters on presence in narrative-centered learning environments. In *Proceedings of the 2008 SIGCHI Conference on Human Factors in Computing Systems*, Florence, Italy, pp. 1511–1520.

Moreno, R., & Mayer, R. E. (2000). Engaging students in active learning: the case for personalized multimedia messages. *Journal of Educational Psychology, 92*, 724–733.

Moreno, R., & Mayer, R. E. (2004). Personalized messages that promote science learning in virtual environments. *Journal of Educational Psychology, 96*, 165–173.

Mullis, I. V. S., Martin, M. O., Foy, P., & (with Olson, J.F., Preuschoff, C., Erberber, E., Arora, A., & Galia, J.). (2008). *TIMSS 2007 international science report: Findings from IEA's trends in international mathematics and science study at the fourth and eighth grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Nasir, N., Rosebery, A., Warren, B., & Lee, C. (2006). Learning as a cultural process: achieving equity through diversity. In K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 489–504). Cambridge, England: Cambridge University Press.

Nelson, B. (2007). Exploring the use of individualized, reflective guidance in an educational multi-user virtual environment. *Journal of Science Education and Technology, 16*(1), 83–97.

Nelson, B., Erlandson, B., & Denham, A. (2010). Global channels for learning and assessment in complex game environments. *British Journal of Educational Technology*, Published online, Jan. 2010. To appear in print, June 2010.

Nelson, B., Ketelhut, D., Clarke, J., Bowman, C., & Dede, C. (2005). Design-based research strategies for developing a scientific inquiry curriculum in a multi-user virtual environment. *Educational Technology, 45*(1), 21–34.

Neulight, N., Kafai, Y. B., Kao, L., Foley, B., & Galas, C. (2007). Children's participation in a virtual epidemic in the science classroom: making connections to natural infectious diseases. *Journal of Science Education and Technology, 16*(1), 47–58.

Newton, P., Driver, R., & Osborne, J. (1999). The place of argumentation in the pedagogy of school science. *International Journal of Science Education, 21*(5), 553–576.

Nguyen, N.-L., & Meltzer, D. E. (2003). Initial understanding of vector concepts among students in introductory physics courses. *American Journal of Physics, 71*(6), 630–638.

Papert, S. (1980). *Mindstorms: Children, computers and powerful ideas*. New York, NY: Basic Books.

Pelletier, C. (2008). Producing difference in studying and making computer games: how students construct games as gendered in order to construct themselves as gendered. In Y. Kafai, C. Heeter, J. Denner, & J. Sun (Eds.), *Beyond Barbie and Mortal Kombat: New perspectives on gender, games and computing*. Cambridge MA: MIT Press.

Raghavan, K., & Glaser, R. (1995). Model-based analysis and reasoning in science: the MARS curriculum. *Science Education, 79*(1), 37–61.

Roschelle, J. (1991). *Students' construction of qualitative physics knowledge: Learning about velocity and acceleration in a computer microworld*. Unpublished doctoral dissertation, University of California, Berkeley.

Roschelle, J., & Teasley, S. (1995). The construction of shared knowledge in collaborative problem solving. In C. O'Malley (Ed.), *Computer supported collaborative learning*. Berlin: Springer-Verlag.

Rosebery, A., Warren, B., Ballenger, C., & Ogonowski, M. (2005). The generative potential of students' everyday knowledge in learning science. In T. Romberg, T. Carpenter, & F. Dremock (Eds.), *Understanding mathematics and science matters* (pp. 55–80). Mahwah, NJ: Erlbaum.

Roth, K. J., Druker, S. L., Garnier, H., Lemmens, M., Chen, C., Kawanaka, T., et al. (2006). *Teaching science in five countries: Results from the TIMSS 1999 video study*. Washington, D.C.: National Center for Education Statistics. NCES Number: 2006011.

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

diSessa, A. A. (1993). Toward an epistemology of physics. *Cognition and Instruction, 10*(2 & 3), 105–225.

Shaffer, P. S., & McDermott, L. C. (2005). A research-based approach to improving student understanding of the vector nature of kinematical concepts. *American Journal of Physics, 73*(10), 921–931.

Snively, G., & Corsiglia, J. (2001). Discovering indigenous science: implications for science education. *Science Education, 85*(1), 6–34.

Squire, K., Barnett, M., Grant, J.M., Higginbotham, T., (2004) Electromagnetism supercharged!: learning physics with digital simulation games. In *Proceedings of the 6th international conference on learning sciences*, p.513–520, Santa Monica, California.

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

Squire, K., Jenkins, H., Holland, W., Miller, H., O'Driscoll, A., Tan, K. P., & Todd, K. (2003). Design principles of next-generation digital gaming for education. *Educational Technology, 43*(5), 17–23.

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

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

Tao, P., & Gunstone, R. F. (1999). The process of conceptual change in force and motion during computer-supported physics instruction. *Journal of Research in Science Teaching, 36*(7), 859–882.

Turkle, S. (1997). Seeing through computers: education in a culture of simulation. *The American Prospect, 31*(March–April), 76–82.

Vygotsky, L. S. (1986). *Thought and language*. Abridged from 1934; A. Kozulin, Trans.. Cambridge, MA: MIT Press

Warren, B., Ballenger, C., Ogonowski, M., Rosebery, A., & Hudicourt-Barnes, J. (2001). Re-thinking diversity in learning science: the logic of everyday language. *Journal of Research in Science Teaching, 38*, 529–552.

Warren, B., & Rosebery, A. (2008). Using everyday experience to teach science. In A. Rosebery, & B. Warren (Eds.), *Teaching science to English language learners* (pp. 39–50). Arlington, VA: NSTA Press.

White, B. Y. (1993). ThinkerTools: causal models, conceptual change, and science education. *Cognition and Instruction, 10*(1), 1–100.

White, B., & Frederiksen, J. (1998). Inquiry, modeling, and metacognition: making science accessible to all students. *Cognition and Instruction, 16*(1), 3–118.

Wieman, C. E., Adams, W. K., & Perkins, K. K. (2008). PhET: simulations that enhance learning. *Science, 322*, 682–683.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9*(4), 625–636.

Wright, W. (2006). Dream machines. *Wired, 14*(04). http://www.wired.com/wired/archive/14.04/wright.html.