Learning the Rules of the Game: The Nature of Game and Classroom Supports When Using a Concept-Integrated Digital Physics Game in the Middle School Science Classroom

Phillip Michael Stewart, Jr.

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
and the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2013

ABSTRACT

Learning the Rules of the Game: The Nature of Game and Classroom Supports When Using a Concept-Integrated Digital Physics Game in the Middle School Science Classroom

Phillip Michael Stewart, Jr.

Games in science education is emerging as a popular topic of scholarly inquiry. The National Research Council recently published a report detailing a research agenda for games and science education entitled *Learning Science Through Computer Games and Simulations* (2011). The report recommends moving beyond typical proof-of-concept studies into more exploratory and theoretically-based work to determine how best to integrate games into K-12 classrooms for learning , as well as how scaffolds from within the game and from outside the game (from peers and teachers) support the learning of applicable science.

This study uses a mixed-methods, quasi-experimental design with an 8th grade class at an independent school in southern Connecticut to answer the following questions:

1. What is the nature of the supports for science content learning provided by the game, the peer, and the teacher, when the game is used in a classroom setting?
2. How do the learning gains in the peer support condition compare to the solo play condition, both qualitatively and quantitatively?

The concept-integrated physics game SURGE (Scaffolding Understanding through Redesigning Games for Education) was selected for this study, as it was developed with an ear towards specific learning theories and prior work on student

understandings of impulse, force, and vectors. Stimulated recall interviews and video observations served as the primary sources and major patterns emerged through the triangulation of data sources and qualitative analysis in the software QSR NVivo 9.

The first pattern which emerged indicated that scaffolding from within the game and outside the game requires a pause in game action to be effective, unless that scaffolding is directly useful to the player in the moment of action. The second major pattern indicated that both amount and type of prior gaming experience has somewhat complex effects on both the uses of supports and learning outcomes. In general, a high correlation was found between students who were more successful navigating supports from the game, the teacher, and the peer and higher gain scores from pre- to posttest. However, students with a lot of prior game experience that found the game to be easy without much assistance did not do as well from pre- to posttest as they did not need as much assistance from the game to do well and therefore missed out on important physics connections to impulse, force, and vectors. However, those students with little prior game experience did not find game scaffolds as useful and did not do as well from pre- to posttest without significant teacher and peer support to bolster or supplant the game's intended scaffolding.

Implications for educators, educational game designers, and games in science education researchers are presented. It is argued that teachers must find ways to extract those scaffolds from the game which are easy to miss or require failure to activate so that all students, even those who find the game easy, are exposed to the intended learning in the game. Ideally, game designers are encouraged to find new ways to present scaffolds such that players of any ability can benefit from the connections from the game to physics.

TABLE OF CONTENTS

Prefatory Pages

Chapters

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ACKNOWLEDGEMENTS

First I would like to thank my committee members, doctors O. Roger Anderson, Felicia Moore Mensah, Sandra Okita, Mary Putman, and Ann Rivet for their support with this dissertation. I would especially like to thank Dr. Anderson for putting up with my endless questions, and of course Ann Rivet for being my mentor through this complicated process: her probing theoretical insights and talent for spotting methodological weaknesses taught me more than I thought I could ever learn writing this thesis. I am forever grateful.

I would also like to thank the school that allowed access to their entire 8th grade class. You know who you are!

To my physics teachers, especially Angelina Winborne (the best high school physics teacher in America); Davidson College professors Dr. Mario Belloni, Dr. John Yukich, Dr. Dan Boye, Dr. Laurence Cain, and Dr. Tim Gfroerer; and my physics professor at Columbia, Dr. Fernand Brunchwig: all of your support and guidance throughout my physics life has led me to this moment and I credit you for cultivating my love of this field.

My friends and colleagues at Columbia have been tremendously helpful, not just in troubleshooting issues regarding the dissertation, but for helping me relax with laughs, lunches, and late afternoon coffees. I would especially like to thank Jess Mezei, Alissa Berg, and Jess Hammer for their interest, help, and invaluable advice. Jess Hammer's class on video games and education was the seed for all of this work, and I will never forget that.

My family deserves more thanks than can adequately fit on this page: my mother for her patience and support; my grandparents for their positivity and generosity of spirit; my brother Robert for making me laugh even on the darkest days; my sister Merritt for letting me complain; and my dog Rufus, who spent more time with me writing than anyone else.

Finally, I want to express the deepest love and gratitude to my wife, Ellie, for her taking on every role simultaneously: cheerleader, reality check, shoulder, breadwinner, editor, and best friend. I truly could never have done this without your boundless support and sacrifice. I love you more every day.

For my grandparents
Pauline (Woo) and Robert (Bobby) Johnson

For their love, patience, strength,
and endless support
throughout my entire education

> "Is it possible to suppose that a child's behavior is always guided by meaning, that a preschooler's behavior is so arid that he never behaves with candy as he wants to simply because he thinks he should behave otherwise? This kind of subordination to rules is quite impossible in life, but in play it does become possible; thus, play also creates the zone of proximal development of the child. In play, a child is always a head taller than himself. As in the focus of a magnifying glass, play contains all developmental tendencies in a condensed form; in play it is as though the child were trying to jump above the level of his normal behavior."
>
> –Vygotsky (1967), from *Play and its Role in the Mental Development of the Child*

> "The great game of science is modeling the real world, and each scientific theory lays down a system of rules for playing the game. The object of the game is to construct valid models of real world objects and processes…The main objective of science instruction should therefore be to teach the modeling game."
>
> -Hestenes (1992)

CHAPTER ONE

INTRODUCTION

Science education researchers in the United States have been embroiled consistently in a thorny conversation about *content*. Traditionally, classrooms focused full attention on the facts and figures of science rather than exposing students to the actual processes behind discovering those facts in the first place, but even from the early 20th century, thinkers questioned the emphasis on content in the school (e.g., Dewey, 1910). Indeed, current national standards for science education (e.g., National Research Council, 2012) eschew the traditional focus on content in favor of richer science experiences. In the report *Taking Science to School* (Duschl, Schweingruber, & Shouse, 2007, pp. 36-41), current perspectives on the best ways to learn science are distilled into four strands:

> Students who are proficient in science:
> 1. know, use, and interpret scientific explanations of the natural world;
> 2. generate and evaluate scientific evidence and explanations;
> 3. understand the nature and development of scientific knowledge; and
> 4. participate productively in scientific practices and discourse

State standards documents stem from these recommendations, but the reality of multiple-choice testing results in assessments which emphasize knowledge over both processing skills and conceptual understanding. On paper the goals for science education align with the current best thinking on science learning, but classroom approaches fostering these new goals lag behind. Well-designed digital games are uniquely suited to support science experiences which closely align with the *Taking Science to School* recommendations and hold much promise in the shift to providing a more robust science experience that moves beyond rote memorization of facts (Clark, Nelson, Sengupta, & D'Angelo, 2009).

Many fields have seen a spike in interest of using games in research, including but not limited to the military, medical science, organizational psychology, and of course education. With a wide array of fields also comes a variety of theoretical perspectives and research approaches, but the actual knowledge gleaned from thousands of studies has been relatively scant as the studies do not meet empirical standards (O'Neil, Wainess, & Baker, 2005). As well, most larger-scale research on games and science education is more focused on how games foster scientific inquiry rather than learning of the specific science concepts. Instead of continuing along the line of proof-of-concept and general exploration of much games and learning research, scholars recommend a "heavier emphasis on rigorous analysis of qualitative and quantitative data of what exactly is being learning, by whom, and how" (Clark et al., 2009, p. 54).

This study seeks to address the lack of classroom-based information regarding the use of games for physics education specifically by focusing on one game's use in an independent school in southern Connecticut. The concept-based learning game SURGE (Scaffolding Understanding by Redesigning Games for Education) was created by researchers at Vanderbilt and Arizona State University "to help students build on their tacit intuitive understanding of motion gained through engaging game play and translate that into the formalized explicit language and ideas used in school-based contexts" (D'Angelo, 2010). It was chosen for this study because it was developed from a robust theoretical framework and attends to many of the effective design characteristics the games and learning literature has identified.

This study aims to uncover the nature of game-based (mechanical) and classroom-based (cultural) supports when using a game in the classroom, as well as the interaction with the aesthetic game experience, vis-à-vis the formal definition of a game. A support is anything provided by an external entity that helps someone learn something, either indirectly by improving the conditions for learning, or directly by providing content or links to content. Game-based supports include the actual components of the game which connect game features to science explicitly, supports which increase student motivation, as well as the game's accompanying problem-based scaffolding (pen-and-paper examples and contextualized problems). Classroom-based supports include interactions with the teachers and with peers. Two conditions will be examined: one where students work independently with the game and one where students work in dyads supporting each other's play. Two questions guided this research:

1. *What is the nature of the supports for science content learning provided by the game, the peer, and the teacher, when the game is used in a classroom setting?*
2. *How do the learning gains in the peer support condition compare to the solo play condition, both qualitatively and quantitatively?*

There has been empirical work looking specifically at supports around science simulations in the classroom (e.g., Linn & Hsi, 2000), but not concept-based science games. There has been work looking at the learning potential of specific game-based scaffolds (e.g., D'Angelo, 2010; Sun, Wang, & Chan, 2011), but not in the classroom. There have also been studies around meaning-making while playing commercial off-the-shelf games with others (e.g., Abrams, 2010; Hung, 2008), but not in classrooms. This study aims to inform researchers in games in science learning, future learning game designers, and those who choose to use digital learning games in their classrooms. Researchers in games and learning are certainly interested in the sociocultural interactions surrounding play (e.g., Abrams, 2010; Hung, 2008; Juul, 2003; Salen & Zimmerman, 2003) and classroom-based research is needed (National Research Council, 2011). The second group, educational game designers, needs to know what kinds of game supports are the most powerful in helping students and teachers play the game most effectively and how those supports are used by students and teachers (or not). The final group, teachers and others using games for learning, need to know how best to support students in a classroom where a game is being used for learning. Teasing apart the way different levels of support interact with one another in the way the students play the game can help teachers support their students more effectively around play.

All classrooms planning to use games include the game and the teacher, but it is ultimately up to the instructor whether or not to put students in pairs when working at the computer. While little research exists on collaboration in single player games, research does exist on collaboration around computer simulations (Dillenbourg, Baker, Blaye, & Malley, 1996). Linn & Hsi (2000) report significantly higher learning gains from students working in dyads while using science simulations, and indeed many science educational games have emphasized collaborative game design components in open-world games like Quest Atlantis (Barab, Thomas, Dodge, Carteaux, & Tuzun, 2005), but collaboration has been all but ignored in studies of single player games, even though financial restraints keep most schools from having one computer for each child. This literature gap is important to fill because it is important to both explore the nature of peer support while playing single player games, and how adding a level of peer support changes the way the players engage with the learning game and with other forms of support. This study sits precisely at the nexus of theory and practice.

CHAPTER TWO

REVIEW OF THE LITERATURE

In reviewing the literature, I begin by taking a look at work that has been done in science education around supports and technological resources, especially simulations. Next, I present a working definition of digital games to distinguish them from simulations, and then look at how researchers have conceptualized learning in games. I push further to provide a rationale for using educational games to teach science specifically, including a review of the work that has been done on learning science content from games. Finally, I specifically describe how SURGE and the supports provided by the game link game learning to content learning, and conclude with a description of the need for research.

Supports, Scaffolding, and Technology Tools

Digital gamers in general do not read game manuals (Gee, 2007), which should not come as a surprise considering that most of us never read a rules document before we play a sport. Game players exist in a social community of gamers and share strategies in several ways: inside the game in Massively Multiplayer Online Games (MMORPG) like World of Warcraft, outside the game through forums and blogs (Steinkuehler & Duncan, 2008), and even side-by-side in play sessions. However, games are also learning machines (Gee, 2005) because talented game designers are very good at using scaffolding theory (Wood, Bruner, & Ross, 1976) to teach the rules of the game through play. Great games can be understood without any help outside the game, though outside help can make understandings of play more robust. The games and learning literature is currently investigating how we can design games that promote similarly autonomous learning of content through play via game supports. Before looking deeply into the supports that well-designed games provide players, a closer examination of the characteristics of scaffolding in general is in order.

Supports from Teacher, Peer, and Tool

Vygotsky's (1967) quotation that opens this thesis directly connects play to the idea of the Zone of Proximal Development [ZPD]: a liminal space where learners may venture to the periphery of their knowledge, conceptualized as the gap between what a learner can know with and without help. It is here that the teacher, or an entity with more knowledge or experience, can provide support to create new learning and allow the learner to increase his or her individual knowledge. The idea of the ZPD is central to many social constructivist theories of learning. Supports for learning are often called scaffolds and the idea of "scaffolding"[1] was first described by Wood, Bruner, and Ross (1976) in their work around tutoring; it is what we call that bridge between the ZPD and new learning.

Teacher Scaffolding

Wood et al. (1976, p. 98) delineate six types of scaffolding provided by an instructor to a learner:

> (1) *Recruitment:* This type of scaffold engenders interest or buy-in on the part of the learner, thereby improving the conditions for learning.
> (2) *Reduction in degrees of freedom:* This scaffold type places black-boxes on certain aspects of the task at hand, reducing the complexity of the problem and lowering the number of steps required to reach a final solution.
> (3) *Direction maintenance:* Learners can go off track, trying to achieve aims not intended by the problem at hand. This kind of scaffold keeps the learner on task.
> (4) *Marking critical features:* This scaffold highlights salient aspects of tasks while de-emphasizing extraneous aspects.

[1] For the sake of variety, supports and scaffolds should be considered interchangeable throughout the paper.

> (5) *Frustration control:* The presence of this scaffold limits frustration by providing guidance, support, or relief, thereby improving the conditions for learning.
> (6) *Demonstration:* This type of scaffold is actually modeling of the solution in some way, whether it be performing the solution fully or by taking a partial solution put forth by a learner and idealizing it in the hopes that the learner will imitate the solution back in a more normative way.

While the work of Wood et al. (1976) was specifically about supports provided by a tutor in a one-on-one session with learners using building blocks, they still subsume the types of scaffolding which can be provided by a teacher to his or her students. But supports need not be provided only by the teacher; they can come from specially designed cognitive tools.

Digital Cognitive Tools and Integrated Game Scaffolding

There has been much work in how digital resources can be turned into digital cognitive tools (Songer, 2007) by including scaffolding features with the resources; it is prudent to look at the work that has been done around supports using other technological representations, especially simulations. This research has been done with extensive longitudinal work in the science simulations literature by embedding representations such as animations and simulations into technology environments which provide scaffolding to support teachers and students in the use and analysis of the representations, like the Web-based Inquiry Science Environment (WISE: Linn, Clark, & Slotta, 2003). Design principles extracted from this work, which takes a knowledge integration perspective (Kali & Linn, 2007; Linn et al., 2003), are the guiding principles science education researchers go to when taking technology into the classroom or when designing a digital cognitive tool: (1) Make science accessible, (2) make thinking visible, (3) help learners learn from each other, and (4) promote autonomous lifelong learning. More than 70 individual supports for learning have been identified in the research under these umbrella meta-principles (Kali & Linn, 2007). A couple of examples of successful supports identified by the WISE work are presented in Table 2.1.

Table 2.1

Examples of successful support structures in WISE (Kali & Linn, 2007).

Meta Category	Example One	Example Two
Make science accessible	Using personally relevant examples	Authentic contexts or tasks
Make thinking visible	Knowledge representation tools	Enable 3D manipulation
Help learners learn from one another	Automated peer strategy sharing functionality	Learners are prompted to reflect on a dilemma, propose an initial solution, and collaborate around a revised, synthesis solution
Promote autonomous lifelong learning	Enabling manipulation of factors in models and simulations in suitably structured ways to scaffold use of increasingly sophisticated strategies	Showing clear interactions of variables through an elegant interface promotes autonomous experimentation and learning

While studies of commercial physics-based games have shown some clear results in terms of students gaining intuitive notions of force and motion, these notions are not formalized knowledge pieces about science content (Clark et al., 2011). The goal of SURGE is in the name itself: Scaffolding Understanding by Redesigning Games for Education. The game literally embeds scaffolds like commercial game designers, but instead of those scaffolds helping the player merely learn the game, they are also helping the player connect his or her physics notions to physics content. These game scaffolds come in two varieties: endogenous and exogenous. Exogenous scaffolds include the storyline of the game, the graphics, and the game structure itself--these are all in the interest of recruitment (Wood, Bruner, & Ross, 1976), or getting the buy-in of the player and their interest piqued. Endogenous scaffolds include game design elements which directly scaffold learning the game and, in the case of concept-based games, science content (explored in detail in the section on SURGE design characteristics later in this chapter).

Peer Support

In social constructivism, it is believed that task context, including the presence of classmates or working partners, interacts with the learning process. A classical Piagetian would not call someone of equal mind a potential scaffold, but Vygotskian thought allows for a looser definition of who can provide learning scaffolds. In games, for example, a player may be more adept at digital game play in general and may have more skill than his or her partner, therefore becoming a demonstrator (Wood et al., 1976) for the learner.

In a meta-analysis of groups working together with computer technology, several conditions were found to significantly enhance the condition of learning in dyads, including that students had group work experience, group size was small (e.g., dyads), and students were relatively low or relatively high ability compared to the general population of students (Lou, Abrami, & D'Apollonia, 2001). When students work in dyads using computer technology under these conditions they gain more individual knowledge when working together than when working with the technology alone (Lou et al., 2001). Additionally, it was found that while students working individually at the computer tend to complete their tasks faster, interact with the computer more, and get more help from the teacher, those in groups benefitted more from cognitive interaction with peers, used appropriate learning strategies, and had better task perseverance, and better individual performance. The meta-analysis reveals that when the positive conditions for group work are met then a "moderate positive effect of social context (mean ES = +0.66) may be expected" (Lou et al., p. 477). Of course, peer interaction must be supported either by the teacher or the digital cognitive tool, or both.

There are a variety of game-play arrangements supported by different kinds of digital games. In the typical game-play arrangement of a game like SURGE, each student has his or her own play experience and the game does not intrinsically require the cooperation of other simultaneous players nor does it specifically support peer collaboration (unlike a conceptually-embedded game like Quest Atlantis does). That said, bringing a game like SURGE into a school context puts the game in the social plane of the classroom. When watching others play and discussing play with a peer, players can see other approaches and strategies which may inform or even improve their own play. Likewise, while games can provide "just-in-time" scaffolding to individual players, that scaffold may not be appropriate for all learners--teachers or peers can respond more sensitively to students' play experiences and inform their own play and, as a result, their learning.

One particular study which highlighted the interaction of peers and game learning was done with the commercially-available Korean MMORPG *Gersang*, which simulates the Korean economy from the 19^{th} century (Kim, Park, & Baek, 2009). The researchers briefly trained 132 ninth graders in three metacognitive strategies to use explicitly while playing the game: self-recording, thinking aloud, and modeling. Self-recording was an individual strategy which involved personal, guided, purposeful reflection on play. Thinking aloud involved students in cooperative gaming groups discussing their plan for play during the day and explaining the play process to the group out loud while playing. Modeling involved identifying and then watching a fellow classmate for 10 minutes during a 45 minute play session, taking notes, and making observations to use in individual play. They used these strategies to facilitate the metacognitive process of self-planning, monitoring, and evaluation while playing the game. Outcomes were measured by performance on a pre-/posttest and final game score. The study found that the strategies of thinking aloud and modeling provided a statistically significant path to achievement in learning, while the more isolated activity of self-recording did not. The authors hypothesized that the social nature of the modeling and thinking aloud strategies accounted for these findings. The study was limited, however, in that the researchers relied on students' self-reporting of their use of the meta-cognitive strategies.

Before delving into the process of bringing a game like SURGE into the classroom, I provide a rationale for considering games as different from any other digital resource and also provide a rationale for using games for science learning. Then I describe the game SURGE specifically as well as its integrated scaffolding.

What is a Game and How is a Game Distinct from a Simulation?

Many scholars have suggested working definitions for games (e.g., Caillois, Salen, & Zimmerman, 2006; Dickey, 2005; Huizinga, Salen, & Zimmerman, 2006). In science education, precisely defining a digital learning game becomes especially tricky because of the prominence of computer-available simulations as digital resources (Songer, 2007). Cruickshank (1980) defines a *simulation game* as one "in which participants are provided with a simulated environment in which to play" (p. 22) whereas a straight simulation is a product which is a manipulable simulacrum of some other process, phenomenon, or object. In other words, a simulation is not a cognitive tool until it has been scaffolded for the classroom using a design framework such as the WISE meta-principles (Linn et al., 2003) or the Cognitive Tools Framework (Songer, 2007). Because of the prominence of simulations in science education and the relatively mature research base informing both simulations' design and delivery as cognitive tools, Clark et al. (2009) advocate defining "digital games specifically in terms of their relationship to simulations" (p. 25).

Simulations are similar to digital games in that games usually involve some kind of interactive model which the player can manipulate directly or indirectly by changing parameters, but they are different in the sense that games typically elicit some level of enjoyment and motivational engagement as a core design feature (Clark et al., 2009). The relative enjoyment of games and simulations is of course subjective, so more readily identifiable and applicable markers are required to truly distinguish the two.

The essential ingredient of "gameness" (Juul, 2003) is the "magic circle", a self-contained space the player enters where game rules reign supreme (Huizinga, 1938). It is here that players work together or independently toward achieving the goal of the game, iteratively refining their play in response to either game-provided scaffolding and feedback, or via feedback provided by fellow players or leaders/coaches/teachers. Synthesizing the work of Huizinga and other prominent thinkers around games (e.g., Caillois et al., 2006; O'Neil et al., 2005; Salen & Zimmerman, 2003), a more robust theoretical definition of games has been developed by Juul (2003). Rather than seeing a game as a stand-alone artifact, Juul defines a game along three dimensions: (1) the game as a formal system, (2) the player and the game, and (3) The game and the rest of the world. In other words, a game is simultaneously a mechanical and sociocultural entity and as such must be defined within a cultural context. In fact, while the gaming experience may be distinct, bounded by the "magic circle", it is certainly informed by sociocultural factors (Abrams, 2010) as explored in the section on bringing games into classrooms later in this chapter.

A game has six definable features which are classified into one of the three dimensions of a game (see Table 2.2): (1) Rules, (2) Variable, quantifiable outcomes (3) Value assigned to possible outcomes, some positive and some negative (4) Player effort, (5) Player attached to outcome, and (6) Negotiable consequences (Juul, p. 35).

Table 2.2

Essential game features categorized along three dimensions (Juul, 2003, p. 35).

	The game as formal system	The player and the game	The game and the rest of the world
1 Rules	x		
2 Variable and quantifiable outcome	x		
3 Valorization of outcomes		x	
4 Player effort	x	x	
5 Player attached to outcome		x	
6 Negotiable consequences			x

The first feature refers to the fact that all games must have rule sets which govern progression in the game. The second characteristic refers to the fact that each game play session will have different outcomes depending on play decisions--these outcomes might be represented by, for example, character death or opening up a new level; these outcomes need not be endgame outcomes but can be micro-outcomes throughout the game session. The third characteristic, value assigned to possible outcomes, refers to the feedback mechanisms which games employ to give the player an indication of how he or she is doing at a given moment, including health meters or running scores. These feedback mechanisms can also originate from fellow players giving one another feedback on performance. The fourth characteristic, player effort, refers to the fact that players feel challenged enough to exert some sort of effort into achieving the desired outcome(s). Players also must feel connected to the outcome, like feeling happy when they do well or unhappy when they do poorly. The final characteristic, negotiable consequences, refers to the ability of a game to have real-life consequences (or not). This final characteristic is what separates things like noble war and traffic from games: they have all the other characteristics of games, but they always have real-life consequences destroying the magic circle that separates a game from the real world.

Clark et al. (2009) also posit that games involve rules which must be followed in the interest of reaching a specified goal, usually including rewards or a scoring system along the way to provide real-time feedback of game performance. In the interest of building upon prior research in science education and games, this thesis defines digital games in the same way as Clark et al.: "Digital games involve: (a) digital models that allow users to make choices that affect the state of those models, (b) an overarching set of explicit goals with accompanying systems for measuring progress; and, (c) subjective opportunities for play and engagement" (2009, p. 26). Note that Clark's definition is fully consistent with Juul's. From these various definitions, it is clear that simulations might be on the borderline of games, but do not have the valorization of outcomes (i.e., a clear signal that the game is won or lost) or an overarching set of explicit goals which a game must.

Proposed Benefits and Drawbacks of Using Games for Education

Despite the wealth of terrible "edutainment" available (Kirriemuir & McFarlane, 2004), the proposed benefits of using games for learning are many as good games are designed with many supports for learning how to play the game autonomously (Gee, 2007; Sun et al., 2011). For example, it is suggested that games can enhance learning through visualization, experimentation, and creativity as visualizing is a central component of discovery and problem solving (Rieber, Luke, & Smith, 1998), especially in science (Habraken, 2004; Squire & Jenkins, 2003). Learning appears to be more effective when it is fun, and games combine elements of fantasy, curiosity, and challenge to increase the fun level (Malone, Lepper, Snow, & Farr, 1987). Cordova and Lepper (1996) found that students learning with instructional games in math classrooms outperformed students in more traditional settings and that context, challenge, control and curiosity increased motivation, though not necessarily learning. More holistically, Gee (2005) suggests good games are literally learning machines because they empower learners to customize their own experience, foster problem solving by presenting well-ordered problems with appropriate "just-in-time" (Saloman & Perkins, 1989) scaffolding, and promote understanding by requiring systems thinking, an essential 21st century skill (Marx, 2002). Others echo the belief that games can support the mastery of complex problems and concepts (Kelly, 2005; Klopfer & Yoon, 2005; Rieber et al., 1998; Swartout & van Lent, 2003). Sophisticated physics engines in digital games provide physically accurate representations of physical phenomena with few caveats (Price, 2008), and allow players to interact with normally invisible phenomena, like electric fields (Squire, Barnett, Grant, & Higginbotham, 2004). Good games include feedback mechanisms which communicate how well the player is doing at any time; further, good games allow the player to change the level of difficulty to match his or her level of skill to keep the game pleasantly frustrating (Gee, 2005; Kirriemuir & McFarlane, 2004), similar to keeping a learner in the ZPD. When a game is able to keep the child challenged enough to keep playing the game to get better, but not so challenging that the player gets frustrated and stops playing, it reaches the optimal state called 'flow' (Csikszentmihalyi, 1990).

Because of these proposed affordances, educators seek an understanding of the game playing process since games promote a level of attention in the player that educators wish they could harness to inform educational approaches (Kirriemuir & McFarlane, 2004). But playing a game at home or in a research lab is different from playing a game at school because of the altered cultural milieu (Abrams, 2010; Heeter et al., 2003): this study is particularly interested in how classroom supports and game supports interact with one another.

The current culture of schools is not a hospitable place for games, especially commercial ones (e.g., Hammer & Crosbie, 2006; Heeter et al., 2003). Using and learning new games in the classroom takes valuable time away from high-stakes assessment preparation (Hammer & Crosbie, 2006), so the educative features of a game must be obvious to the teacher without the irrelevant content of the game overpowering what is to be learned (Kirriemuir & McFarlane, 2004). Obviously, this need would also hold true for games designed specifically for education. These challenges must be considered because unresolved tensions can destroy research projects on games in schools (Hammer & Crosbie, 2006). In science education specifically, a concern in using technology to support learning is the background knowledge of the student and whether he or she has enough to understand the representations in the innovation (Edelson, Gordin, & Pea, 1999). Additionally, innovations must be robust and have enough fidelity to sustain open-ended inquiry, while also providing the motivation to engage in the inquiry (Edelson et al., 1999). Games like SURGE respond to these issues by providing carefully calibrated, sequential experiences designed to support the player's experience with each object/mechanic before adding a new one while providing the motivating context of a video game (D'Angelo, 2010). The game also addresses these challenges by building on students' intuitions of motion developed from other game experiences and experiences with SURGE (Clark et al., 2011).

Rationale for Exploring Games and Science Education

After Sputnik, science education moved away from positivist instructional models where knowledge and skills were transmitted to the students, into instructional models where thinking skills were taught alongside content so that students could actually use their scientific understandings (DeBoer, 1991; van de Akker, Fraser, & Tobin, 1998). Science knowledge is not viewed as an accumulation of discrete knowledge pieces, but rather a web of interconnecting understandings which are situated in the context in which they were originally learned. The more contexts in which a learner actively constructs his or her knowledge, the richer the learner's interconnections of knowledge become, and the more easily currently held conceptions can be brought to bear on unfamiliar situations (Bransford & Schwartz, 1999). Indeed, current national reform agendas (e.g., Duschl et al., 2007; NRC, 2011) emphasize sociocultural approaches to learning. These approaches recommend that teachers engage their students in developmentally appropriate inquiry by helping them actively construct ideas in personally meaningful ways through a variety of curricular patterns. But, Chinn and Mahotra's (2002) research found that many so-called "inquiry" activities in schools engendered scientific habits of mind antithetical to the epistemological underpinnings of inquiry. Since the proposed value of supports in games looks like a natural match with the supports for authentic scientific inquiry using technology like in WISE (Linn et al., 2003), while at the same time engaging 21st century literacies (NRC, 2011), there is a strong need for empirical research in games and science learning, particularly at the classroom level (Clark et al., 2009; Squire & Jan, 2007).

Most large-scale research efforts on games in education (Barab, Dodge, Jackson, & Arici, 2003; Barab & Dede, 2007) center around the idea that an "academic discipline is not primarily content, in the sense of facts and principles" (Gee, 2007, p. 22). In fact, a domain is actually "a lived in and historically changing set of distinctive social practices" (p. 22). Content is created in these practices and then written about, discussed, transformed, etc. Gee (2007) argues that outside of school, good games motivate the player to learn *how to play the game*, where experience leads to the abstraction of the procedural architecture and patterns driving the game play. Basketball serves as the perfect example. If I wanted to teach you basketball, I would not give you a textbook and ask you to understand the game; I would play the game with you. But Gee argues that this process is precisely what traditional science classrooms are doing with students—students come to the books to get the knowledge, even though that process has no connection whatsoever to how the content was created in the field in the first place.

Building from the situative tradition of learning as enculturation into a community of practice (Lave & Wenger, 1991), situativists see the process of getting good at a game as participating in a semiotic domain, where both discourse and actions take on distinctive meanings. This constitutes "active learning", a cyclic process of (1) probing the world, (2) forming hypotheses “about what something (a text, object, artifact, event, or action) might mean in a usefully situated way” (p. 88), (3) reprobing the world with that hypothesis in mind to see what happens, and finally (4) using feedback from the world to accept or rethink the original hypothesis. This process is the basis of reflective practice for any expert in a complex semiotic domain (Gee, 2007), and builds off of the notion that active learning across contexts leads to the abstraction of knowledge and transfer (Collins, Brown, & Holum, 1991).

Clark et al. (2009) delineate a categorical scheme for organizing learning game characteristics in science education along three dimensions (see Table 2.3, adapted from Clark et al., 2009, pp. 27-28). The authors are careful to note that the boundaries between categories within each dimension are porous; games can span multiple categories and are not always mutually exclusive. Science education learning games can respond to all four strands of the Taking Science to School report and exemplary games can be organized under each strand according to their focus: conceptual understanding (strand 1), process skills (strand 2), epistemological understanding (strand 3), and scientific attitudes and identity (strand 4) (Clark et al., 2009). The game used for this study, SURGE (described in more detail below), is more geared towards strands 1 and 2. Table 2.3 classifies SURGE and its use in this study along each of the three dimensions delineated by Clark et al. (2009).

Table 2.3

Organization of science game types. SURGE characteristics in each dimension are italicized.

Dimension	Genres/Categories
Nature of science learning connected to the game	Inquiry/argumentation as the primary goal *Simulation-based science content and processes learning in the game* Inquiry/argumentation/design/engineering learning among members of a community outside the game Familiarity with other disciple specific representations, tools and processes Science content knowledge
Duration and nature of the game participation	*Short interaction casual games* Longer duration finite games organized with specific start and stop time On-going participation type games in which players become members of a persistent community in and around the game
Intended purpose of the game along an entertainment /curricular spectrum	Fully recreational *Serious game for informal context that maintain design elements of recreational games but with more purposeful curricular focus* *Serious games designed for formal instructional contexts** Assessment games that are designed to assess existing knowledge rather than serve as a learning platform

*Note the game straddles two categories in the third dimension, intended purpose, because the game can be used both outside and within the classroom (Clark et al., 2009).

Specific Concept-based Games: The Evidence

In well-designed concept-based games, students learn the rules relatively quickly because they are rewarded for behavior that the designer desires, but unless the rules are sufficiently sophisticated, the learner will not learn the science idea at the appropriate level (Facer et al., 2004). Indeed, "The main challenge to designers is to develop sufficiently sophisticated game rules, and sufficiently focused challenges, in order to encourage the children to attempt different strategies to overcome these problems" (Facer et al., 2004, p. 407): note this is another angle from which to describe optimal flow with a game (Csikszentmihalyi, 1990).

ThinkerTools (White, 1993) is an early example of a computer-based simulation game used as part of a conceptual change approach (Strike, Posner, West, & Pines, 1985) to science instruction. The game-like simulation engages middle school students in authentic inquiry where the primary focus is to create and then apply causal models of force and motion as part of a larger instructional design where students cyclically (1) question, (2) predict, (3) experiment, (4) model, and (5) apply. Players are presented with two-dimensional modeling tools where they can create experiments which are impossible to run in the real world (simulation games allow one to turn off friction and gravity, for example). Using a 'dot-impulse' model, students can apply impulses to the object of interest in the x and y directions and actually interact with the extreme cases impossible to explore in real life. The researchers see the computer and the student as similar:

> The computer is not the real world; it can only simulate real-world behavior by stepping through time and using rules to determine how any forces that are acting (like friction or gravity) will change the dot's velocity on that time step. Thus the computer is actually using a conceptual model to predict behavior, just as the students will use the conceptual model they construct to predict behavior. (White, 1993, p. 15)

But eventually, students actually use their constructed causal models to complete game tasks such as arriving at a certain point on the screen with a certain velocity (White & Fredericksen, 1998). When they fail at the simulation games, they must revise their strategy (and perhaps causal model) to achieve the goal. Because the gaming aspects were embedded in a much larger instructional design, not much can be said about the specific affordances of the game aspect except to say that the intervention was successful.

At MIT, the *Supercharged!* project researchers (Squire et al., 2004) used an experimental design approach (Brown, 1992; Collins, 1990) embedded in existing classrooms to assess the potential benefits of a concept-based game designed specifically for classroom use. This game attempts to "couple the intrinsically rewarding aspects of games with the pedagogical power of simulations" (Squire et al., 2004, p. 3). Building off the research that computer simulations and games can engage students in learning about abstract, complex physical science concepts, Squire et al. designed *Supercharged!* to afford students the opportunity of exploring representations of electromagnetic fields in a game context. The game was designed specifically for a Physics First curriculum (AAPT, 2002), a relatively new approach which may require a conceptual instructional method that is alien to veteran physics teachers. Indeed, many science educators argue that conceptual or naturalistic physics instruction could lead to deeper understandings (diSessa, 2000), and Forbus explains that "students should deeply understand the qualitative principles that govern a domain – including the mechanisms, such as physical processes, and the causal relationships—before they are immersed in quantitative problems" (as cited in Squire et al., 2004, p. 2).

The game is a three-dimensional environment with a first-person perspective where students engage in two phases of play: planning and playing. In the planning phase, students are presented with a three-dimensional electric field arrangement and have to strategically place charges in the space to make their playing phase easier. In the playing phase, the player can switch his or her ship's charge (positive, negative, neutral, or dipole) and directly change the trajectory of the ship with a limited amount of fuel in order to make it to the goal. As the levels increase in difficulty, other objects are introduced like lines and planes of charge, and even electric currents or magnets. Like with many modeling and conceptual change approaches to education (Posner, Strike, Hewson, & Gertzog, 1982), the experience increases in complexity by drawing on prior knowledge while presenting something new to be assimilated via accommodation—this is cognitive conflict, the process of dissatisfaction with prior notions, and then finding newly presented notions to be intelligible, plausible, and fruitful for further pursuits. The game provides the motivation (Pintrich, Marx, & Boyle, 1993) to adapt cognitive structures in order to succeed at the game, as students always know how they are doing in relation to the goal.

The experimental classes that played *Supercharged!* outperformed the inquiry-based control classes. Through interviews, the researchers discovered that the game-based class also held more robust and accurate physical understandings of electric fields, and that “the primary affordances of games as instructional tools may be their power for eliciting students’ alternative misconceptions and then providing a context for thinking through problems” (p. 517). In other words, students used the game’s representations of electric fields as "tools for action" in reflecting upon their conceptions. It is crucial to note that *Supercharged!* was more effective with teacher supports. Students had a difficult time knowing what to do initially—teachers responded by providing students with log sheets to document their play in order to help them notice emerging patterns. The teacher also projected the game on the overhead so that students could, in a forum, interpret what they were seeing in the game. Adding this structure made students focus more on their play, ostensibly leading to deeper reflections (Squire et al., 2004), though no real data was collected investigating the effects of the teacher scaffolding on the play. Importantly, students did not learn vocabulary from the game which means it might be challenging to get students to express the implicit understandings and knowledge they might acquire from a concept-based game.

Like Squire (2006), I believe the most powerful aspect of *Supercharged!* is that the core game mechanics are inextricably bound up in science content. Malone and Lepper (1987) problematize the nature of contexts by labeling them as either endogenous or exogenous. Exogenous contexts are mere embellishments where the content of the game is clearly independent of the game context (like in *Hangman*), whereas endogenous contexts are inextricably bound up with the content, making it hard to separate the context of the game from the required learning which must occur to become good at the game (like in *Supercharged!*). Contextualizing factors in Project Based Science instruction (Rivet & Krajcik, 2008), an example of endogenous contextualization, have been empirically linked to improved student learning, as students who capitalized on more contextualizing factors of instruction (like the anchoring experience) did better on outcome measures. Concept-based games with endogenous contexts are much harder to design but have more proposed benefits in educational contexts (Fisch, 2004; Gee, 2005; Kirriemuir & McFarlane, 2004; Malone et al., 1987), and I believe this is possibly the most important distinguishing feature of a truly effective concept-based game. This sentiment is shared by Squire, Giovanetto, Devane, and Durga (2005) who found in a qualitative study on the commercially available game Civilization III that game learning environments are more effective when the mechanics of the game mimic the kinds of understandings we would like students to have. These endogenous games have been coined as "conceptually-integrated", whereas more open world games like Quest Atlantis, which are virtual worlds where students explore a game space which serves as a backdrop for more specific inquiry activities, are called "conceptually-embedded" games (Clark et al., 2011). SURGE's game mechanics are endogenous to the physics content to be learned and therefore maximize learning potential, making it well-suited for this study.

SURGE: Scaffolding Understanding by Redesigning Games for Education

SURGE is a conceptually-integrated educational physics game developed by a team of researchers at Arizona State University and Vanderbilt University and was designed with an eye towards current learning sciences research as well as the science education literature. SURGE is a simulation-based game emphasizing process and content learning in the game itself. The game has a finite number of levels and is organized to have a beginning and an end. Finally, it is a serious game designed for a formal instructional context.

In the game, players traverse two-dimensional levels[2] with their spaceship heroine named, appropriately, Surge. The goal of each level is to navigate through mazes with as few collisions as possible, while also collecting the “Fuzzies”, who have been enslaved by an evil warlord and trapped in the labyrinths. Using the rescue theme and art design reminiscent of the hugely popular *Sonic the Hedgehog* game series as motivators, players must complete embedded and endogenous physics challenges to advance through each level.

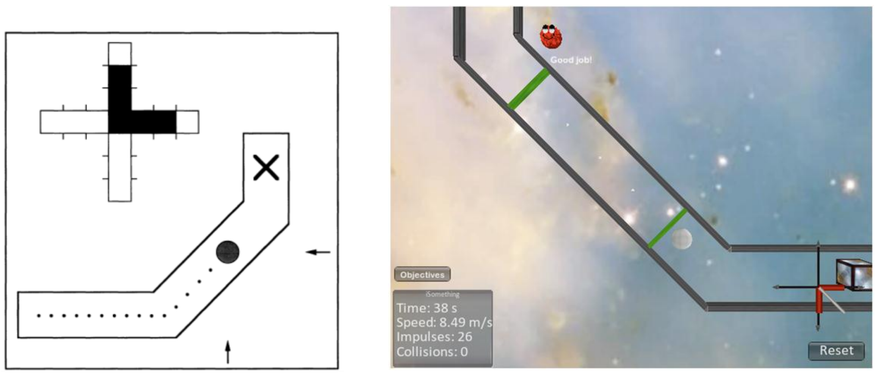
The game scoring mechanism incorporates data including the number of impulses players apply to their ship, the number of collisions with game obstacles,

[2] Note that the two-dimensional construct is a scaffold in and of itself, reducing the degrees of freedom.

how many Fuzzies were saved, the amount of time to complete the level, and level- or task-specific criteria to arrive at a score representing game performance. Each level focuses on one or two physical topics including impulse, inertia, vector addition, motion maps, velocity, and acceleration. Levels also include challenges which must be met to advance, such as navigating through certain parts of the maze while speeding up, or slowing down, or moving at constant velocity. If the condition is not met, way-gates will not open to allow advancement and players will have to try again until they meet the specific criteria. Corridors begin to narrow as well, making traversing these challenge regions more harrowing. The game is intended to help students build more robust intuitive understandings of vectors and Newton's laws, and potentially teach specific physics content.

Core SURGE Design Elements

The game design of SURGE is informed by a variety of learning theories as well as research projects in science education. Specifically, the game draws on the knowledge-in-pieces perspective on student learning (diSessa & Sawyer, 2002) as well as the coordination class theory of how these knowledge pieces form units, which help learners conceptualize different ideas (diSessa & Sherin, 1998). SURGE's core design elements stem from the ThinkerTools work done by White (1993). In ThinkerTools, students use a joystick to provide impulses to a dot on the computer screen. The program maps what White refers to as the "wake" of the dot, or the motion map of the previous positions the dot held at each point in time measured by the computer. The idea of the motion map is not only to indicate where the object was in the past, but also to give a sense of the speed of the object: when dots get further apart, the object is accelerating. In the same way, if the dots are always the same distance apart, the object is moving at a constant velocity. SURGE uses a data-cross similar to the ThinkerTools microworld; learners can see the velocity vector arrows split up into their x- and y-components as they navigate the mazes. These salient design features are included because they were productive for the novice students in the ThinkerTools work (D'Angelo, 2010).



Figures 2.1 & 2.2. Similarities of the data-cross and motion map representations in ThinkerTools (left) and SURGE (right). The data-cross can be seen both pictures, but SURGE adds a numerical display to track conditions and game performance.

White (1993, pp. 49-50) ascribed this learning in the ThinkerTools microworld to seven instructional strategies which:

(1) Employ manipulable, linked representations for key abstractions
(2) Make the phenomena easy to see and interpret
(3) Create scaffolded inquiry activities
(4) Reify the knowledge to be acquired
(5) Foster collaborative learning
(6) Facilitate model evolution by providing model progressions; and
(7) Incorporate learning about scientific inquiry

The SURGE team extracted three of these primary instructional strategies and made them endogenous to SURGE in its design: (1) manipulable, linked representations, (2) easily seen and interpretable phenomenon, and (3) model progression facilitation. Note that these are all examples of supports that cohere with both the WISE framework and the Wood et al. scaffolding framework (Wood, Bruner, & Ross, 1976).

The game has two full modules of six to seven levels each that are considered in this study. The first set of levels employs a constant force control system where holding down an arrow key will apply a constant force to Surge. In addition to the motion map regions where the player has to navigate a zone in a specific way, players can gain feedback from the game environment by noting their current speed, the number of impulses used, and how much time has elapsed playing the level. They can also observe their current x- and y- velocity component vectors along with the resultant vector superimposed on Surge. The second module of levels employs a control system based on discrete impulses such that each time an arrow is pressed, a discrete amount of speed is added to Surge (the character the player is controlling) in the selected direction. Levels gradually ramp up in complexity such that the first level just introduces the player to the controls. Next the player has to go around a few 90-degree turns and down long straight corridors, eventually having to go through the motion map regions at constant velocity, with acceleration, or with deceleration, depending on the level. Most challenging in the first module is when students must navigate through diagonal, 45 degree zones (two dimensional motion). Players must navigate these kinds of zones with both the impulse and the constant force control scheme. The score at the end of the level incorporates the data that students can track during play: number of collisions, number of impulses, and total time taken to complete the level (D’Angelo, 2010).

Other core design feature include a "Did you notice?" page at the end of each level, pointing out something interesting in the game as it relates to physics. Additionally, when the student is having trouble (i.e., running into walls repeatedly) the game stabilizes Surge as frustration control, and displays a piece information to scaffold the player's performance to mark critical features or provide direction maintenance.

In the end, SURGE's aim is deeply Vygotskian. As Clark et al. (2011) explain, "In *Thought and Language*, Vygotsky (1986) discusses the potential for leveraging intuitive understandings from every day experience ('spontaneous concepts') with instructed scientific concepts to build robust understandings” (p. 2180). SURGE aims to leverage the tacit understandings players gain from previous play experiences or experiences in SURGE and transform them into robust instructed concepts.

An international study comparing the implementation of SURGE in Taiwan and the United States found significant learning gains across the two countries on a pre/posttest after playing the game for 55 minutes. It is interesting to note that students did not have a background in vectors and were simply told by their teachers before playing the game, "SURGE relates to the concepts of force and motion, which you will learn later in the textbook" (Clark et al., 2011, p. 2184).

Bringing a Learning Game Into the Classroom

Bringing a digital resource like a game into the classroom requires either a broader instructional frame to be useful for learning, or the game itself must incorporate tenets of digital cognitive tools. The WISE framework (Linn et al., 2003) involves (1) making science accessible, (2) making thinking visible, (3) helping learners learn from one another, and (4) promoting autonomous learning. As described fully in previous sections, SURGE has been designed as a conceptually-integrated game and includes specific scaffolds to help the player make explicit connections from the game to science content. Players have a situated experience in the "magic circle" of the game but that experience becomes nested in the classroom plane. It is important to note that SURGE was not designed specifically to support peer collaboration and the game itself serves some duties normally reserved for the teacher through the design of the game. A useful way to visualize the different supports provided by the game and the social plane and how they interact with game and science learning is represented in Figure 2.3.

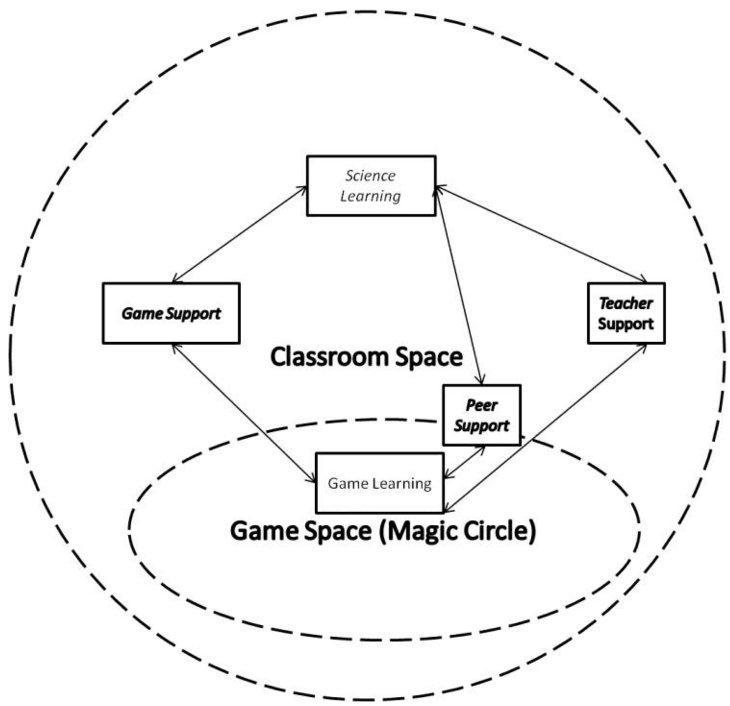


Figure 2.3. A conceptual model of the classroom space when bringing a game into the classroom. Note the external game supports and the teacher sit outside of the magic circle, but as the peer is playing the game with the player, the peer can function both within and without the circle. Porous boundaries of both the classroom and game space are consistent with socio-cultural theories of learning.

Table 2.4 shows examples of the kinds of supports the game, the peer, and the teacher provide using Wood et al.'s (1976) framework of scaffolding processes. However, those processes were empirically developed using a tutor/tutee arrangement and did not incorporate considerations of the classroom environment.

Table 2.4

Wood et al.' s scaffolding process (1976) and some examples of supports provided by the game and the classroom using SURGE. See table 4.1 for a more highly specified version of this table.

Scaffolding Process	**Mechanical (Game-based)**	**Social (Classroom based)**
Recruitment	Ideally, the game is attractive and fun but also accessible because it provides manipulable, linked representations of force and motion phenomena.	Teacher: Framing of the game play to garner buy-in from students. Student: Providing a sense of fun through friendly competition.
Reduction in degrees of freedom	The game is carefully sequenced such that each new concept is used sequentially, gradually building in complexity. (The game is also in two dimensions rather than three, though the effect of this support cannot be gauged through this research as it is pervasive.)	Teachers/Students: Both may set smaller goals within a game to facilitate completing levels, such as stopping at corners.
Direction maintenance	The game directs play through levels of increasing sophistication and scores/gold medals. This includes the velocity challenge regions.	Teacher/Peer: Teacher keeps players on task to complete the assignment and the levels. Peer keeps player on task to earn the highest score possible and earn all gold medals.
Marking critical features	SURGE highlights salient aspects of tasks by showing the vector cross directly on the player's spaceship Surge and the wake trail of Surge in motion map regions.	Teacher/Peer: Both teachers and peers may point out features that the player misses while playing to improve performance.
Frustration control	The game supports players who have trouble by steadying an out-of-control ship and explaining how to avoid going out of control through systematic application of impulses.	The peer may have suggestions or be willing to take over the controls to help with a particularly frustrating moment.
Demonstration	The game shows sample wake trails (in red) in the velocity challenge zones for players to replicate with their own white wake trail dots.	Teacher/Peer: The teacher or peer may model play to provide an additional strategy, or students in dyads may observe their partner use a strategy that they would like try or emulate.

Table 2.5 presents examples of the kinds of supports the game, the peer, and the teacher provide using the WISE framework meta-categories for successful cognitive tools. Since not all aspects of the WISE framework are covered by SURGE, classroom based supports can fill in the gap.

Table 2.5

The WISE design framework (Linn et al., 2003) and some examples of supports provided by the game and the classroom using SURGE.

WISE Framework Meta-Categories	**Mechanical (Game-based)**	**Social (Classroom based)**
Make science accessible	See recruitment, reduction of degrees of freedom, and frustration control in Table 3.	
Make thinking visible	Thinking is made visible in the game only in the sense that players can see the results of their strategies in real time as the game provides feedback through scores, the health of Surge, and the use of gold, silver, and bronze medals.	The players use log sheets and the external game scaffolding to represent the progress and thinking as they play the game
Support students learning from each other	This iteration of SURGE does not include built-in peer supports	Peer: Games involve an element of competition and goal-directedness, providing the impetus for collaborating to get the highest score.
Promote autonomous learning	Well-designed games are learning machines. Surge shows clear interactions of variables through an elegant interface and promotes experimentation.	Peer: the game may provide the motivation to encourage the pairs to work together to experiment until they get the highest score. Reflection opportunities have also been shown to foster autonomous learning.

SURGE presents a model of vector motion and gives constant immediate feedback on the state of Surge (the unit the player controls) and also scaffolds them through progressively more complex levels. The teacher's role is to help students learn from each other and from the game, but the role of the teacher may be quite different working with solo students versus pairs of students. Additionally, games can bring out feelings of competition and working with peer support may be dynamically different from peer support with simulations or other technological interventions. Simulations do not have win conditions and the valorization of outcomes that games do, and as such students come to a game with very different expectations. It could be that working with a partner in the classroom space changes the game and learning experience and use of supports in unexpected ways.

Need for Research

As many authors have indicated, while some believe that games are motivational, educationally effective tools, the empirical evidence to support such beliefs is quite limited and even contradictory, particularly evidence that games are effective for particular, concrete educational purposes (Facer, 2003; Kafai, 2001; Kirriemuir & McFarlane, 2004). In fact, most every study concentrates on motivational rather than curricular and instructional aspects or core academic benefits. There is a great need for studies focused on the actual players and their experiences and practices in and around the game, especially in classroom settings (e.g., National Research Council, 2011; Squire, 2006).

In a review of 99 studies on games and learning, Dempsey, Haynes, Lucassen, and Casey (2002) explain that in most studies learning outcomes are ignored. Even when learning outcomes are considered as part of the research analysis, the quality of the research methodology renders conclusions tenuous at best. A more comprehensive meta-analysis of 15 years worth of simulations and game articles found that under 2% of thousands of published articles met their standard of empirical evidence, and many games and learning studies focused not on specific learning objectives and concepts but rather on more holistic notions of fostering a sense of scientific inquiry and community (O'Neil et al., 2005). There is a need for research that measures if and how students navigate game and classroom supports to learn specific science ideas, while also avoiding, as much as possible, the methodological limitations of previous work.

Researchers have studied classroom supports around simulations extensively (Linn et al., 2003; Linn & Hsi, 2000), but not games. Though simulations share some characteristics of games, a game experience is uniquely bounded by the "magic circle" and is a distinctly different experience from working with a simulation. While research does exist of dyads working on computer-based tasks, the research around pairs of students working at the computer vary substantially given the nature of the computer-based activities. Very different learning outcomes arise through different kinds of interactions, depending on the actual digital interface (Dillenbourg et al., 1996). Lee et al. (2008) explain that working with a partner and technology-mediated prompts (scaffolding from the software) are helpful on their own, but we have relatively little understanding when comparing peer supports to technologically-mediated supports. We also do not know how they interact with one another, or with teacher supports. This study aims to understand how various levels of support interact with the experience of playing a game, one group with peer support and one group without peer support, to see how play experiences and learning change (or not).

A recent NRC (2011) report entitled *Learning Science Through Computer Games and Simulations* outlines a research agenda for games in science education and describes one critical research path as follows: "Investigate how best to integrate games into formal learning contexts (K-12 and higher education)...to enhance learning. This should include studies of how internal scaffolds in the simulation or game and external scaffolds provided by a teacher, mentor, peers, or other instructional resources (either in person or via various online mechanisms) support science learning" (p. 124). The present study begins to address this gap in the literature by analyzing the various interactions of different supports around and inside a conceptually-integrated game in the 8th grade science classroom, exploring students' uses of different supports during play sessions to answer the following research questions:

1. *What is the nature of the supports for science content learning provided by the game, the peer, and the teacher, when the game is used in a classroom setting?*
2. *How do the learning gains in the peer support condition compare to the solo play condition, both qualitatively and quantitatively?*

In the next chapter I describe the mixed methods employed to answer the research questions, including both qualitative measures like stimulated recall interviews and video observations, as well as quantitative measures like a pre-posttest assessment and game scores, and preview the data preparation and analysis processes.

CHAPTER 3

METHODOLOGY

A mixed methods case study design was chosen for this research, including both qualitative and quantitative sources of data. A description of the participants, setting, data sources, and analysis methods are included in this chapter.

Participants and Classroom Setting

The current study considers 8th grade co-educational students from a 7-12 independent day school in the Northeastern region of the Unites States. Students were taking a course called Science 8, a general science course in which all 8th graders were required to enroll, covering content in all major areas of science. Students were not "tracked" in science in middle school, so each individual classroom represented a sample of the general population at the school. There were four teachers of Science 8 with classes of 10-14 students, totaling N=80 students. One student did not participate in the study but participated in playing the game. Another student missed the entire enactment because of a family illness, and two other students missed both days where play occurred, leaving a total of N=76 students. Students at the school were admitted through an application process and most students paid tuition to attend. Science teachers tasked students with weekly group labs and all students worked in dyads or groups at least once a week throughout the year.

Because this study was conducted at a school, random assignment to groups was impossible marking this as a quasi-experimental design. Four classes were assigned to work in peer groups with the game during class, and two classes worked solo with the game. It was decided to use only two solo groups in the interest of getting a breadth of variation in the dyads, as peers work differently together--the only two classes with an odd number of students on the original roster were chosen as the solo groups, splitting the sample for convenience. The number of students in a class section of the same course was completely independent of the ability of a class. See Table 3.1 for a breakdown of the number of students in each condition.

Table 3.1

Group assignments by teacher.

	Mr. C	**Ms. M**	**Mr. A**	**Mr. R**	**Total**
Solo Players		10		12	**22**
Dyad Players	12		14 | 14	14	**54**

Note. Total N=76.

Data collection occurred over one and a half days of class time, as well as time before and after that period for pre- and post-assessments (20-30 minutes each) and individual audio-taped student interviews of target students (30-40 minutes each). The researcher acted as the teacher in the interest of providing as similar an experience as possible to each class of students. After introducing the students to the study (Appendix A), they were administered a background survey (Appendix B) and a paper and pencil posttest on velocity vectors (see Appendix E). On the next two class days (which may or may not have been consecutive, depending on the teacher), all students played SURGE for one and a half 55-minute instructional periods after being very briefly introduced to the concept of vectors through a two minute scripted mini-lesson (see Appendix A). Students were asked to achieve gold medals on each level before progressing to the next one, though on the second day this requirement was loosened to allow students to at least experience every level. Target students (see next section) were observed by video cameras trained on them and the computer screen. Students in both conditions were provided with the SURGE scaffolding problems to serve as guideposts to their play (for examples see Appendix C) and were asked to complete the reflection homework after the first day (see Appendix F). The students in the individual play condition played alone and those in dyads played together for the same amount of time. During the second half of the second day, students took the posttest, which included the same questions from the posttest as well as extra questions probing the students' uses of the game in answering the questions.

Table 3.2

The sequence of events in the study for each class. This process ran on a cycle for a period of three weeks to allow all six classes to go through the process.

Time	**What Occurred**
Day 1 Several days before the study	• Students introduced to the study, given signature forms. • Teachers collect forms as they come in.
Day 2 The class day before play commences (not necessarily the day before).	• Last call for permission forms. • Students take intro survey and pre-assessment in the latter half of their science class.
Day 3	• Students are given a short introduction to vectors and then play the game for the full period. • Students fill out the reflection for homework.
Day 4 The next calendar day after Day 3.	• Students hand in reflection, play the game for 30 minutes, and then take the post-assessment. • Some target students are interviewed later in the day.
Day 5 The next calendar day	• The remaining target students are interviewed.

Target Students

Selection

Students were asked as part of their background survey if they would be willing to participate in an audio taped interview after the second day of play. From the pool of students volunteering, the teachers of each class eliminated any students who had attendance issues or that were having academic difficulty (e.g., performing at a C level or below in science class). From that remaining pool, equal numbers of boys and girls were randomly selected as target students (10 boys and 10 girls). Each 8th grader had at least one hour of study hall a day, and all student interviews took place during study hall times in the science building. No interview occurred more than 24 hours after the end of a student's last play session.

Two target students were selected from each of the two solo classes for a total of four solo student interviews. Two target pairs were selected from each of the four dyad classes for a total of 16 interviews. Note, however, that students in dyads were interviewed separately to provide candid descriptions of their experiences. These students were interviewed outside of class time to provide a richer description of the game play experience and their thinking around the game while playing, vis-à-vis the three levels of support: teacher, peer, and game (see Appendix B). One target student became ill during the school day and had to go home before she could be interviewed, though she did participate in all other aspects of the study; because this student's partner was also a target student, it was decided that the dyad's study-related materials would still be used for analysis, but that they would be treated like non-target students, leaving 18 target students total. A brief description of each target student or target student dyad follows.

Solo Students

Elliot.

Elliot was a very quiet student while playing the game with intense focus, never once asking for teacher support. He reported playing games three to six hours a week on average and finished the game before other students. As a result, he became very interested in how to maximize his score by running experiments comparing time of completion to number of impulses/amount of fuel used to complete the level. He fully mastered control of the spaceship and became adept at only accelerating when absolutely necessary.

Greg.

Greg reported that he was a gamer, playing video games more than 20 hours a week during the school year and more than 80 hours per week in the summer. Greg's gaming habits tended towards the creative and role-playing genres: for example, he was actively constructing a replica of the entire school and its grounds in the game *Minecraft*. Greg also finished early and spent his extra time obsessively playing impulse level six over and over again to have a "perfect run". His play was confident and fast.

Maddie.

Maddie was a self-described casual gamer, only playing short games designed for portable devices like the iPod, or her cell phone. Her favorite game was *Angry Birds*, a wildly popular cell phone game which uses trajectory physics as its primary game design element. She talked to herself and to those around her regularly while playing, laughing at the "Fuzzy" characters in the game. She played very cautiously and slowly and after some teacher support, she collided very rarely. She did not finish early.

Wallace.

Wallace was an extremely thoughtful student who played seven hours of games per week, all sports games like *Madden NFL 2011*. He had a lot of trouble adjusting to the controls of each respective set of levels, often cursing at his own play under his breath and calling his performance a "disaster" on the first levels of impulse control. He too found success in slower and cautious play after teacher supports, but then became more confident and started playing with speed by the end of the play session. Wallace used the full time to play through the levels.

Dyad Students

Allen and Hank.

Allen reported playing 18 hours of games per week, mostly physics-based action or puzzle games like *Portal 2*. Hank reported playing strategy games, like *Age of Mythology*, less than one hour per week during the school year but much more in the summer. They both characterized their partnership as cooperative and they watched each other's play with careful attention to the game mechanics and user interface. Though Allen displayed more dexterity with the controls, neither had too much trouble completing the game with minimal teacher support. They used their remaining time to determine how to get the best possible score on constant force levels.

Annie and Lara.

Annie and Lara were an awkward pair who were not friends outside of class and had never before worked together on a class assignment. They played amicably but did not interact very much when watching the other play, though they did find watching each other helpful to the understanding of the game. Annie reported playing one hour of games per week, typically platforming-type games like *Kirby Returns to Dreamland* or *Super Mario Brothers*. Lara said she played games one to two hours a week after homework, usually casual cell phone games like *Temple Run* (not physics-based) and *Words with Friends* (a casual word game not unlike the board game Scrabble). They played through the full game and did not finish early.

Bentley and Lidia.

Bentley and Lidia had a very friendly and mostly cooperative partnership, though a couple of one-point difference match ups on levels brought out a congenial competitive streak. Bentley said he plays around three to six hours of games a week like the adventure/shooter *Half-Life 2*. Lidia said she plays absolutely no games during the school year, but that her favorite game to play is *Scribblenauts*, a creative thinking game where you can write pretty much any noun and the object will appear on the screen to help you complete the given objective. They played through the whole game and did not finish early.

Carter and Siobhan.

This partnership marked the widest gap in videogame play experience between the players. Carter played first person shooters like *Halo: Reach* for about five hours a week, enjoying the cooperative online mode with his friends to destroy enemy targets. Siobhan was a very inexperienced gamer relative to the population, having only played the game *Temple Run* about ten minutes a week for a couple of weeks (i.e., she played her first video game two weeks before this study). Carter grew audibly and visibly frustrated with Siobhan's slow and inexpert play, and did not have much patience for guiding her to more fruitful approaches. Siobhan changed strategies and improved play, but was unable to get a gold medal on every level before the end of class. Carter also did not get a gold medal on impulse level six (the hardest level) and surprisingly did not change strategies very much and generally used a guess-and-check approach through the whole game.

Damien and Nathan.

Damien and Nathan were both regular gamers (3-6 hours a week each) and extremely competitive players during class time. They jocularly taunted each other and strived to beat each other's scores, though speed was more important to Damien than Nathan. In fact, Damien was a driving/racing game fanatic, placing *Need for Speed: The Run* and *Midnight Club 4* among his favorite games. Damien used techniques he dubbed "hugging the wall" and "cutting corners" to achieve the most efficient times possible. Nathan played slowly and carefully at first, but started trying to go faster in order to keep up with Damien's higher scores; Nathan was a fighting game aficionado. Both students completed the game with gold medals, but spent much time replaying levels trying to one-up each other.

Danya and Ione.

Danya and Ione were both gregarious and extremely cooperative while playing. One typically took on the role of cheerleader while the other played, reading on-screen prompts and warning the other of looming dangers. Ione asked me numerous game and physics-related questions while Danya was playing and would then explain what I had said to Danya after she finished the levels. Ione said she played about one hour per week, and her favorite game by far was the physics-based game *Angry Birds*. Danya was a self-described casual gamer with about one hour of play a week; she loved *Temple Run* and the physics-based game *Osmos*. They completed the full game but were unable to earn gold on impulse level six.

Enzo and Jeremy.

Enzo and Jeremy were both regular gamers. Enzo played 7 to 14 hours a week but claimed no favorite genre or game because he plays games until he gets bored and then moves on to the next one. Jeremy was a fan of the real time strategy genre, playing *League of Legends* (not physics-based) for six to seven hours a week. Their partnership was cooperative and they supported each other in maximizing scores. However, their play was mostly restricted to game-elements and they did not talk about physics much while playing the game. They completed the full game.

These target students were the focus of the digital videotaped play observations and participated in individual, 30-minute stimulated recall interviews, introduced in the next section.

Data Sources and Collection

In many case study analyses, one or two methods of data collection dominate while the others "play a supporting role in gaining an in-depth understanding of the case" (Merriam, 1998, p. 137). Videos of play experiences and stimulated recall interviews are the main data sources to answer the first research question under study, but other sources inform these main data. The second question uses data from pre- and posttests and other survey data along with the quantified qualitative data from research question one. A description of each of the data sources is presented, roughly in the order used in the study. Fully detailed descriptions of data collection and analysis methods are in each of the results sections (Chapters 4-5).

Introductory Game and Science Survey

All students took a short survey on their digital game history as well as their previous science courses (see Appendix C). This data source is primarily descriptive to give a better idea of the population and items like previous play experiences and gender are considered when triangulating data sources and when checking for possible covariances in the quantitative data set.

The Pre-Posttest

In this study, student learning is partially defined as measuring gains on the conceptual physics questions from the pre- to the posttest (see Appendix G), though qualitative measures are also considered and triangulated with the more traditional quantitative measures as this is a mixed-methods approach. The pre-posttest was developed as part of the pilot study for SURGE and was subsequently modified based on pilot data for a dissertation investigating undergraduate students' uses of two different versions of vector representations in the game, as well as the differences in the types of scaffolding provided to students directly following play (D'Angelo, 2010). Questions from the pre-posttest include items from the seminal Force Concept Inventory [FCI] (Hestenes, Wells, & Swackhamer, 1992) and other items from well-regarded published studies on student understanding of vectors (Flores, Kanim, & Kautz, 2004). It should be noted at the outset that students typically show minimal improvement on the FCI after a full semester of physics. The reliability of the test using Cronbach's alpha is .74 (D'Angelo, 2010), an acceptable value for this study.

The posttest has additional items which ask students to provide information about if and how the game helped them answering the test questions. Because students tend to write much less than they may actually feel, these extra questions were further probed in the stimulated recall interviews with selected target students.

Reflection Pieces: *Game Score/Data and Homework*

Software created by Clark and his colleagues specifically for SURGE keeps an in-game data log, but because of a server issue at the school site the laptops were unable to be connected to the internet while students played and students had to use a local version of the game. Having prepared for such an issue, all students were provided with tracking sheets (see Appendix D) to note their scores on their trials for each level and other salient information (e.g., number of collisions). Students tracked scores for both days of play, and all aspects of the score were considered--overall score, color of medal earned for each level (bronze, silver, or gold), number of collisions, and time. To assure that the scores were accurate, student data sheets were spot checked with videotape for a subset of the data corpus. These checks revealed that solo or pair monitoring of score was not a reliable measure as students tried many more trials than they actually recorded on their data sheets, favoring successful trials. For this reason, data logs primarily drew the attention of the students away from the physics content presented at the end of each level: they paid much more attention to the scores. These and other issues with the data logs are presented in more detail in Chapter 5.

As homework after the first day of play, students were asked to articulate their version of the rules of the game and also any strategies they used or developed while playing the game, as well as rate their enjoyment and level of challenge on two Likert scale items (see Appendix F). Reflection is an important part of the instructional design as reviewed in Chapter 2, and the intention was for students to think about the game on the night between their two play sessions. The Likert data was used to further describe the population's affective responses to the games as a whole and by group. Additionally, prior empirical research suggests that forced play on students will only have the desired effect if students like the game enough to choose to play it outside of class (Heeter, Lee, Magerko, & Medler, 2011). For this reason it is important to gauge student's affective response to the game to consider the enjoyment score when interpreting the data.

Video

Filming students playing has been a common methodological choice for recent studies of student play (e.g., Hung, 2008; Lee & Probert, 2010). Two small Flip Video HD digital cameras with attached wide-angle lenses and miniature tripods were positioned on the lab table just to the side of the two target students (or the two student dyads), to capture gestures and other human interactions such as the teacher coming by to point something out or one student taking over the mouse from another student. The camera also captured the computer screen. These videos were analyzed in NVivo 9, as described in the analysis section.

A third video camera captured the entire classroom space to allow for a coarse, but full, view of all interactions, the arrangement of desks and lab tables, and a general view of the flow of the classroom experience. This decision allowed for reporting of actual class time spent playing the game, any moments of technical difficulties, all teacher/student interactions, and any instances where one group may go over to another, or one student may go to another, to get advice on advancement in the game. Importantly, as the researcher acted as the teacher in the classes, field notes from the class experience were impossible to capture adequately without the use of video to supplement memory.

Target Student Stimulated Recall Interviews

After taking the posttest, target students were asked to engage in one 30-40 minute interview with the researcher (dyad target student interviews were generally longer as the cuing clips were longer). These stimulated recall interviews (Gass & Mackey, 2000) included the student watching researcher-selected video clips from the student's play session alongside the researcher while being encouraged to describe exactly how they felt or what they were thinking while playing (see Appendix B). The focus was on recalling their previous play experience in class to ascertain the kinds of intuitions and prior experiences they brought to the game experience, as well as their understanding of game mechanics as related to content learning. The questions concentrated primarily on the students' thinking around supports provided by the game, the teacher, and the peer (if applicable). The audio for these interviews was recorded by a Flip camera, which was simultaneously filming the laptop screen on which the student and the researcher were watching the classroom play video.

Data Management

Data was managed in QSR NVivo 9 in a three step process recommended by Reid (1992): data preparation, data identification, and data manipulation.

Data Preparation

The preparation process involved putting the data into a usable form. This included uploading audio and video information into the computer as well as writing the descriptions, observations, and interactions from the videos of classroom play right inside NVivo 9. In other words, rather than creating whole summaries of each class, video observations were written alongside the video clips such that each observation was mapped to the appropriate point in the video. A similar process was followed for the stimulated recall interviews.

Descriptions of the play experiences as well as interactions were noted alongside each level segment for the observation videos. Strategy switching and number of trials attempted for each level were noted, consistent with another study on using a game for kinematics (Holbert & Wilensky, 2011). Other events of note were also described in the observation videos and interviews in a way consistent with Merriam (1998) such as who asked questions, what the questions were about, what students pointed to on the screen and how often, what game characteristics they noticed and talked about, what gave them trouble or what they found especially easy, and interactions with the teacher/researcher. In other words, the descriptions concentrated on interactions--player/game, player/teacher, player/peer, and game/teacher. Careful attention was paid to the differences between solo play and peer play in the observations.

Data Segmenting

A convenient way to segment the observation clips proved to be by game level (e.g., constant force level one play, per student, would be one segment), while stimulated recall interviews were segmented such that each time the play video was stopped, that began a segment, and each time the video was restarted, that ended the segment. Surveys and pretest/posttest solutions as well as homework paragraphs for all students were also imported into NVivo 9 after converting them to a form interpretable by the program (either as typed documents or spreadsheets of survey/assessment data).

Data Coding

After preparing and segmenting the data, the final step involved coding meaningful chunks and segments which subsequently could be indexed and searched for, queried, sorted, or rearranged in the final step of data manipulation, which led directly to the analysis.

Analysis

A description of how each research question is addressed using the data is presented here. Findings in the study are presented as episodes as well as in the form of assertions, about the nature of supports and the ways students' experiences are shaped by them, which emerge from a recursive analysis of the data corpus, similar to a process used by Lee and Probert (2010). See table 3.3 for a visual representation of the research questions and the associated data sources used to answer those questions.

Table 3.3

Data sources matrix. Primary sources are in italics.

Research Questions	Data Source(s)	
	Quantitative	**Qualitative**
What is the nature of the supports provided by the game, the peer, and the teacher, when the game is used in a classroom setting?		*Videotape recordings, interviews,* descriptions, background survey
How do the learning gains in the collaborative condition compare to the solo play condition, both qualitatively and quantitatively?	Pre-posttest gains	Effective support scores, reflections, survey

Research question one: What is the nature of the supports provided by the game, the peer, and the teacher, when the game is used in a classroom setting?

To address this question, the videos of target student classroom play sessions and the stimulated recall interviews served as primary sources. The videos were imported into QSR NVivo 9 and all uses of supports from the game, teacher, and students which were visible from the video were coded using the scaffolding theory of Wood et al. (1976). Supports were first identified as either coming from the peer (P), the teacher (T), or the game (G). Then, each scaffolding moment or exchange was coded as one of the six scaffolding processes: recruitment (1), reduction in degrees of freedom (2), direction maintenance (3), marking critical features (4), frustration control (5), and demonstration (6). Not all supports had the intended effect or were even noticed. For example, some players skipped over the story elements between levels to get back to playing quickly, but physics content was described in the context of the game in these introductory screens. A peer may have demonstrated a perfectly suitable strategy that the partner immediately failed to even attempt. A teacher may have pointed out a critical feature in the game environment but the student may have misinterpreted it. Moments like these were coded in the same way as other scaffolds, but with an added (NE = not effective).

Student interpretations of supports were not always visible or were unclear in the videos, which is why stimulated recall interviews suited this study and were an effective primary source to help triangulate and validate the coding of the observation videos. The stimulated recall interviews were fully transcribed and coded in a similar way, also in NVivo alongside the interview video, to corroborate the observation video codes or disconfirm them, as well as add more instances of support to the corpus of data. The cumulative matrices of codes for each interviewee were compared, by group, to make assertions about how different levels of support interacted with one another. Analytical memoing was used to uncover two underlying patterns in the data, which were explored at length.

Research question two: How do the learning gains in the collaborative condition compare to the solo play condition, both qualitatively and quantitatively?

I quantitatively expressed the relationship between the pair condition and solo science learning as measured by pre/posttest assessment gains. A repeated measures ANOVA (time x group) determined whether or not groups improved their scores (both pre/post score and game score) significantly more than the other group. Item analyses using generalized linear models (Binomial Logistics) were also performed to determine which assessment items had significant changes, by group. The findings from research question one were also revisited in the context of the quantitative findings to make assertions about content learning vis-à-vis support usage. This triangulation with qualitative findings validated the quantitative findings, while the qualitative findings gave voice to the quantitative findings (Creswell, 2007).

Target students only represented a quarter of the whole sample, so the reflection data source offered a more complete picture of the entire population and further informed the quantitative findings. The intention was that these reflections on the rules of play and strategy, given as homework, were to be coded looking specifically at whether students used science language or game language to describe play. Most students wrote short responses like "go slow" for strategy or "Save the Fuzzies" for the rules. As the responses were not illustrative of physics concepts used, the reflections were used to support descriptive findings.

Findings for the two research questions are organized into the following two Results Chapters. Student interactions involving game, teacher, or peer supports for play or learning, which are illustrative of claims or assertions in the findings, are transcribed verbatim to paint a portrait of the enactment across the different conditions. Patterns, assertions, and claims are refined into a number of recommendations for support for game designers, teachers, and educational researchers to provide to students who play concept-integrated games in classrooms. More thorough descriptions of data management and analysis specific to each research question are included in the findings (Chapters 4-5).

CHAPTER 4

THE NATURE OF SUPPORTS IN AND AROUND A GAME USED FOR CLASSROOM LEARNING

This chapter involves an exploration of the first research question: what is the nature of the supports provided by the game, the peer, and the teacher, when the game is used in a classroom setting? Video data were prepared, segmented, and coded, looking especially closely at target students' effective uses of game, peer, and teacher supports through interactions with each. Relative frequencies of use of each type of support were tabulated in order to compare and contrast the kinds of support used by solo and dyad students.

The actual result of these analyses is presented in several parts. Data collection and analysis methods are covered in greater detail, including data preparation, data identification, and data manipulation. Next, I give an overview of the enactment in both the solo and dyad classes. I present the results of the analysis looking at the solo target students first, who had most of their support coming from the game itself; I then turn to the dyad target students and compare their experiences and uses of support to the solo students. Third, I provide a numerical representation of the effective support from the game, the peer, the teacher, and overall for each target student, and I describe characteristics of students at the high and low end of effective support. Finally, from the analysis of the data, patterns regarding the nature of supports emerged across both conditions through analytical memoing (Merriam, 1998). These patterns are described and illustrated through examples directly from the data corpus.

Data Collection and Analysis Methods

Data collection methods are described more fully in the following section. All data entry and manipulation was handled using QSR Nivo 9 as powerful new features facilitate working with video sources in more efficient ways. Methods for handling each of the primary sources are fully described as well as the rationale behind aligning the observation videos and the stimulated recall videos. A fuller description of coding decisions for both primary sources is provided.

Data Preparation: Transcribing, Segmenting, and Alignment

The two primary sources, video observations and stimulated recall interviews, were handled in similar ways, though important distinctions are made here.

Observation videos

Three Flip video cameras were placed in each classroom. One camera recorded the entire classroom space: this camera began by filming the class from behind the last row of desks to ensure that the vector introduction given to each class was identical. As students moved to the lab tables in the back of the classroom to begin play, this camera was moved by the regular classroom teacher to capture the lab table area. It was ensured that the student who did not participate in the study was off-camera.

The full classroom observations were more generally described to give a sense of the classroom atmosphere and an indication of the flow of the class. This gave a proper sense of what each actor was doing in the space, though lacking the specificity of the target videos because of the usually indistinguishable audio. These videos were not coded but rather used to describe class episodes and the overall enactment of the study.

The other two cameras were trained on target students or target student dyads, using a wide-angle lens to capture faces, hand gestures, and of course the on-screen action. This angle did not always capture a teacher interaction, but voices were loud enough to transcribe those interactions whether they appeared on screen or not. Before uploading the observations into the software, each pair of target student(s) videos were merged into one larger file in the video editing software Windows Live Movie Maker. Each of these larger 90 minute videos were imported into NVivo 9 for description, transcription, and analysis. After uploading the target student observations into the software, a meaningful way of chunking the videos became clear: each level of play would constitute a single chunk. Transcription involved not only writing what was said by students and teachers, but also describing the actual approaches to play and the special events in play, like starting levels over or switching strategies. All events taking place in an observation were placed in [brackets].

Stimulated recall interviews

It was clear from reviewing observation videos on the night after the first day of play that students were most interactive with one another and the teachers during the first couple of levels of play and when the game's control scheme switched from constant force (the first six levels) to impulse (the last seven levels). It was also clear from the classroom that impulse level six got the largest rise out of the players. For these reasons, target students were shown common clips from their play sessions, adding a level of consistency to the interviews: the first two levels of each control scheme (constant force levels 1-2, and impulse levels 1-2), and the sixth level of impulse. In addition to these five levels, three additional clips were chosen for each target student when an interesting support event happened outside of those 5 levels. Most of these extra clips were chosen from day one as some stimulated recall interviews occurred just minutes after the second day of play (though there were exceptions when I made a personal note to include a clip of something I observed in the second day, noting the time so that I could efficiently find the clip to show the student).

Chunking was more obvious with the stimulated recall interviews. Any time the researcher or the student stopped the video, that began a chunk. The end of the chunk was demarked by beginning the video again. In between these moments of time was usually an interviewer question and then a student response. There was no fishing for answers from students, though clarification follow-up question were asked when appropriate (i.e., "You said 'inertia', what did you mean by that?") and students sometimes talked about several supports during a single chunk. Additionally, clips were generally longer in the dyad interviews as we often watched both the interviewee and his or her partner play each level, especially on the first level of each control scheme. For this reason, dyad interviews often took a bit longer (up to 40 minutes total).

It is important to note that the video was paused many times during some clips. One of the interview questions involves asking the student what they were thinking when certain game features appeared, like the opening screen, the wake trail, the vector arrows, etc. For this reason, some clips (especially Constant Force Level 2) were paused multiple times.

The stimulated recall interviews concentrated both on use of supports in the classroom and game as well as the student's game experience. Because stimulated recall interviews used actual video clips from the observation videos as the stimulus for response, descriptions and dialogue from the clips which were shown to the target students to cue their thinking were reassigned in the observation video coding with {curly brackets} to distinguish those observations from ones that were not shown to the student to cue a response, which were in regular [brackets]. All student and researcher dialogue in the videos was transcribed verbatim and linked to the cuing clip from the observation videos. See Appendix H for a sample data chunk.

Alignment

Though not done literally, NVivo was used to create one master description/interview document to code for each student, including all descriptions and interactions during classroom play in [brackets], actual observation descriptions of cuing events in {curly brackets}, and finally the interview material placed in line with the descriptions in normal type. This facilitated coding across the two primary sources, as described in the next section.

Data Identification

Data chunks varied in size depending on the number of relevant events that happened in each chunk. Some chunks included a cue, a researcher question, and a simple "No" or "I wasn't thinking about anything" as a response, constituting a short chunk. However, the chunk presented in Appendix H is relatively large and includes multiple references to supports. For this reason, chunks were further segmented to facilitate common coding across the two researchers doing comparative coding. Appendix I shows how the exemplar chunk in Appendix H was segmented to prepare for data coding. Each segment represented one main idea, game element, and/or support moment. These segments were coded according to the rules set forth in the next section.

Data Coding

At the surface level of coding, a node was created for each target student as well as node for "student response". This made it possible to highlight all non-interviewer dialogue in the interview and code it as a student response, therefore facilitating searches of only student responses when looking for word patterns across the target students, for example when looking for mentions of a specific game design element like the "yellow arrow" or the "white dots". Additionally, entire student interviews and observations were coded at the dyad or solo node, depending on condition. This decision facilitated queries focused on differences between the two groups.

At the finer-grained level of coding, each segment was labeled according to the framework set forth by Wood, Bruner, and Ross (1976). In the initial segmenting phase it became clear that Degrees of Freedom (DF) and Direction Maintenance (DM) would be challenging to differentiate. For example, when Maddie, a solo target student, mentioned that the impulse level controls were really different and that, "I realized you really have to slow down or stop on a corner cause the walls are more narrow than in constant force," it was hard to tell whether the game's reduction in the possible degrees of motion or the fact that the game forced more directed paths as direction maintenance led her to change her approach to play. When trying to achieve inter-rater reliability, the potential problem was confirmed as my fellow coder and I would often debate whether to code one event as Degrees of Freedom or as Direction Maintenance when I was describing the planned coding algorithms with small data sample. Though there were times when the difference was distinct, in too many instances the decision was muddled; it was decided that collapsing the codes was a more prudent path. However, it should be stated that other games might have more clear distinctions between these two types of supports and that keeping them separated may be useful. This new code was called Refocus (Ref), to envelope the two meanings: degrees of freedom refers to methods like breaking down problems into sub-parts, to have you focus on what is important, and direction maintenance refers to keeping the learners eye on the task at hand, helping them see the next salient challenge or goal. Table 4.1 explicates exactly how the kinds of supports from the game and classroom were coded in the way they were. Recall that a support is defined as something that helps one learn or helps improve the conditions for learning.

Table 4.1

Evidence of game supports and peer/teacher supports for each scaffold code.

Code	*Game*	*Peer/Teacher*
Recruitment	- Having fun - Motivated by the story, high scores, or better medals	- Motivated by cooperation or competition with partner - Motivated by teacher
Refocus (combination of Degrees of Freedom and Direction Maintenance)	- Describes changing strategy because of a game design element - Describes differences in levels' designs and the impact on play - Describes changes in the controls	- Peer or teacher encourages breaking a problem down into steps, for example restricting motion to one axis. - Peer or teacher brings player back on track or keeps them on track with real- time advice
Marking Critical Features	- Describes using the wake trail - Describes using the vector arrows or speed display - Uses the map to plan approach - Just-in-time scaffolding - Introductory/"Did you notice?" screens	- Points out or explains a game design element to the player
Demonstration	- Describes the use of the red dots (which demonstrate the proper solution to the velocity regions)	- Teacher or peer takes over controls of the game to demonstrate a technique. - Teacher or peer says which controls to use out loud to the player - Watcher benefits in some way from watching partner play.
Frustration Control	- Uses the stabilize button to bring ship under control or resets a level - Uses the yellow gates to warp back to the start of a velocity challenge region, if failing it the first time	- Teacher or peer reassures or calms down player - Teacher or peer helps the player by clicking the stabilize button.

The last step before establishing inter-rater reliability was to describe the conditions under which a scaffolding moment would be deemed effective versus ineffective. A scaffold was deemed effective (and therefore not coded again) if one of the following segment conditions was met: (1) Student changed approach to play due to the scaffold and was more successful (fewer impulses/collision, etc.), (2) Student used the scaffold to describe normative physics content, (3) Student used the scaffold to connect to another context, like a different game or something in real life. Scaffold segments were deemed ineffective if (1) Student mentioned the specific scaffold did not help them, (2) Student did not see the scaffold, (3) Student used the scaffold to describe improper physics, (4) Student attributed understanding of something to another source other than the scaffold. Because some of these descriptions sound a little abstract, it is worth examining sample student statements which met each of these criteria (Table 4.2).

Table 4.2

Example student responses fitting each condition.

Segment Condition	***Example***
Student changed approach to play due to the scaffold and was more successful	"I went really slow the second time because if I used as little impulses as possible that would make me get a gold medal and it worked for this level." (Greg, solo)
Students used the scaffold to describe normative physics content	"If there's no friction then you need an equal but opposite force in the other direction to counteract the red arrow." (Elliot, solo)
Students used the scaffold to connect to another context, like a different game or something in real life	"I thought those white dots were like if I'm walking on a path and keep dropping bread crumbs trying to find my way back from grandma's house. If I'm walking they will be closer together but if I'm running they will be farther apart." (Maddie, solo)
Student mentions the specific scaffold did not help them (NE)	"I didn't think anything about those things while I was playing [just-in-time, pop-up scaffolding cues from the game] except how annoying they were--I had to be focused on my ship so I couldn't even read them." (Damien, dyad)
Student did not see the scaffold (NE)	"Whoa I never noticed that those red lines were always at a 90 degree angle I thought they curved." (Jeremy, dyad)
Students use the scaffold to describe improper physics (NE)	"If one of the red arrows [x- and y- component velocity vectors], like the length, if each of the arrows become really long then the yellow arrow [the resultant], well I'm pretty much sure it will always be longer, the yellow one. The yellow is the sum of the two red arrows." (Annie, dyad)
Students attribute understanding of something to another source other than the scaffold (NE)	"While I was playing I noticed the white dots and I understood them before you said anything. I knew the farther spaced they were the faster you're going and the closer together you are the slower they are going." (Enzo, dyad)

Much consideration was given to developing a scoring metric to award more "effectiveness" to scaffolding segments with direct connections to physics and less to scaffolds which made playing the game easier, or that made the player link to another context or game. A Boolean decision was more appropriate for a couple of reasons. First, there is no prior theory to suggest *degree* of effectiveness; instead, this study is more interested in how the relative uses of different scaffolds interacted with play and learning. Second, there are already a number of codes and adding more degrees on top of the current structure would introduce too many complications.

Once codes were established, my fellow researcher and I coded a subset of data including two full stimulated recall/observation videos (1 solo, 1 dyad) which meant coding two full 90-minute observation videos and three stimulated recall interviews (1 solo student, 2 dyad students), which represented a bit more than 10% of the appropriate data corpus. After ensuring that all names in the software were pseudonyms, I coded the appropriate documents and then my fellow coder followed by logging in as a separate user--she was unable to see my codes. Inter-rater reliability of 90% was achieved on this subset of the data for each level of coding. I commenced coding the entire corpus using the same coding procedures outlined above.

Data Manipulation

Coding in the described way made it relatively straightforward to run queries on the data to look for patterns. Each target student had his or her supports tallied across game, peer, and teacher, looking at the raw number of supports mentioned in the interviews or clearly and explicitly encountered in the observations. These raw numbers were not comparable because some students talked more than others and as much as I strived to keep the interviews as similar as possible, it was impossible to control students talking more or less than average or stopping the video over and over again to describe their thinking during the game (or never stopping the video). For this reason, an effective support metric was devised: one score for each type of support, and an overall effective support score.

Calculating the score was fairly straightforward: to calculate effective game support score, I queried the database for a coding matrix with target students as the rows and "game support" as column one and the intersection of "game support" and "not effective" as column two. I deleted column two from column one and then divided by column one to arrive at the game support score (essentially the ratio of effective game supports to all game support described/experienced). Game plus peer support scores were calculated, as well as an omnibus score including all forms of support.

In addition to arriving at three scores for each student, frequency counts of each type of support were tabulated, by group, by type, and by student, to search for patterns in the target student data set. Characteristics of students and supports with highly effective scores could then be explored along with the characteristics and supports with less effective outcomes. Exploring the data in each node included the process of analytical memoing, the result of which was two emergent patterns of support. Manipulating the data in these ways led directly to the analysis. The results from these methods are presented in the next section.

The Nature of Supports In and Around a Game Used in a Middle School Classroom Setting

The results of the analysis are presented in three parts. First, I compare and contrast the overall classroom enactments in the solo and dyad conditions. I then look at the different kinds of supports used by students in each class, starting with the solo classes and then looking specifically at how the dyad classes differed qualitatively. I then present the support scores for each target student and describe characteristics of the students with the highest and lowest effective support scores. Finally, patterns observed throughout the overall enactment across conditions are presented and described.

Overall Enactment

Students in every class were in their seats and ready to learn before the bell rang to indicate the start of class. Videos confirmed that I did not go off script delivering the vector introduction, so each student across the enactment heard the same introduction to vectors. Though some students had heard of a vector before, none reporting knowing what one was or what vector meant. Students were not familiar with the term impulse in physics, though some knew the word from other contexts (As Lidia questioned, "An impulse is a surge of something that you feel, but it's also a physics thing?"). The majority of the students had studied Newton's Laws in some capacity.

Solo Classes

As soon as students were released to go play the game, differences between groups became instantly apparent. In the solo classes, the room was very quiet; after all students had the game up and running, the lights were turned off to facilitate better image quality in the videos, and this choice may have contributed to the library-like atmosphere. One solo target student, Wallace, muttered to himself regularly saying things like, "Oh this is a disaster!" when colliding enough to die on some levels. Maddie, another target student, was enamored with the Fuzzies; "I like Fuzzies they are so cute!", she said aloud to laughter. She also thought aloud sometimes while playing. A full 80% of the solo players on camera exclaimed a variant of, "Oh, I get it" (i.e., "Oh, I see", "Ohhhh", "So it works like that!") within the first ten minutes of playing. This "Ah-ha" moment was when the student realized that once they imparted speed to their spaceship, they must decelerate by pressing the arrow key in the opposite direction in order to stop, since it does not happen automatically like in many common games. It should be noted that most students in the solo condition read the introductions to the first couple of levels of each control scheme, as it was visible that the student would stop playing to read the screen for 20-30 seconds before the start of each new level.

After this ice-breaking quiet period, the classroom became much more jovial and eager to discuss the experience, and students called me over to show off a score or explain how they figured something out. For example, Miles, not a target student, explained that he finally figured out how to get through a constant velocity challenge region:

> I can't move up or down either, I can't move at all because that will change my direction which changes my velocity. You're changing speed or direction when you apply outside forces on it.

By the third constant force level, Miles had connected pressing arrows to applying forces. My role in the classroom varied from troubleshooter to individual tutor to cheerleader to announcer. Students craved comparison points and often asked me, "Is this a high score?" Some students liked to brag to me about their high scores, and some boys (though not target students) would compare their scores for each level, in fun--I encouraged students to feel proud of gold medals and obtaining higher scores than they did the first try. I also made class-wide announcements which were not physics-related (e.g., "Hit alt-enter to make the game fill up the whole screen."). Most students asked for help at least one time, though sometimes that included simply getting the computer working. I mainly refocused students when they were having trouble avoiding constant collisions. The most common advice given was to try and restrict motion to either left-right, or up-down (I typically avoiding using horizontal and vertical unless the student used the words first). Sometimes saying that was simply enough, but other times students needed further advice (e.g., "If you're going too fast in one direction, try pushing the arrow key in the opposite direction.") or requested an outright demonstration of how to do it right. Some students were observed not asking for teacher help at all, though I did circulate to all students in the classroom at least twice to ask about progress and if they felt they needed any help to perform better in the game.

Mostly affective moments were captured during the wider observation, such as one student saying, "I like playing video games in class" and Maddie replying, "Yeah, we should do this more often guys!" Interestingly, students often commented on their experience and asked questions while playing and without looking away from the computer screen. As far as further peer interactions involving science, very little actual cross-talk between students occurred, though several instances of one student pointing out something or explaining something did occur. For example, one non-target nearest to the camera asked, "What are these white dots?" while I was busy helping another student get logged in. The girl adjacent to her came over and explained that they get further apart when you go faster because the spaceship is covering more distance in the same amount of time. Then both girls laugh as the ship crashes and the game gives them the pop-up message: "In constant velocity, the dots behind you are evenly spaced."

Interestingly Elliot, a target student, got to impulse level six first and started exclaiming aloud that it was very difficult. That level became the talk of the class and once students got to the level they would excitedly say, "I'm at level six!" The exact same phenomenon happened in the other solo class: Robert, a non-target student, got to impulse level six and verbally decried its difficulty, leading the class to commiserate with great amusement. Indeed, no solo student earned a gold medal on level six on his or her first attempt.

Over all, students in the solo classes had very positive reactions when asked about their enjoyment on a five point Likert-scale on the reflections (M=4.75, SD=.44). Only one student rated the game below a three. Students also rated the challenge of the game very near to the challenge of a typical science class (M=3.05, SD=.83, Challenge = 3 indicates the game's challenge is about equal to that of a typical science class).

Dyad Classes

After releasing students to the back of the classroom, choosing a partner seemed a foregone conclusion and students were on task immediately, with apparent animation and positivity. In fact, all four dyad classes were talkative throughout the enactment and it was often impossible to parse specific audio from the full observation videos.

Students spoke chiefly with their partners, but competitions did erupt between groups (usually at the same lab table) to achieve higher scores or better times on levels. One partner would typically communicate while another was playing--either the watcher providing commentary on play, or the player verbally describing his or her experience. Partners would also taunt each other (e.g., "My grandmother goes faster than you!"), act as cheerleader (e.g., "Don't worry you got this you're going to make it go go go go go!"), or provide guidance (e.g., "There's a constant velocity thing [challenge zone] coming up here so make sure you don't change speed or direction.")

My role as teacher was very similar to the solo enactment, however one marked difference was that a student would typically wait to ask a question while his or her partner was playing the game. I was also used to adjudicate arguments over perceived faulty game mechanics as comparison of scores was clearly more important (e.g., Spencer, a non-target student lamented to his partner John, "I ran into the wall after getting the last Fuzzy that's cheap that shouldn't count against me, right? I definitely did better than you!"). In two classes, I spent an inordinate amount of time with two students who were having play difficulties though their partners were not. For example, non-target student Iris was having enormous difficulties progressing and her partner (Sheila) was growing frustrated. I ended up demonstrating play for her (after trying verbal scaffolds and visual scaffolds with my hands) and then had her mimic back my techniques while I watched. She soon got the hang of stopping in one dimension, but the process repeated itself when she had to go diagonally through a constant velocity challenge zone. She had extreme difficulty lining up her ship so that it would go through without colliding, and never was able to consistently make her fingers do what her brain wanted. These two students in the dyad condition were the only students in the sample with next to zero prior game experience.

Again impulse level six emerged as a huge challenge. Interestingly, one non-target student was absolutely determined not to have any collisions in the game whatsoever and his determination paid off--he played slow and steady and was the only student across the sample to earn a gold medal the first time he played impulse level six. One class contained a dyad of boys (not target students) who were ultra-competitive and they contributed to a class atmosphere of trying to get the highest score on level six. This meta-challenge became a focus for three pairs of boys. Obviously it was expected that talking would occur within dyads, but it was interesting to see markedly more interaction between dyads than there was interaction between solo students. Playing the game was a primarily solo experience for solo players and most commonly a shared experience in the dyad condition, but level six brought classes together around a true challenge and elicited both competition and dialogue across the enactment.

Overall, students in the dyad classes had statistically similar reactions to the solo group when asked about their enjoyment of the game on a five point Likert-scale ($M=4.64$, $SD=0.53$). They also rated the challenge very near that of a normal science class ($M=2.92$, $SD=0.77$). There is no statistically significant difference between challenge and enjoyment scores between the two conditions.

Effective Supports

After completing the coding procedures outlined earlier in the chapter, data queries revealed frequencies of effective supports across scaffold type and student group. Similarities and differences in supports emerged from these queries and supporting evidence from the corpus is incorporated into the analysis to illustrate the findings.

Sources and Types of Effective Supports for Solo Students

Figure 4.1 displays the effective support from the game, peers, and the teachers for the solo students. Clearly, and perhaps unsurprisingly, the majority of support came from the game and for Greg and Elliot, both gamers, teacher support was not requested or redundant with game support.

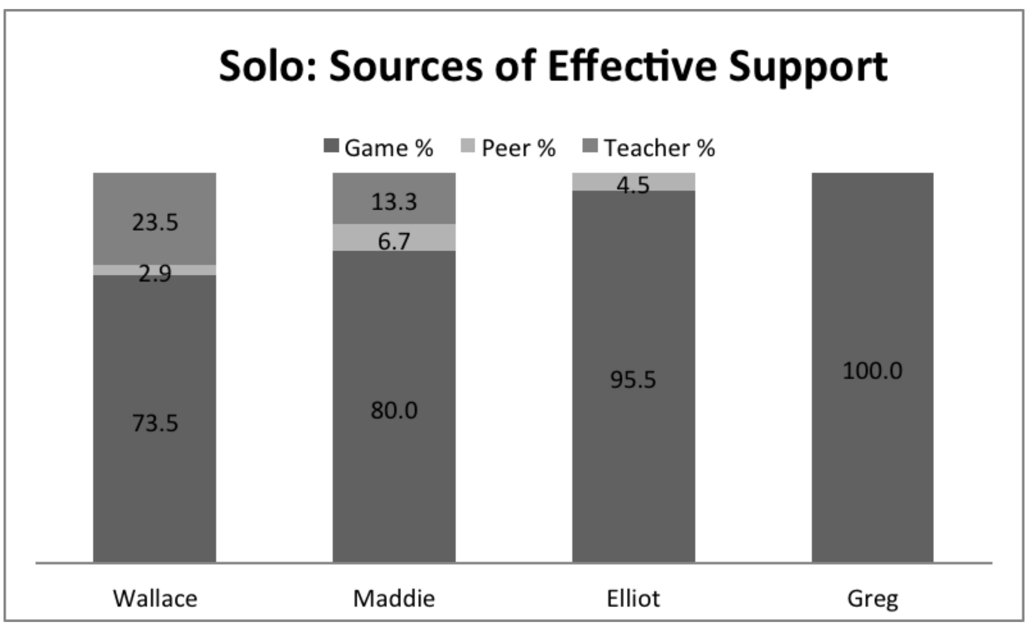


Figure 4.1. Proportion of each kind of effective support for solo target students.

Game supports.

Figure 4.2 provides a visual display of effective game scaffolds by type for the solo players. Marking critical features and refocus dominated, while the other forms of scaffolding were less present but still important in the overall portrait of supports. Recruitment (8%) appears low probably because motivation from the game was less apparent from the video observations than with the dyad groups. All four students mentioned in their interviews that they were happy when they received gold medals, but that the score did not really matter so much to them.

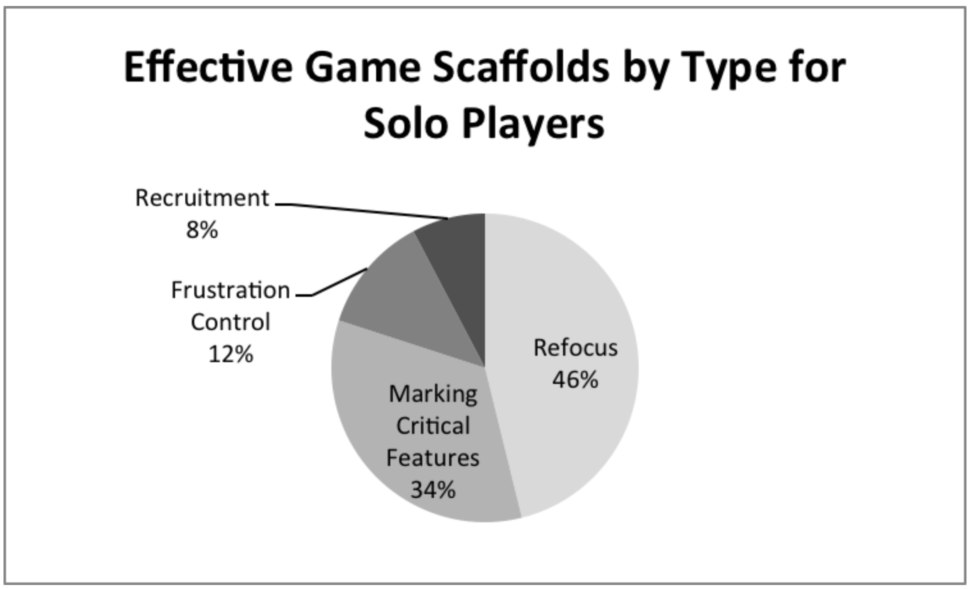


Figure 4.2. Pie chart depicting solo students' relative uses of each effective game support type.

The only aspect of the game that demonstrated anything to the players was red dots lining each velocity challenge region, modeling what the white wake trail dots should looks like when traversing the velocity challenge zones. The target students either missed them all together, or thought that they were boundaries, which is why demonstration (0%) was coded as ineffective in the solo target student sample for game support. Frustration control (12%), the final lesser-used scaffold, involved use of the stabilize or reset buttons to bring an out-of-control ship back in control, or allowed students to restart the levels. Additionally, hitting the exit wall after failing a challenge region reset the player directly before the challenge region, eliminating the need to double back and get in position again. Elliot liked the exit gate mechanic because it saved him the headache of "going back through that narrow region perfectly, only to turn around and do it perfectly again." Wallace, in his own words, "abused" the reset button as he would get very frustrated on the impulse levels and just start all over instead of watching himself die over and over. He said, "That helped calm me down--otherwise I would have gotten really angry and just want to stop playing." Other students never got unduly frustrated and did not need help keeping on goal.

The game was very good at marking critical features effectively (34%). Wallace had the most trouble of the target students, but like all students he used feedback from the game mechanics to inform his action as he said in his interview when we watched him fail to clear the first constant velocity region:

> "We learned that velocity is speed and direction so we learned that but I thought that I was smarter than the game and that it wouldn't see small changes so I thought if I changed direction just a little bit it wouldn't notice. I was surprised that hitting the wall counted against me, but basically I knew then that I couldn't change the yellow arrow in there." (Wallace, solo)

Evidently, the game marked the critical feature of the resultant velocity arrow and that helped him realize that any motion would change either speed or direction (or a combination of the two).

In addition to using the yellow arrow, students also reported reading the introductory screens, which contained direct physics information about velocity, impulse, and vectors, and also saw how the red vectors and yellow vectors interacted. All four students mentioned that the yellow arrow was "between" the red arrows (x- and y-components); one student, Elliot, explained, "I noticed that the yellow arrow kind of went in the average direction of the two red vectors." The just-in-time scaffolding windows, which popped up when players had trouble doing something or had too many collisions, elicited the most negative response. Though all students reported reading at least the first one, their regular appearances were considered "a nuisance" (Wallace), "really annoying" (Greg), "repetitive" (Elliot), or "in the way" (Maddie). Overall, they digested the scaffolding pop-up window which displayed after they died, but found their presence while playing to be problematic and disruptive to the play experience.

Critical features were also marked in the "Did you notice?" screens at the end of each level. A research design decision was noted as the reason this scaffold was not as effective as it might have been. Though target students uniformly read the first one, they were too distracted logging their scores after each level and neglected to read most of them, potentially missing out on important, explicit connections to physics. Similarly, students noticed the speed indicator in the lower left hand corner, but none used it for anything and never thought about it more than simply noticing its presence on the screen, except for Elliot who was the only target student in either condition to actually use the speed indicator to inform his play. This exception is explored more in Chapter 5.

The game was excellent at refocusing (46%) the attention of the student when the control scheme changed or when levels forced new strategies. All target students had strongly affective responses when the controls switched over to impulse: Maddie found the controls immediately easier, labeling them more "accurate":

> "The impulses were just these short bursts of speed, they happened the minute you press the button, but in constant force you held it and you slowly increased the speed over time. The control is different but easier for me."

The other target students described the controls as "restrictive" (Greg), "tighter" (Wallace), and "less fluid" (Elliot), but all students noted qualitative differences between what pushing an arrow did in constant force versus impulse levels. Impulse levels had much tighter corridors to traverse, so students were observed to adjust their strategies accordingly and systematically. This adjustment took more time for some rather than others: Maddie, who liked the new control scheme, adapted relatively quickly and stopped moving in two dimensions whenever possible in the very first impulse levels. Elliot and Greg went through a process of (begrudgingly, as they liked going as fast as possible) stopping completely at turns and then gradually got good enough never to stop, instead taking turns while changing directions with great fidelity. Wallace had the hardest time adapting and needed my help several times to learn how to restrict motion to one dimension, and then how to move at exact 45 degree angles, but he eventually settled on the stop and go strategy. The velocity challenge regions made students more aware of their speed and direction simultaneously: as the challenge regions got narrower and narrower, students had to be lined up perfectly before the region started in order to successfully get through without hitting a wall.

Peer supports.

Peer support was all of the marking critical features variety, and a very minor part of the overall support matrix. Maddie, Elliot, and Wallace all learned of a game feature from another student in the classroom who mentioned or described that feature out loud. For example, Wallace learned about the stabilize feature from a non-target student and used it to his ward off frustration when he started having trouble.

Teacher supports.

Teacher support was also much less prevalent than game support. The reason for this is quite obvious: target students are only four of the 22 total solo student players and I rotated among everyone. As a rotating presence, my scaffolding was not always needed or, in the case of ineffective teacher scaffolds, redundant with something already discovered by the student directly from the game. A pie chart displaying effective teacher supports is presented in Figure 4.3.

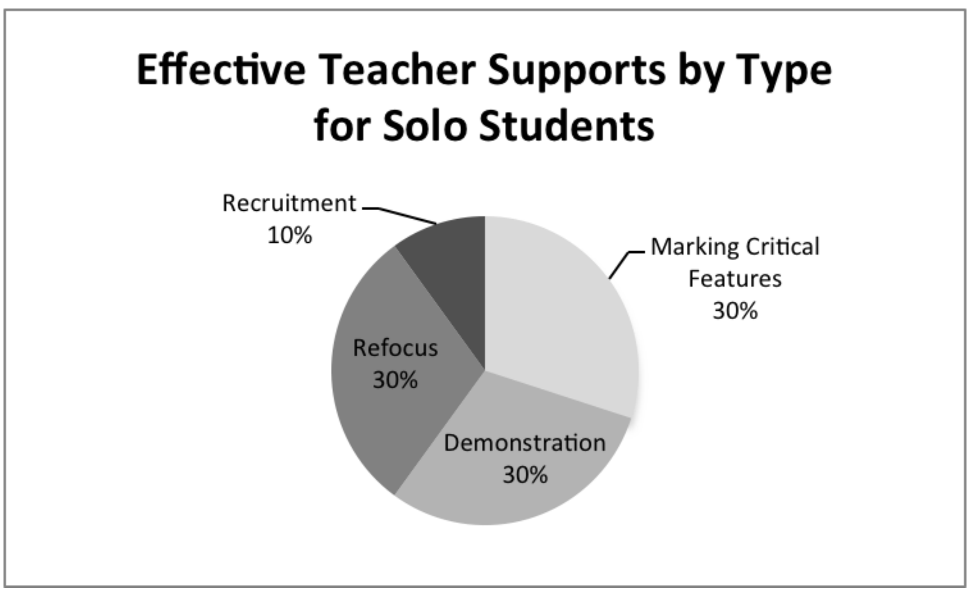


Figure 4.3. Pie chart depicting solo students' relative uses of each effective teacher support type.

Wallace used the most teacher support, needing both demonstration and marking critical features support as he struggled once he got to the impulse levels and had difficulty finding a strategy that ensured safety through narrow 45 degree and 90 degree corridors. He asked for some guidance and wanted me to show him how to stop. I demonstrated for him how to go forward and then undo that acquired speed by applying impulses in the opposite direction. As he took control back he asked me, while playing, how to deal with an upcoming narrow 45 degree corridor. I told him to apply impulses in equal amounts up and right from a stopped position to go precisely 45 degrees up and to the right. After failing to do what I told him, I had Wallace turn to me as I showed him using my hands how equal speeds right and up would result in a 45 degree angle. His reply to me extended the idea, indicating true understanding:

> "Oh so if I pressed 4 [units] up and 2 [units] left I would be going more up than left." (Wallace, solo)

Explaining to him how to navigate safely proved challenging while he was playing, but was relatively straightforward while not playing. He was able to combine the strategy of stopping, which I showed him directly, with the technique I showed him with my hands how pushing up and right would combine to a 45 degree angle.

A summary of the supports sources and their percentage effectiveness is presented in Table 4.3, for solo target students only. Refocus support from the game was 100% effective, indicating that the level design and different control schemes, restricting motion with impulse and allowing complete control with constant force, were extremely effective in making students rethink their approach to success. Strategy switching was apparent in all students. Wallace and Maddie reached the point where they could consistently travel diagonals and 90 degree turns without trouble, though they always stopped at these turns with their final strategy. Elliot and Greg, who are perhaps not coincidently self-described gamers, pushed their strategy even further and were capable of doing even impulse level six without stopping. As Elliott said of his strategy on level six, "Well I didn't need to stop here I could just use my speed to help make the turns more efficiently. It's like Newton's First Law of Motion: unless something's motion is opposed by something like friction or gravity, it will keep moving...yeah, inertia."

Table 4.3

Percentage of effectiveness for each observed scaffold for solo target students.

Solo Condition					
Support Source	Refocus (%)	Demonstration (%)	Frustration Control (%)	Marking Critical Features (%)	Recruitment (%)
Game	100.0	0.0	53.3	70.4	66.7
Peer	N/A	N/A	N/A	100.0	N/A
Teacher	100.0	100.0	N/A	60.0	100.0

Encountering narrow corridors meant careful navigation was paramount, so players who easily acclimated to the looser controls of force and played with great speed early on (Elliot and Greg) had to completely change their play style when entering the less forgiving impulse levels. Players refined their approach as a result of these level changes numerous times, making refocus the most common and most effective scaffold from the game. Marking critical features was the other most common effective support despite the lower 70.4% effectiveness rate. This phenomenon occurred because students did not notice all of these scaffolds (e.g., the "did you notice" screens), or found them to be annoying and unhelpful (e.g., the pop-up just-in-time scaffolding), but marking critical features was still a very effective way for the game to scaffold players in their physics understanding, connecting the game to the outside world, or helping them play better, when the conditions were right. Effective features tended to be ingrained into the play experience, and directly manipulable by the players (e.g., the wake trials, the vector cross). The game's recruitment effectiveness (66.7%) reflects that students were driven to succeed by the gold medals, but they did not so much care about the final scores.

Teacher scaffolding had higher success rates than game scaffolding, but the effect is not as strong as it may appear. Firstly, teacher support was only used by half of the solo target students and even then, much less than game support. The likely reason the supports were more successful is because a teacher does not leave the student's side until the student signals a level of changed understanding or a more effective play strategy. Note, however, that the teacher's marking of critical features was not as successful (60%), in effect because students had already noticed or understood what I was pointing out by gleaning that from the game. With the solo condition target students, peers only served to mark critical features and only then very rarely as peer support accounts for the least amount of support received by solo target students. Game support was paramount, and was supplemented by mainly teacher support, but some peer support.

Sources and Types of Effective Supports for Dyad Students

Figure 4.4 displays the effective support from the game, peers, and the teachers for the dyad students. From the display it is readily apparent that students in dyad classes had a more diverse support experience as peer and game support dominate for most of the target students, though some students like Bentley (41.7%), Siobhan (38.9%), Ione (30.8%), and Lara (50%) show significant proportions from the teacher. There was a richer interplay between support sources in the dyad classes. Students talked much more and the social component of play is clearly apparent by looking at the distribution of supports. Table 4.4 displays the percentage effectiveness for each support by source for the dyad students. This table will serve as an important reference when considering the results of the following analysis.

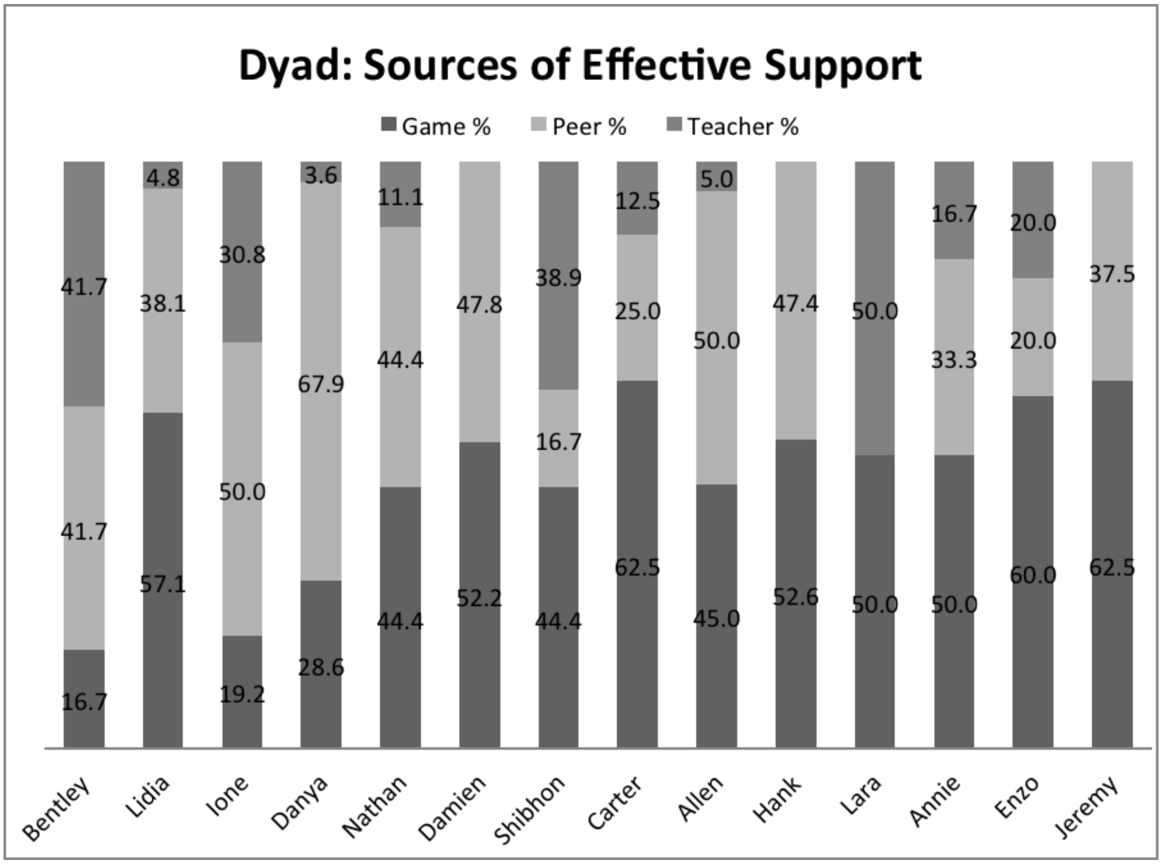


Figure 4.4. Proportion of each kind of effective support for dyad target students. Note that students are presented adjacent to their partners and roughly in the order of percentage of game experience.

Table 4.4

Percentage of effectiveness for each observed scaffold for dyad target students.

	Dyad Classes				
Support Source	Refocus (%)	Demonstration (%)	Frustration Control (%)	Marking Critical Features (%)	Recruitment (%)
Game	89.7	16.7	33.3	51.0	60.0
Peer	75.0	84.6	80.0	88.9	77.3
Teacher	100.0	100.0	100.0	77.3	100.0

Game supports.

Game scaffolds behaved in a way very similar to game scaffolds in the solo condition (see Figure 4.5). Again, students failed to notice the red dots, the only aspect of the game that demonstrates a solution outright--demonstration accounts for only a very small proportion of the overall game support (3%), and even then it is not especially effective (16.7%). Again, the game does not put demonstration scaffolding front and center so a low presence is expected, but it is surprising that so few students understood the intention of the red dots in both conditions in light of the number of students who did come to understand the meaning and use of the white wake trail dots.

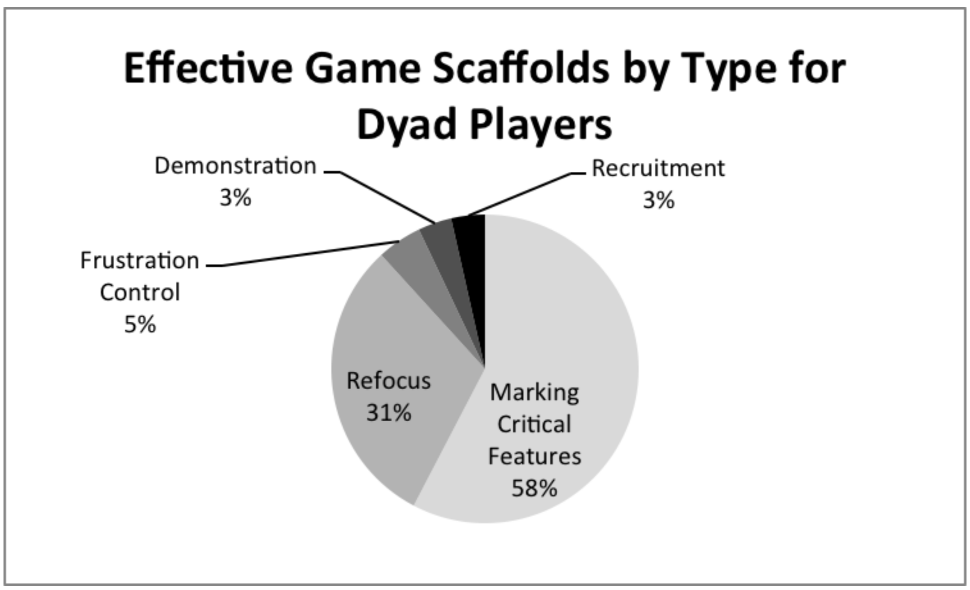


Figure 4.5. Pie chart depicting dyad students' relative uses of each effective game support type.

Some students reported being motivated by getting a gold medal, but again not in as high a proportion as the solo condition, where all students were driven to get the best medal. Marking critical features proved to be the most prevalent level of support (58%) but with a lower effectiveness (51%). The refocus elements (31%) like the velocity challenge regions and the 45 degree corridors led to strategy shifting, but not in as high a proportion as the solo students (46%) nor quite as effectively (89.7% rather than 100%). Frustration control was also used less in the dyad classes (5%, 33.3% effective), possibly because arriving at effective strategies happened more quickly in the dyad classes due to converging to the most effective partner strategy, so resets and stabilizers were not needed as much. All of these phenomena are a direct result of being in dyads, which is explored more fully in the next section by comparing game supports and peer supports of the same types and looking at how they interacted.

Peer supports.

Peer supports were as effective as game supports for many students, and a depiction of the peer supports by type is in Figure 4.6. It was more straightforward to do the analysis of the solo condition because the vast majority of the support came from within the game, but the dyad condition was more varied and support spanned across all three sources for most (10 out of 14, or 71%) dyad target students. The main qualitative differences between groups were apparent in the way peer support interacted with game support, so it is beneficial to consider Figure 4.6 in the context of Figure 4.5.

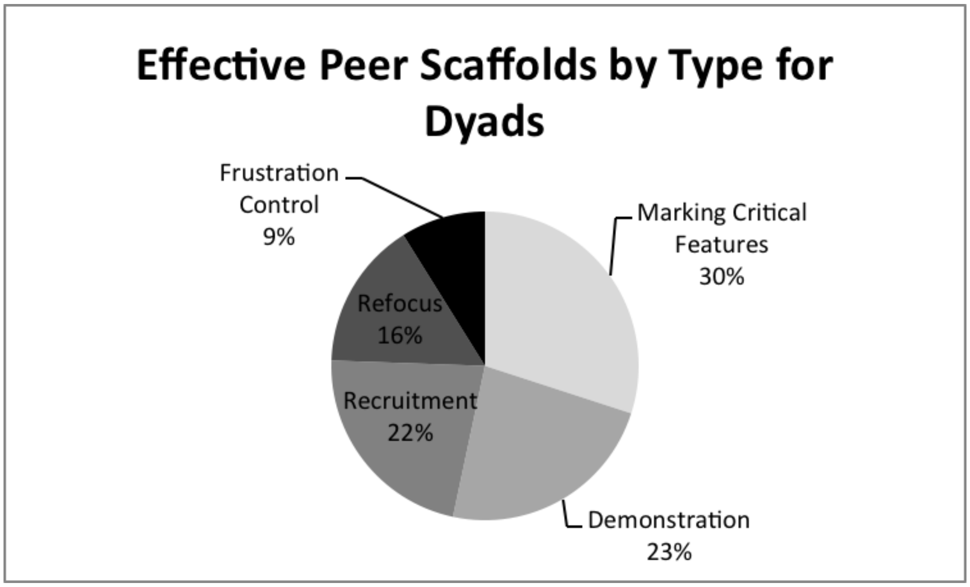


Figure. 4.6. Pie chart depicting dyad students' relative uses of each effective peer support type.

For example, recruitment is much higher here than in the solo condition because students were more typically interested in the goal of helping their partner succeed, beat their partner's score, or a combination of the two. In other words, recruitment came more from the peer interaction than the game. As an example, Lidia said, "I didn't care so much about the medals, but there were two levels I beat Bentley by one point. It was beautiful, beautiful! But usually we were just working together." Other groups competed, like Damien and Nathan: they had an intense rivalry and engaged in much jocularity. Damien was a racecar game enthusiast and this made them eventually concentrate the most on getting the fastest possible times. His desire for speed was infectious for Nathan who exclaimed, while going for top speed, "I'm allowed to crash once in a while, that is allowed!" even though he was only able to get through the harder levels with much greater caution.

We can also see that frustration control originated from the peer more than the game for the dyad condition. Dyad students were observed to use the stabilize button more frequently. This was a result of the fact that the game is controlled with the arrow keys, but the stabilize button requires a mouse click. Solo students did not have time to take their hands off of the controls to click "stabilize". By the time they went to click, they were already dead. For some students in the dyad condition, the watching partner was at the ready if an emergency stabilize was necessary.

The more interesting differences occurred with the marking critical features and refocus types of support. Peer groups used more of the critical features scaffolds for one major reason. The phenomenon resulted from the difference in the act of playing and watching. While watching their partners play, many students saw and read the scaffolding pop-up windows that were missed by the solo students, because they were focused on play. Importantly, they would not always read aloud or explain to the playing partner, though sometimes they would. For example, Bentley and Lidia were highly cooperative and mutually benefitted from watching each other play and articulate ideas about what the game features signaled: for example, while Bentley was playing, Lidia saw the scaffolding message that explained that white dots are equidistant when moving at a constant velocity. She immediately made a connection and said, while Bentley was playing:

> Lidia: Oh so the farther spaced they are the faster you're going and the closer together you are the slower you're going.
> Bentley: Stop talking I'm trying to concentrate!
> Lidia: Ok, ok jeez!
> (Bentley earns a silver medal)
> Lidia: Did you understand what I said?
> Bentley: No I wasn't listening.
> Lidia: It's like, if you're running and you're dropping breadcrumbs at regular moments, if you go faster they will be farther apart.
> Bentley: Oh, so that's what those dots are. (pause) Whoa I see, that's really cool, let me try that again.

Bentley not only commented that he liked the analogy, but proceeded to go to a velocity challenge zone and experimented with making the white dots further apart and closer together. He even tried to make them as far apart as possible by increasing his speed as high as possible. Lidia then proceeded to try and make a perfect square out of dots, which is not a very straightforward thing to do with the control scheme.

A similar pattern was discovered for the intro screens for each level--dyads were observed to read many more of these than solo students, probably because solo students were looking down writing down their scores but only one person in a dyad had to keep track of score, so the other partner was free to read the screen as the recorder was writing down the score values. It is worth mentioning again that the research design decision to have students keep track of their scores had an unintentional interaction with the potential effectiveness of the game's "Did you notice?" marking critical features support after each level.

Peer supports were also different in another major way. Much less strategy switching was observed for the dyad students overall. This seemed to result from one partner finding a fruitful strategy and the other mimicking that strategy, and indeed the partner with more game experience was typically the one to exhibit the best strategy sooner. While Hank (a strategy gamer, but only in the summer) was playing the very first level he explained to his partner Allen (an avid gamer),

> Hank: When you're going straight and you try to turn you keep going straight a little bit. It's like that question on the test [pretest] about the spaceship going in space and which path it takes. I said it was going to go like a line but here it curves, so constant force curves.
> Allen: I think when you move the arrow right, the lateral, you get the horizontal speed and then that combines with the vertical speed for a total.

Because Hank already developed a strategy of control and Allen was able to not only understand Hank but also extend his understanding of the controls, Allen exhibited a winning strategy at the very start. He never overshot a Fuzzy in the first level, always bringing his velocity down enough to capture each one with minimal wasted moves. When Hank picked the controls back up for level 2, he was immediately following Allen's strategy of increasing velocities and decreasing them systematically such that he was never going too fast in one direction. Allen was the more experienced gamer, and his strategies converged quickly to his preferred style, possibly because it was fruitful. However, both students contributed to the overall strategy using the game's refocus supports.

A similar phenomenon was evident with Jeremy and Enzo. Enzo played the first impulse level first and intuited that if he counted out loud he could always be sure he would stop. Jeremy echoed that strategy and can be heard in the videos audibly counting his impulses. When Enzo played the first level with 45 degree turns, Enzo explained to Jeremy: "You have to set up a 45 degree angle by pressing up once and right once." When Jeremy took back control, he replicated the strategy perfectly. Enzo reported playing games more than 12 hours a week, while Jeremy played half that amount. As Jeremy always played after Enzo, and did have facility with games, he was easily able to replicate Enzo's play patterns. However, impulse level six tested the limits of Jeremy's understanding of the strategy and he had a lot of difficulty getting through the level safely. He seemed unable to anticipate the moves required as quickly as Enzo; Enzo had developed the strategy on his own and when merely replicating the strategy, Jeremy had a hard time adapting to a harder situation.

Teacher supports.

Teacher supports (see Figure 4.7) were much less common than both game and peer supports for the same reason as in the solo condition. The main role of the teacher in the dyad class seemed to be as gap filler, which I will illustrate with two examples.

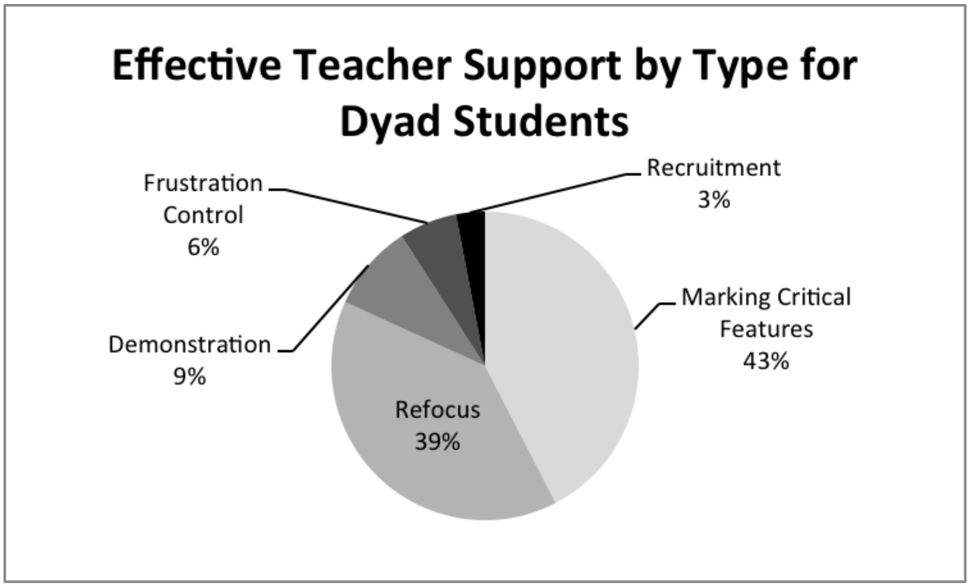


Figure 4.7. Pie chart representations of the effective teachers supports used by the dyad group target students

Looking back at Figure 4.4, it is clear that Bentley and Lidia received support in different ways, despite being paired (most pairs had more similar effective game supports). It was already established in the prior section that Lidia noticed and understood scaffolds from the game while playing and while watching Bentley play, hence her high game support percentage. Bentley found it very helpful watching Lidia before playing himself (though the game did inform his tactical strategies through refocusing his attention in velocity challenge zones, etc. as he mentioned that those were hard to understand until you came upon them yourself). Bentley called me over a number of times while Lidia was playing to ask me about his approach to playing and how it could be improved. For example, he could not perfect going diagonally and I showed him the idea of going up and over the same amount to go 45 degrees. When asked about this in the interview, Bentley said,

> "Um, yeah, that made me think about playing better. I was thinking that I would have to go sort of like stairs because it was so narrow, but then I saw

> that I could go diagonally safely at that exact angle, I just didn't realize it was so exact until you said that. "

Bentley seemed more at ease getting direct information and hints from a teacher and his partner than he did from the game itself. As teacher, I filled in gaps that Bentley missed from the game either because he did not notice them while playing, or he was not paying as careful attention to Lidia's play as she did to his play.

One instance of a non-gamer paired with a more experienced player (Iris and Sheila) was already described in the description of the dyad enactment, and a target pair exhibited a similar issue. Carter (an experienced gamer) was partnered with Siobhan (an especially inexperienced gamer) and was often annoyed at her slow play style. He reported learning only a little bit from watching her (she played the levels first) and that playing the game helped him much more, as she was unable to develop her strategies at a pace that matched Carter’s. Unsurprisingly, observations of Siobhan’s play revealed the lowest percentage of effective peer support (16.7%) of all dyad students, with her more effective use of teacher (44.4%) and game supports (38.9%). My standing by her side and aiding her through refocusing her attention down to a single dimension at a time, and demonstrating effective approaches to her supplemented the fact that Carter was, for the most part, unhelpful to her.

In terms of the effectiveness of teacher support, again the ratings were higher than game support chiefly because the teacher did not leave the student until new understanding or changed play resulted. Again, marking critical features was observed to be the less effective support from a teacher, because students already figured out the features from playing the game: they did not need additional help from the teacher to see the features. In the dyad condition, game and peer support dominated, while teacher support filled in the gaps, though some exceptions were noted where certain students required substantially more teacher support.

Summary

In this section I described the different kinds of supports available to students playing a game in middle school science classrooms, discussing the effectiveness of each kind of support from the game, the peer, and the teacher, and how those supports interacted. Generally, the teacher served a similar role in solo and dyad conditions, helping students with very little gaming experience learn the game more quickly and helping others when trouble or confusion arose. Peer support similarly supplemented game support but many instances were recorded of the observing player effectively using game scaffolding (e.g., the just-in-time feedback screens) or the non-data recording partner effectively using the "did you notice?" screens at the end of each level. The next section of this chapter looks more closely at how each individual target student effectively incorporated support from the game, the peer, and the teacher, looking at characteristics of the students with the highest and lowest scores.

Effective Support Scores for Target Students

Effective support scores were calculated for each student following the process outlined in the data analysis section at the beginning of the chapter. Results are listed alphabetically in Table 4.5. In this chapter, these values are used to qualitatively discuss difference between students in each group, but in Chapter 5 these values are considered in light of student gain scores from pre- to posttest.

Table 4.5

Effective support scores for target students coming from the game, the peer, and from all sources combined (game, peer, and teacher).

Student	Game	Game + Peer	Overall
Allen	0.69	0.83	0.83
Annie	0.50	0.56	0.57
Bentley	0.40	0.64	0.75
Carter	0.56	0.47	0.50
Danya	0.62	0.84	0.85
Damien	0.67	0.79	0.79
Elliot (solo)	0.91	0.92	0.92
Enzo	0.55	0.50	0.56
Greg (solo)	0.83	0.83	0.83
Hank	0.83	0.90	0.91
Ione	0.46	0.67	0.72
Jeremy	0.60	0.69	0.60
Lara	0.38	0.55	0.64
Lidia	0.80	0.87	0.88
Maddie (solo)	0.63	0.65	0.65
Nathan	0.50	0.67	0.69
Siobhan	0.73	0.69	0.78
Wallace (solo)	0.81	0.81	0.83

Organizing these results in a more meaningful way facilitates discussion of the characteristics of students with higher and lower support scores. Figure 4.8 displays the game support in order of least to most effective.

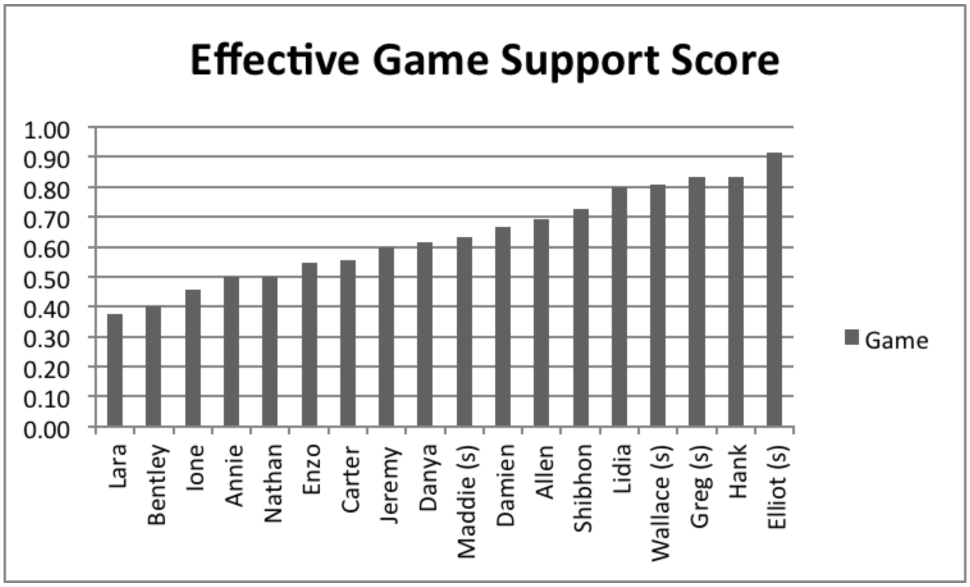


Figure 4.8. Effective support score for target students, least to most effective.

It is immediately obvious that solo students had higher game support scores, which is to be expected considering they did not have peer support to supplement or replace game support. Since this is in essence an unfair comparison between groups, it makes more sense to look at the data including total support from game and classroom sources.

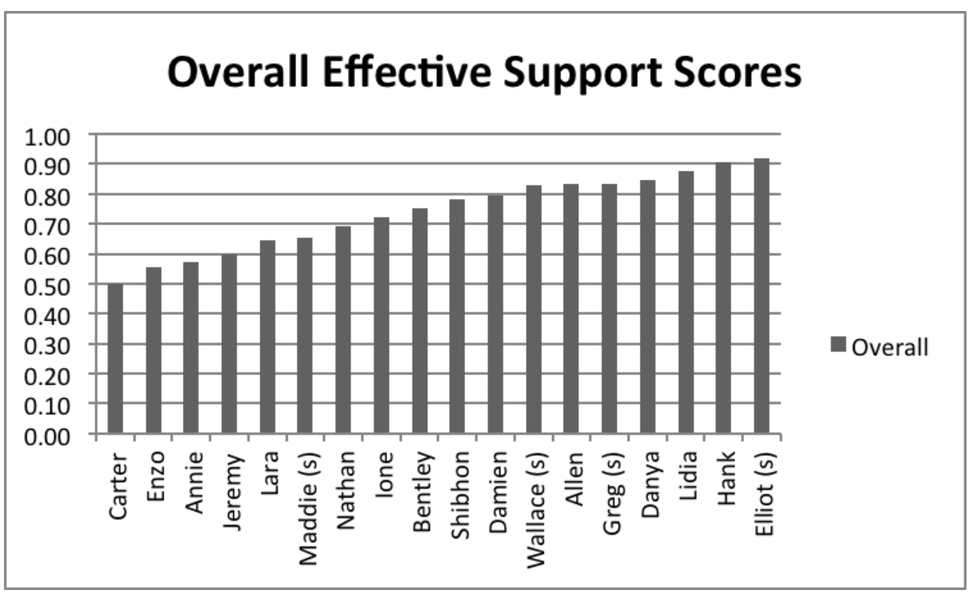


Figure 4.9. Overall effective support scores for target students.

The overall scores (Figure 4.9) reveal interesting characteristics shared by students on the low and high ends. Looking at the high end, Elliot from the solo classes had the highest proportion of effective supports. Recall that Elliot finished playing through the game early and spent time running experiments trying to obtain the highest scores possible. He controlled variables like number of impulses and saw how that interacted with time to figure out how the score at the end of the level was determined. These extra trials and additional strategies increased his understanding of the game's mechanics. In much the same way, Hank and Allen also finished first in their classroom and spent time doing similar experiments on the constant force levels. For Allen, game supports were not as effective because he attributed some of his understanding of some of the physics content in the game to prior knowledge (Note: Chapter 5 illustrates how Allen had one of the highest pretest scores and therefore had less room to improve than Hank). Lidia and Danya had the most fruitful partnerships. Lidia was a careful observer of her partner Bentley's play and allowed that to inform her play. Though Bentley's score was not as high, it is much higher than it otherwise would have been without a partner as seen by the larger that average jump from game support only to overall support. Danya and Ione had the most amicable partnership and while neither had a very high game-only support score, they were able to pull each other up with peer support. Just like Bentley, Ione's jump from game support to overall support was much larger than average, even though she did not score as highly overall as Danya since Danya was better at using supports from the game directly to inform her play.

Both Hank and Allen were game players, but so were Enzo and Jeremy who appear on the low end of the support scale. A closer inspection into their partnership reveals the reason for this. Like Allen, Enzo had a high posttest score--in fact, his score was the highest among the target students (10 out of 13 correct). His interview revealed that his father is a physics teacher and that he knew a lot of these ideas already from doing independent learning at home. He had never studied vectors before, but he had strong understandings of Newton's first law. This explains why his support score was so low: he was already confident with a lot of material covered in the game.

Both Enzo and Jeremy were avid gamers, though Enzo played about twice as much (12+ hours a week) as Jeremy (5-6 hours a week). Enzo played each level first and very quickly discovered effective strategies which Jeremy followed, as explained in the previous section. In fact, both Enzo and Jeremy were shown to change strategies as a result of the game less than anyone as Enzo found viable strategies so quickly without much trial and error. Jeremy bumbled through impulse level six and was never able to earn a gold medal as he never broke out of simply copying Enzo. In fact, in the interview, Jeremy admitted that he "never really understood" the velocity challenge zones because he would just do what Enzo was telling him (e.g., "Speed up now! Slow down now!").

Annie and Lara were partners and were quite low on the overall effective support. As mentioned when initially introducing target students, both girls played games casually (about 10 to 20 minutes a day) and had an awkward partnership. They communicated very little but did both report that watching the other was helpful. They engaged in very little strategy switching and seemed to be playing with mostly guess-and-check methods throughout the levels, slowly but surely beating each one. Though I helped both of them understand the function of the white dots, and demonstrated to Lara how to navigate at 45 degree angles, neither of these led the girls to critique their play in context of science. The supports for Lara and Annie were initially confounding when held next to the results for Danya and Ione since they all had similar game play habits outside of school. Danya and Ione listed their favorite games as *Osmos* and *Angry Birds*, and both Danya and Ione mentioned these games in the observations and interviews, connecting motion control in those games to the controls in SURGE. Perhaps Danya and Ione were prepared more adequately by their prior experiences in other physics-based games. Danya serendipitously described the game as "pleasantly frustrating" when trying to figure out how the controls differed from *Osmos*,[3] a phrase that actually appears in the literature (Gee, 2005) and which Ione agreed with by exclaiming, "Totally!"

When a game is not too hard and not too easy, it has the quality known as flow (Csikszentmihalyi, 1990), and Danya (and to a lesser extent Ione) exemplified the Vygotsky quote that introduces this thesis: "In play it is as though the child were trying to jump above the level of his normal behavior." The girls combined the game experience with their prior experiences in other physics games to overcome the frustrations because new understandings were possible to reach with the supports granted by the game and classroom. Perhaps Danya and Ione were prepared more adequately by their prior experiences in other physics-based games. Annie and Lara did not have such reference points, or at least did not mention them during the enactment of the research, and had trouble ever breaking past a guess-and-check strategy.

[3] This coincidence nearly knocked me off my feet and I immediately asked Danya if I had said "pleasantly frustrating". She assured me I hadn't and it originated from her, and her claim was validated by the observation videos. She said it spontaneously.

The lowest support score was for Carter, which in context of his partnership with Siobhan may be expected. What is so surprising is that Siobhan's support score is so much higher than Carter's, given that Carter has so much more gaming experience, playing Call of Duty: Modern Warfare 3, a first person shooter, cooperatively online regularly. Carter seemed to be the type of student who enjoyed gaming for the cooperative experience ("It's fun to play with friends," he said in his survey about what he likes about playing his favorite game). Closer inspection of the observations and interview reveal that Carter was visibly annoyed and even moderately angry about the amount of time it took for Siobhan to learn to play the game. She required multiple, extended interactions with me, and as a result he had much less time with the game than other players. Additionally, Carter was quite confident in his abilities to play the game in context of Siobhan's obvious novice-status. Carter's arrogance was not beneficial as he completed levels without much thought for strategy as he could complete them much more easily than Siobhan. Even at impulse level six he was using the same start and stop, jerky strategy that he started with early in the game. Carter may have benefitted from a partner that would have challenged the effectiveness of his strategy.

So why was Siobhan's score so high compared to Carter, given her almost complete lack of game experience? The other girl in the sample with almost no game experience, non-target student Iris, reported the only enjoyment Likert-score less than three, rating the game as a two, less enjoyable that a normal science class. She declined three points from pre- to post- test, the most of any student across both groups. Indeed, Iris said in her reflections, "Frankly, I never enjoyed video games. I really don't play them for fun and I would rather learn by reading and hearing lectures." Even though Siobhan also had essentially no game experience just like Iris, she loved the game and rated her enjoyment and challenge level as the maximum of five. She wanted to be good at the game and used the game's representations and teacher scaffolding to evolve her strategies and play, eventually using more sophisticated (yet always slower) methods than her faster, more experienced partner Carter.

Summary

Exploring the data from each condition, I attempted to show how students' uses' of supports available to them through the game, the peer (if applicable), and the teacher varied by type and by group. The analysis revealed ways in which supports interact, for example how students in dyad groups see more critical features because they watch the play at times, but also how they engage in strategy switching less often because they converge to the most successful strategy between the pair. In general, students in the solo groups used more supports from the game than those in dyads, who used a mixture of peer and game supports to play. Teacher supports were used when students needed help controlling their character, or when they needed help understanding a game design element. The game was most effective at refocusing the attention of the player through level design, including the constant velocity regions and the narrow 90 and 45 degree turns, all of which required strategic alterations to play. Marking critical features was the most common support element overall, but those features which appeared "just-in-time" were not as effective for active players, and "did you notice?" screens were not typically useful for solo players or partners taking down data (unless the peer explained them to the partner after recoding data). Peer demonstration was used prevalently in the dyad condition as students found it helpful to watch their partners play levels to plan ahead for their own play experience; while watching their partners these students were more likely to read the just-in-time scaffolding as well.

I also looked at the effective support scores for students and tried to pick apart similarities and differences between students at the high and low end. From the results of these analyses emerged two patterns of support that occurred across both the solo and dyad conditions as the game was played in the classroom setting. The first pattern explores the need for content scaffolding to happen outside of play. The second pattern explores how prior gaming experience interacts with the use of and effectiveness of supports.

Pattern One: Variable Effectiveness of Just-in-Time Scaffolding in Solo vs. Dyad Conditions

Just-in-time scaffolding is a powerful tool when students work with simulations (Cox, Belloni, Dancy, & Christian, 2003). But one marked difference between a simulation and a game is the fact that simulations allow incremental control, often letting the user rewind time, watch in slow motion, look in incremental steps, or some other variant of manipulating time to observe a cause and effect relationship. A game does not normally allow for this type of manipulation because it takes the player out of the game experience (though time manipulation can certainly be a purposeful game mechanic, as in the popular game *Braid*). Indeed, dyad target student Allen remarked that "it would be cheating" if the game somehow paused the game action to allow you to read the pop up just-in-time scaffolding.

Across all support types, unless the student was directly manipulating the support mechanisms to traverse the levels (e.g., vector arrows, wake trail pattern) they were ineffective. Much of the content explicitly connecting physics to the game is presented before each level and after each level. However, just-in-time scaffolds appear at the moment they would ostensibly be most helpful for the player. When the spaceship has two collisions in quick succession, a window pops up providing advice on how to control the ship more effectively. For example, one scaffold says "Use opposite impulses to slow down" among a number of others. Across both groups, players said they usually read the first but the others were ineffective. In pairs, the non-playing partner was able to read these and found them effective. For example, Hank from the dyad group explained,

> "I never actually read those things when I was playing but when I was watching I looked at them. I had to be focused on my ship. But that's how I knew that I needed to apply an opposite force to slow down, reading one of those while Allen was playing before I did."

Allen echoed,

> "I could never read those while I was playing, they would be better as little panels during the objectives. I think having them pop up during the game play was distracting and detracting from the experience. I could only read them when it was Hank's turn to play."

So it is not that the scaffold in and of itself is ineffective, but in the way it is presented. Peer supports worked in precisely the same way. For example, while Siobhan was having a very hard time navigating through her first diagonal region, her partner Carter said,

> "You have to stay constant in that area you can't change like that, you keep changing every time. You have to start going and then stop pushing anything until you're through it because you have to keep your speed *and* direction constant. You have to set it up carefully."

Everything that Carter was saying was true, but Siobhan was fully in the act of playing and trying to correct her motion and stop herself from dying as a result of too many collisions. When asked about this exchange in her stimulated recall interview, Siobhan explained,

> "What he's [Carter] saying makes sense, I mean, he's right. But at the time I didn't know how to control it and to be honest I just was ignoring him. I thought when you pressed down it would go down, you could just press down once, and then I didn't know how to stop it. After the level he told me, yeah he helped me, he was the one that actually told me that if you want to stop you have to press in the opposite direction."

This moment is particularly illustrative of the pattern because it shows us that the state of play determines the effectiveness of the scaffold. Carter's help did not register at all while she was playing, though Carter helped her understand play by pointing out a crucial game mechanic after she was not actively playing anymore.

Though teacher supports were more rare across the sample, the same pattern recurred. An example was already presented of working with Wallace in the solo groups, where I tried to explain how to come to a complete stop and then move at a 45 degree angle, but was unsuccessful in relaying that to Wallace until he actually stopped playing and looked at me. In the dyad group, when Enzo asked me to explain the wake trail dots to Jeremy because, as Enzo said, "I don't know quite how to say it right, but I do understand it." Importantly, Jeremy was playing the game while I explained to him how it works. Jeremy replied with an "Ohhh...I see, I get it now" but sounded suspiciously unsure despite repeating that he understood. I knew to mark this moment for his interview and indeed, he confirmed my suspicion:

> Me: What did you think when I explained that to you here?
> Jeremy: To tell you the truth I wasn't really listening.

Support within the game that actually guides the player through levels or supports which are directly manipulable by the player are effective because they are a necessary part of the play experience, as they are endogenous features. The design of levels can help draw the attention of the player away from extraneous information or other distractions to focus in on important pieces of the game to affect strategy switches. However, other scaffolds more directly connect to pieces of physics content or present strategies in a way that contextualizes the game with normative physics. This latter type of scaffold must be presented while the player is not engaged in actual play, but rather between levels or when a player is not in a state of directly controlling a character. This may not be such a problem in some types of games where starting and stopping is more common, but in a more action-oriented game like SURGE, the method to make the pop-up scaffolding more effective is not as clear-cut. One student suggested that every one that would have appeared while you played a level should appear after you die, and you should have to click on each one to acknowledge that you read them. This is only problematic in that it does not allow immediate and *in situ* diagnosis of the player's errors, but they do not seem effective in the current state, especially in solo play, and a new delivery system for the information in these scaffolds is desirable.

This pattern lends credence to Huizinga's theory of the "magic circle" of a game (1938). Game design elements that are central to the play experience prove useful to the player, but additional scaffolds that are overlaid on play, either literally through on-screen scaffolds, or when teachers and peers try to support play while the other is playing, are ineffective because they disrupt the sacred "magic circle" enveloping the player and the game. Another way to interpret this pattern is through the conceptual change model. A typical requirement for conceptual change is said to be that the learner must find a newly introduced idea or model fruitful for future pursuits (Pintrich et al., 1993). Students playing the game were recruited by the game and, if applicable, their peers, to achieve the goal of the game and get gold medals. Following this "hot" model of conceptual change (Pintrich et al., 1993), any newly introduced idea or support must be fruitful in helping the student get closer to attaining the final goal, in this case a higher score or a gold medal. Pop-up scaffolding was merely a distraction for players as it hindered rather than improved play. Indeed, across both groups and all three sources of support, scaffolds were only effective when they were directly useful to the player in better attaining game goals or physics understanding, and when occurring outside of active play.

Pattern Two: Prior Gaming Experience Interacts with Both the Use and Effectiveness of Supports

The prior analyses include many examples of supports interacting with prior gaming experience. To examine this perceived pattern, I split the group by overall effectiveness score and considering scores "high" for 70% effectiveness and up, and "low" as less than 70% effectiveness. Game experience was gauged based on prior play reported in the introductory survey: 0-2 hours per week was considered "low", 3-6 hours was considered "moderate", and 7+ hours was considered "high" (no students fell between these number sets). There were two exceptions to this rule: Siobhan played her very first video game just two weeks prior to the enactment of the research, and only then for about twenty minutes total over two weeks. Her game experience was marked as "N/A" as she had much lower experience than even students marked "low". Hank was the other exception, as he reported only playing games for less than an hour a week during the school year, but many more than that (ten or more a week) during the summers. As such, his play was averaged as "moderate".

Table 4.6

A glimpse at the interplay between prior game experience and support score. Students are paired with their partners where applicable. Dyad students are grouped together.

Target Student	Game Experience	Support Score
Annie	*Low*	*Low*
Lara	*Low*	*Low*
Maddie (solo)	*Low*	*Low*
Danya	*Low*	*High*
Ione	*Low*	*High*
Lidia	*Low*	*High*
Bentley	*Moderate*	*High*
Carter	*Moderate*	*Low*
Siobhan	*N/A*	*High*
Elliot (solo)	*Moderate*	*High*
Damien	*Moderate*	*High*
Nathan	*Moderate*	*Low*
Hank	*Moderate*	*High*
Allen	*High*	*High*
Enzo	*High*	*Low*
Jeremy	*High*	*Low*
Greg (solo)	*High*	*High*
Wallace (solo)	*High*	*High*

Moderate or high play experience tends to align with a higher support score, but somewhat complex interactions emerge when looking at the cases where a moderate or high score results in a low support score (Nathan, Carter, Enzo, Jeremy), or when a low experience rating aligns with a high support score (Danya, Ione, Lidia).

Moderate-High Game Experience with Low Support Score

Enzo was already explained in the previous analysis: he had prior experiences which rendered many supports redundant with his current conceptions, and despite a low support score he played the game at the same level as the very best players (e.g., never stopping or colliding).

Carter was also explored previously: his partner Siobhan's lack of experience frustrated him, and when I was helping Siobhan he did not feel the need to listen and benefit from my descriptions because, to his mind, he did not need help playing the game. Admittedly, Carter played much less time than other students because Siobhan needed so much more help. Carter picked up the game much more easily than Siobhan, but he had no urgent need to improve his play as he perceived that he was so much better than his partner. Indeed, after watching a moment where I was helping Siobhan play by describing how to move at 45 degrees I asked:

> Me: What did you think of what I was saying to Siobhan here?
> Carter: I was just really getting frustrated because she couldn't play right and it was taking forever.

This apparent feeling that solid play was proficient enough not to need extra support extended to Nathan and Jeremy.

Both Nathan and Jeremy also exhibited the same pattern of a lower support score with more game experience. Closer inspection reveals this resulted for similar reasons. The case of Jeremy was already introduced: recall that he quickly followed Enzo's lead and never developed a strategy of his own to play, and because he so easily picked up the game play he did not have much reason to critically reflect on his approach, as explored below.

To start, Jeremy mentioned in his interview that when he and Enzo were filling out the scaffolding sheets that went along with the game, he never really understood them and that Enzo just filled them out while Jeremy was playing. When he got to impulse level six and had a really hard time, Jeremy said, "I was just trying to get through it, I didn't really know how to control it". Impulse level six adds the very first instance of a narrow, 135 degree angle turn, which required new adaptation from the simpler 45 degree and 90 degree turns. Jeremy was able to get through safely by abusing the stabilize button, and therefore never had to adapt his strategy. In other words, Jeremy's prior game experience harmed him because he was adroit at spaceship control and did not need the game's representations to learn to play better. He was naturally fast at picking up games because of his experience so Enzo's support was mainly restricted to affective statements of cheerleading: " Up up up up up, not too much not too much!". Additionally, Enzo was observed to read the introductions to several levels aloud very quickly. I asked Jeremy:

> Me: What did you understand about what Enzo read here?
> Jeremy: I don't know, nothing really.

Enzo was assuming a lot about his partner's understanding. Enzo could have shared his knowledge with Jeremy, but because Jeremy was an experienced player he was able to mask his low level understanding of the physics involved. There was no real, fruitful need for the content because he was doing fine without it. When he got to the one place where he had trouble (the 135 degree angle) he simply bore through it by using stabilize and never revising his method.

Nathan followed the same pattern as Jeremy, but for different reasons. Nathan was extremely competitive with his partner Damien. It frustrated Nathan (a fan of fighting games) that Damien (a racing game enthusiast) picked up the game so fast. Nathan played the first couple of levels slow and steady despite his partner's speed, but by the time they started playing the impulse control levels, Nathan was doing his best to mimic Damien's speedy play. During our interview, Nathan got so animated while I was fast-forwarding to a clip that he excitedly asked me to stop the video by exclaiming,

> Nathan: Oh stop here and watch this one! Damien was so bad at this one and I was so good.
> Me: Damien played first?
> Nathan: Yeah watch how much better I am! I was really motivated by the score. See how I'm hugging those corners?

What is interesting about this segment is that Nathan is using the phrase "hugging those corners" that he heard Damien invoke while speeding through the previous level. Damien himself made profound connections from the game to his own play experiences. When I asked him what strategy he was using to make 90 degree turns, Damien replied, "That's just instinct from, I really like Go-Kart and racing games and you can take turns kind of fast but you always have to slow down a bit first, and in this game you do that by pressing back instead of pushing the brake button." Car games specifically require the physically normative behavior of slowing down to stop in a specific direction.

Nathan did not have such experiences to draw upon, or at least did not mention them. He was not making the same connections as Damien and did not get the same benefits from the game, connecting its scaffolds to his play. He was merely trying to match Damien's abilities and begged me to watch the one instance where he defeated Damien's score. As Damien said, "The only reason you beat me is because you kept hitting the reset button until you had a perfect run." Basically, Nathan was "brute-forcing" the game in a desperate act to stay competitive with Damien, and by focusing so much on the competition he failed to capitalize from support. After all, he did not need the support because he was able to do very well just following Damien's lead, much like Jeremy.

Damien found watching Nathan play instructive because he learned more about the game: "I saw that Nathan was getting 20-30 less points than me by going slower using less impulses, so I figured out a way to go faster and still use less impulses by not stopping all the way at corners." Nathan said of watching Damien, "When I saw that I knew that I just needed to go faster."

Low Game Experience with High Support Score

Danya and Ione were already described in the previous analysis: while their game experience is characterized as low in frequency, they used that time to play physics-based games *Osmos* and *Angry Birds*. *Osmos* is a momentum-based casual game where the player ejects a small amount of mass from a suspended central bubble in a desired direction, making the larger bubble go off with the equal but opposite momentum of the smaller ejected mass. The aim is to direct the bubble into other bubbles, making as large a mass as possible, without using so many ejected mass impulses that you lose all of your mass (which is the losing condition). *Angry Birds* involves projectile motion using a very simple slingshot mechanic where birds are ejected at the desired elastic tension into a construction of objects. The goal is to knock all appropriate objects off the construction with as little bird ammunition as possible.

Both girls were observed to talk about these games in direct comparison to SURGE, and each mentioned *Osmos* in the interviews, comparing the impulse mechanic in SURGE to the control mechanic in *Osmos* (a highly apt comparison, though *Osmos* is more complicated in that mass changes complicate the momentum changes). Clearly having this prior experience to contextualize their game experience supported their play. Annie and Lara, the pair with low experience and low support scores, did not have such experiences to draw upon. It is also prudent to mention that Annie and Lara were an awkward partnership, communicating the least of all dyads, and that Danya and Ione were one of the most talkative pairs. Surely this contributed to the large difference in peer effectiveness for the pairs.

Siobhan was explored at length as well in the previous analysis. Because her attitude about playing the game was so positive she was receptive to its support and she felt that learning the content and strategies improved her ability to reach the game's goals, and she used supports to help learn how to play the game effectively. In fact, she did not have trouble with the content in the game. For example, after watching her play the first level, I asked:

> Me: What did you notice about how the controls work here?
> Siobhan: When I first started off I didn't know how to control it. I thought when you pressed down it would go down, you could just press down once, and then I didn't know how to stop it. I figured out that if it's going to the left too fast I needed to press right, but I couldn't do it exactly right.

In other words, Siobhan began to intuit realistic inertia was present in this game which she was not expecting. She already started realizing that she needed to press right to undo a leftward motion, but because of her slight diagonal motion she was unable to translate her thinking into the game; her lack of game experience limited her ability to translate her understanding into complete control of the spaceship. This realization planted the seed that would eventually grow when she noted that Carter "was the one that actually told me that if you want to stop you have to press in the opposite direction." She learned through a combination of supports that motion in any direction required a force in the opposite direction to "undo" that speed. Siobhan learned the game mechanics out of necessity to give herself a fighting chance of finishing the levels, whereas Jeremy, Nathan, and Carter ignored some game supports because they did not need them to play well enough.

Finally, Lidia is another example of a student with low game experience and a high support score. Lidia proved unusually observant while her partner Bentley played, always reading the pop-up just-in-time scaffolding and even connecting it to prior experience, to both her and her partner's benefit (e.g., the "breadcrumbs" connection). Most other watching partners simply read the pop-up windows rather than talking about them after play, reporting they were helpful in the stimulated recall interviews but not really poring over the information with their partners in the moment. Lidia also picked up how to play the game much faster than casual gamers, like Danya, Ione, Annie, and Lara. It was observed that despite her report of playing fewer than one hour of games per week, she mentioned in the observations and also in the interview that when her cousins are home in the summer, they play Nintendo Wii games all the time. I never thought at the time to ask for amount of time played in the summers, but it seems likely that Lidia had more game experience than she let on in the self-report introductory survey. The Wii specifically uses motion as its central control mechanism which could have prepared her for some aspects of motion in this game.

Other Cases

The other observations show that students with low game experience had low support scores and students with high and moderate experience had high support scores. Taken together with the other instances described in this section, it is clear that game experience effects the use and effectiveness of supports in and around a game used in the classroom. Because of these qualitative observations, game experience is fully examined as a possible predictor of pre- to posttest gains in Chapter 5.

Summary

In this chapter I explored the uses of and interactions between different sources and types of support in and around a physics game used in middle school classrooms. Students in dyads used a broader range of supports to supplement game supports, though some solo students also used teacher supports when supports from the game were insufficient at attending to the student's needs. Students' effective support scores were examined carefully to note obvious differences in support use for students at the high end and low end.

Two major patterns emerged as part of this analysis. It was discovered that real-time scaffolding overlaid on top of the play experience is ineffective compared to scaffolding which is ingrained in the play experience or which is supplied outside of active play. The second pattern involved the rich interplay between previous gaming background and support use, positing that gaming experience, and indeed the type of gaming experience, can influence the play experience in terms of which supports are used and how effectively. These patterns have profound implications for teachers, researchers, and game designers, and they will be discussed in Chapter 6 which looks at the results across the entire dissertation.

CHAPTER 5

THE RELATIONSHIP BETWEEN GAME SUPPORTS AND STUDENT LEARNING ACROSS GROUPS

In this chapter the second research question is explored: How do the learning gains in the collaborative condition compare to the solo play condition, both qualitatively and quantitatively? Student learning was assessed formally through the pre-posttest to arrive at a numerical gain score from pre- to posttest. Student gains are compared overall and for each group individually. I undertake an exploration of the data with an eye towards game experience differences, as differences in game experience emerged as a major pattern in the data in Chapter 4. Then, I determine if there is a relationship between effective support scores and pre-posttest gain scores.

After analyzing the data in this way, individual subsections of the pretest are compared to see which areas had improved learning gains: impulse, constant force, and/or vector understanding. Next, I take a closer examination of the items that resulted in significant gains, comparing to previous studies. Finally, I undertake an examination of the items that resulted in significant gains, as well as some that did not, in the context of the qualitative data.

Data Collection and Analysis Methods

Quantitative data sources included the pre- and posttest, student data sheets, and any numerical value on the introductory sheets and reflections sheets (e.g., hours of game play per week, Likert scale challenge report, Likert scale enjoyment support). Unfortunately, not only were the data sheets a distraction to play, as explained in Chapter 4, but also an unreliable measure of student performance. Data sheets were only to be used if there was a problem with the server collecting data on play. Because of a school-related issue, laptops could not access the game properly while online--computers had to be taken offline to play the game locally. This necessitated the use of the data forms, but from spot-checking the data reports of the target students, comparing their logs to the actual videos, major problems surfaced. Students often wrote numbers in the incorrect place, they would regularly remove one or two collisions to make their data look better, and most importantly, they did not keep track of trials where they started over. Because the data lacked any kind of reliability, it was not used as part of the analysis.

The multiple-choice assessment items (Appendix G) were scored in aggregate on a scale from 0-13, where no partial credit was granted for incorrect answers. Student scores and all other information were input into a large data table in IBM SPSS for analysis. Group placement, target student status, and gender were input as nominal Boolean variables. Game experience was input at the three ordinal levels indicated in Chapter 4: 0-2 hours per week, 3-6 hours per week, and 7+ hours per week. Challenge and experience were input as one of five ordinal levels as they were Likert-scale items.

Before any analysis was attempted, Levene's test for the Homogeneity of error variances was run on the data. Levene's test was negative; in other words, the error variance on pre- and posttest scores was homogenous across the solo and dyad groups. It was also established that target students and the rest of the population had homogenous error variances, so we can be sure that target students were generally representative of this population of 8th graders. In much the same way, no single classroom differed from any other classroom, meaning each class of the six class sections was representative of the overall population.

A repeated measures ANOVA was used to analyze the learning gains from pre- to posttest, as well as the gains on each subsection of the test (i.e., impulse, force, and vectors). Game experience was added to the repeated measures model to determine if there was a main effect on score gains as a result of prior gaming experience. Next, the effective support scores from Chapter 4 were considered along with the gain scores from pre- to posttest for the target students. Correlations were examined to determine if higher levels of effective support correlated with improved learning. Binomial logistic regression was used to look at single items at a time to see which specific items had significant pre-posttest gains.

Learning by Group

Solo Classes

A repeated measures ANOVA was run on the solo student data and the effect of the treatment was significant, $F(1, 21) = 34.67$, $p < .001$, showing that student learning occurred as a result of the intervention. With such a small sample size, adding other parameters to the model is problematic as 20 observations per model variable is the absolute minimum recommended (Judd, McClelland, & Ryan, 2009) but in adding gender, game experience, and level of challenge reported by students to the model one at a time revealed no significant fit improvement.

Dyad Classes

A similar process was followed for the dyad student data and the effect of treatment was significant, $F(1, 53) = 65.38$, $p < .001$. Because there are more observations in the dyad group, adding a predictor to the model is a statistically viable option. When gender is added to the model as a predictor, it is significant

F(1, 52) = 13.02, *p* = 0.001. Taking a closer look at marginal means reveals that male students improved less than female students, but had significantly higher pre- and posttest scores (see Figure 5.1).

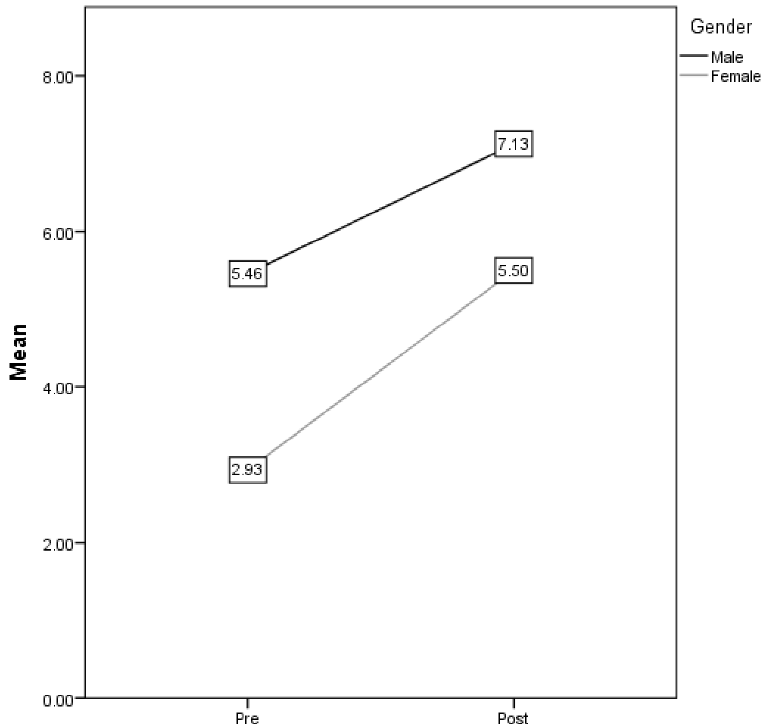


Figure 5.1. Gender differences on the pre- and posttest for dyad classes.

While level of challenge experienced reported by students did not have any statistical significance, using game experience as a predictor did significantly improve the model from the mean-only version, F(7, 46) = 2.56, *p* = .026. An inspection of the frequency counts of students' reports of their approximate time playing games each week reveals a distinct difference between male and female students, meaning that gender and game hours per week (the indication of game experience in this study) are confounding variables and cannot be examined simultaneously in the model as they are correlated (r = .619, *p* < .001). Indeed, most girls reported playing under one hour per week and most boys reported playing 3-6 hours per week (Figure 5.2). Therefore it is unclear whether gender or game experience had more of an effect on outcomes. These results contradict the findings of Clark et al. (2011) in their study on Taiwanese and United States middle school students, where neither gender nor game experience were found to have an effect on learning outcomes. One reason for this difference could be that this population has a much larger proportion of students with less game experience. In Clark et al.'s study, about 12% of the sample reported playing games for under one hour per week--that number is 26% in this study. Additionally, a much higher proportion of those students playing fewer hours are girls (17 girls versus 3 boys) than in Clark's study.

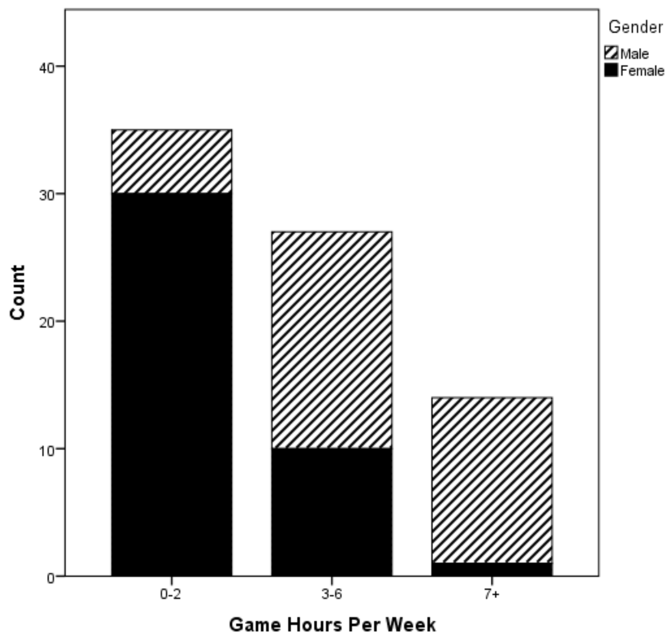


Figure 5.2. A bar graph of game hours per week by gender.

Overall

A repeated measures generalized linear fit on the full data set with group placement as a predictor does not fit significantly better than a mean only model, $F(1, 74) = 1.84$, $p > .05$. Therefore, we fail to reject the null hypothesis that there are differences in learning based on group placement. See Table 5.1 for a summary of the data for solo, dyad, and overall.

Table 5.1

Chart of score gains by group.

	Students (N)	Pre Mean (SD)	Post Mean (SD)	Gain (SD)	Eff. Size (Cohen's D)
Overall (13)[1]	76	4.16 (2.32)	six.54 (2.41)	2.38 (2.09)	1.01*
Solo (13)	22	4.41 (1.87)	7.32 (1.81)	2.91 (2.31)	1.58*
Dyads (13)	54	4.06 (2.49)	six.22 (2.56)	2.15 (1.98)	0.86*

$^*p < .001$; [1]Number of items.

Adding gender to the full population model also proved significant, $F(1, 73) = 12.53$, $p = .001$. Unsurprisingly, an investigation into the marginal means reveals a similar pattern to the dyad classes in the full results (Figure 5.3). Even though gender was not significant in the solo enactment, the larger population of the dyad classes made it significant overall.

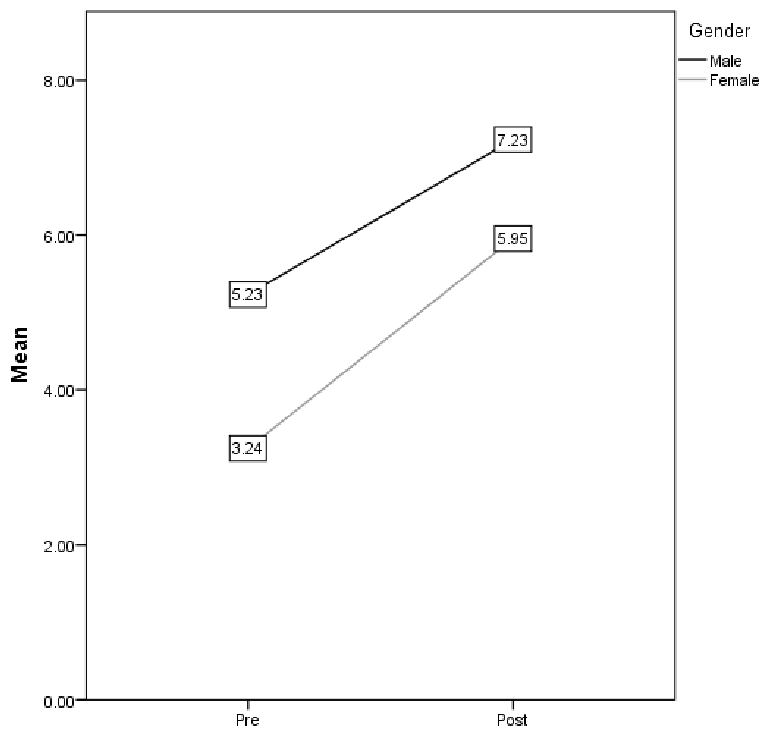


Figure 5.3. Gender differences on the pre and posttest overall.

There was a significant difference in pre- to posttest scores among students with different levels of game experience for the full sample, F(1, 73) = 6.16, *p* = .015, improving the overall fit of the mean only model. Figure 5.4 displays the average pre- to posttest scores for students with low, moderate, and high levels of game experience. It is interesting to note that students with less experience started with lower posttest scores: in fact, Levene's test for equality of variances showed a significant effect on posttest score for experience, F(2, 73) = 5.68, *p* = .005, indicating that students with more gaming experience potentially have better initial intuitions for these specific testing items on impulse, forces, and velocity.

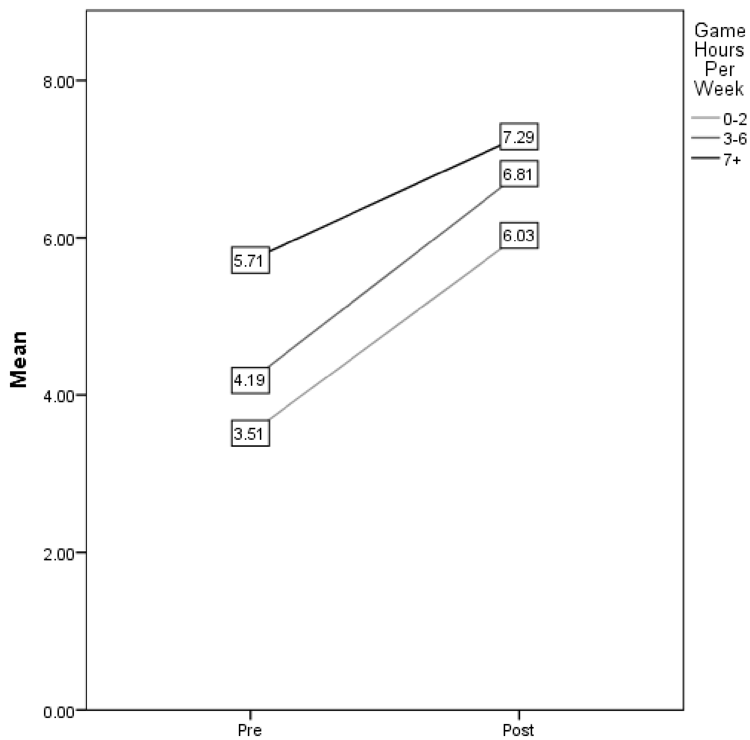


Figure 5.4. Pre- to posttest scores for different levels of gaming experience.

Using a Bonferonni post-hoc test (p < .05), the group with 7+ hours of games per week scored significantly higher than the students in the 0-2 hours per week category. There was no statistically significant difference between the low and moderate groups, or the high and moderate groups. Students with moderate gaming experience ended very near where students with high game experience did, but students with low game experience had just slightly higher posttest scores, on average, than the high game experience players had on the posttest. A pattern emerged in Chapter 4 which pointed to differences in the way students used supports depending on their prior game experience. This statistical analysis confirms a statistical difference in learning outcomes as well. It is prudent to investigate how support effectiveness correlates with learning gains, as game experience has been shown here to have interactions with each.

Comparing Effective Support Score and Pre/Posttest Gains

Statistical analysis was conducted to explore target students' effective support scores as predictors for their gain scores from pre- to posttest. Gain scores and support scores for each target student are presented in Table 5.2. Correlations were calculated for the 18 target students to compare gain scores from pre- to posttest with game support scores. A graphical representation of this relationship can be seen in Figures 5.5. When looking at game support with gain score, the scatter plot indicates a correlation between the two ($r = .63$, $p = .005$) with nearly 40% of the variance in students' gain scores accounted for by effective game support score.

Table 5.2

Game support score, overall support score, and gain score for each target student.

Target Student	Game Support	Overall Support	Pretest Score	Posttest Score	Gain Score
Allen	0.69	0.83	9	11	2
Annie	0.50	0.57	2	3	1
Bentley	0.40	0.75	3	6	3
Carter	0.56	0.50	2	3	1
Danya	0.62	0.85	3	7	4
Damien	0.67	0.79	4	7	3
Elliot (solo)	0.91	0.92	3	10	7
Enzo	0.55	0.56	10	11	1
Greg (solo)	0.83	0.83	7	7	0
Hank	0.83	0.91	6	10	4
Ione	0.46	0.72	2	3	1
Jeremy	0.60	0.60	3	3	0
Lara	0.38	0.64	3	3	0
Lidia	0.80	0.88	3	8	5
Maddie (solo)	0.63	0.65	6	8	2
Nathan	0.50	0.69	5	5	0
Siobhan	0.73	0.78	4	7	3
Wallace (solo)	0.81	0.83	2	6	4

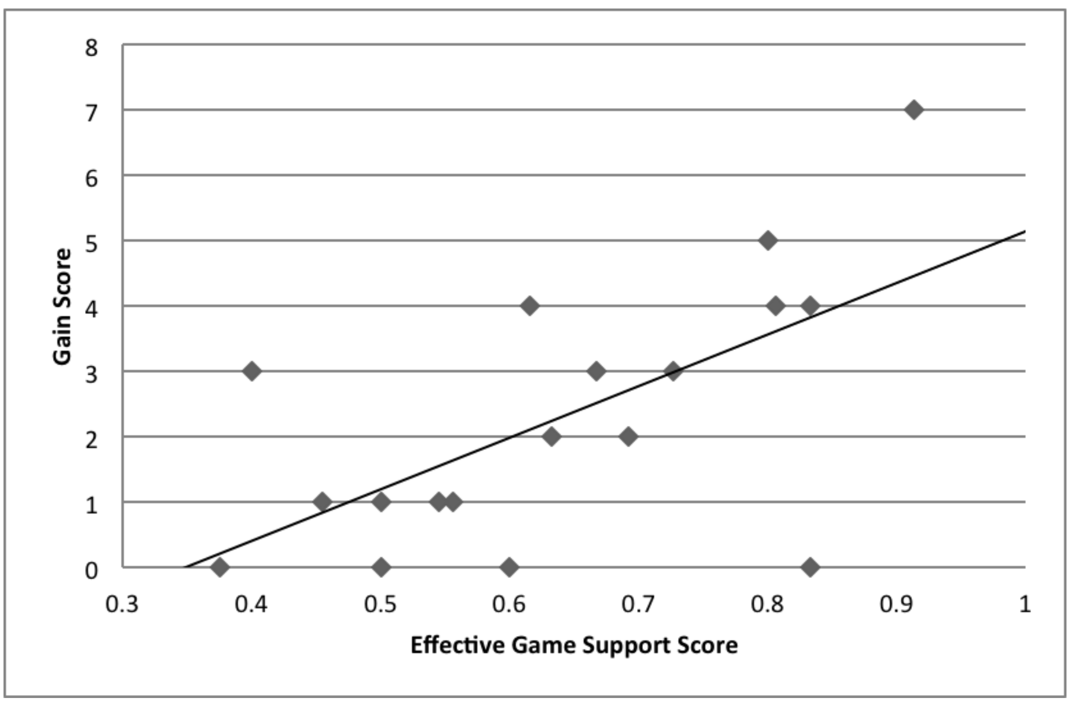


Figure 5.5. Graphical representation of the correlation between target students' effective game support percentage and pre/posttest gain scores.

Statisticians always remind us that correlation is not causation, and that is important to recall while interpreting these results. However, the correlations jibe with the theorized benefits of game supports described in Chapter 2. Prior theory suggests that games are learning machines that teach the learner how to play the game (Gee, 2007). Further, when game mechanics are a simulacra of normative physics processes, it is theorized that learning to play the game will improve physics understanding (Clark et al., 2009). These correlations support both of these theories. This analysis only accounts for supports from the game, so a more well-rounded account of supports comes from the overall effect support score, which includes support from the peers and the teacher.

When looking at the relationship between the overall support score and gain score, a scatter plot indicates a strong correlation between the two (r = .704, p = .001), with about 50% of the variance in student gain scores accounted for by effective support from the game, the teacher, and the peer (Figure 5.6).

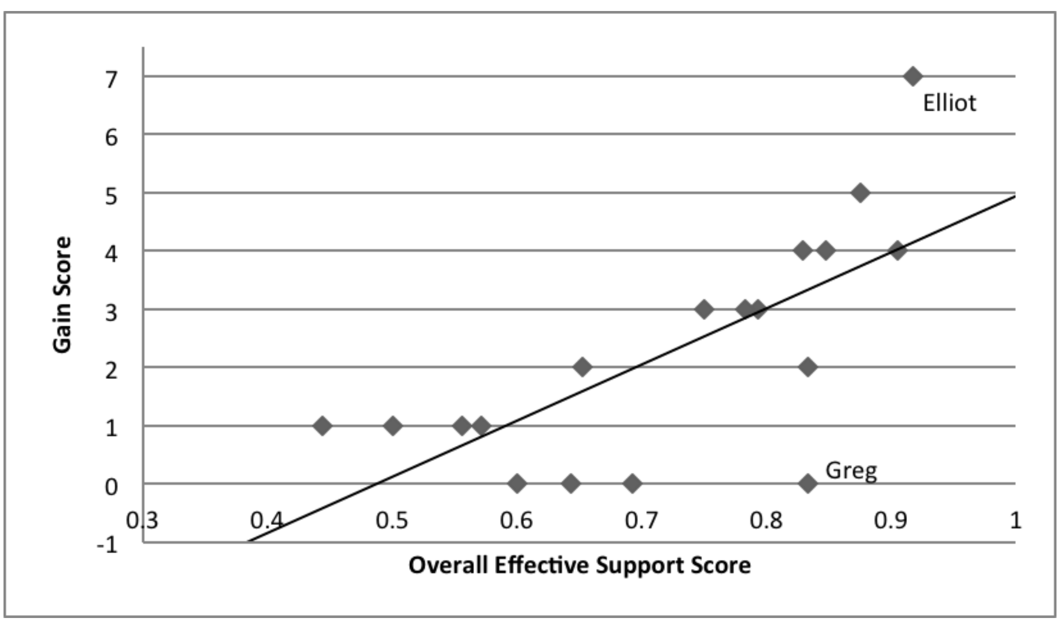


Figure 5.6. Graphical representation of the correlation between the target students' overall effective support and gain score. Potential outliers are marked.

These analyses point to positive correlations between both students' game and overall effective support scores and their gain score from pre- to posttest, which is particularly notable due to the relatively small number of students included in the correlation. This result suggests that the supports from the classroom environment further improved conditions for physics learning while playing a game in the classroom and that the game is more effective with additional supports from the teacher and/or a peer. Other models could potentially fit these data better (e.g., a quadratic) but without sufficient data at the low end of effective support it was difficult to make reliable conclusions about such models.

Figure 5.6 shows two potential outliers. The first potential outlier is Elliot from the solo class, who had a higher gain score than the rest of the sample (7) and also the highest effective support scores for both the game (0.91) and overall (0.92). As mentioned throughout this work, Elliot was the most thoughtful player, having finished first and then running many tests on the game to determine how to get the highest scores. His language was sophisticated in the stimulated recall interview, connecting the game to his play experiences and to new content understanding which he picked up from the game. For example, Elliot was the only target student to spontaneously describe a relationship between the yellow resultant vectors and the red x- and y-velocity vectors:

> Me: What were you thinking about when you were trying this out?
> Elliot: I like, noticed that the yellow arrow was always longer than the two red arrows, but like, if you just go in one direction, the yellow arrow and the red arrow are the same. And then I noticed the speed and that helped me figure it out on the posttest. So, like, if you are going, like, to the right and then add a little bit of speed by pressing up, you'll be going up and right, faster than you're going up or right individually. The yellow is always going to be bigger than the two of them.

Elliot used his extra time not only to run tests but to make inferences that helped him on his posttest. He was the only student to mention playing the game with the posttest in mind during the interview.

Greg is the other student outlier; recall that Greg had the most game experience of anyone in the sample. In Chapter 4, students like Jeremy, Nathan, and Carter were exposed for ignoring game design elements because they did not need them in order to play acceptably well in their partnerships. Greg displayed a solid grasp of the game design elements during play and reported supports from the game were helpful in making connections to other games and physics content. Indeed, Greg had one of the highest effective support scores (.83) but a zero gain score. Like Elliot, Greg finished playing early, but he spent his leftover time obsessively running impulse level six to get as high a score as possible. Unlike Elliot, his approach was not systematic and he was just going for maximum risk to get the highest score, hitting the reset button after any collision so that his score would be perfect if he got through it unscathed. This difference between the two behaviors might account for the discrepancy, but closer inspection also reveals that Greg had a relatively high pretest score (7) and answered in the same way on every question for his posttest. Five of the seven items which Greg got correct on both tests were the five items which showed significant gains when looking at each item individually (see next section). It could be that Greg already knew much of what the game was good at teaching and he did not need the additional information to succeed. Indeed, Greg was the only person in the entire sample that reported that the game did not help him change his thinking while taking the posttest. Elliot may have been able to play well without the additional learning, but he found it fruitful for his experiments and for the posttest performance. Greg just wanted to have a perfect run on level six.

Learning by Item Type and Individual Item

The pre-posttest assessment contained three sections, but that segmenting was invisible to the student. The first section was on impulse and included four items. The second section was about constant force and it also contained four items. The final section covered vectors and it had five items. See Figure 5.7 for a depiction of the pre-post test scores by item for the full sample. Pre- and posttest means are presented for each condition as well as overall, including an effect size calculation (Table 5.3). All gains were statistically significant for each group, and overall, though again no main effect for group was statistically significant.

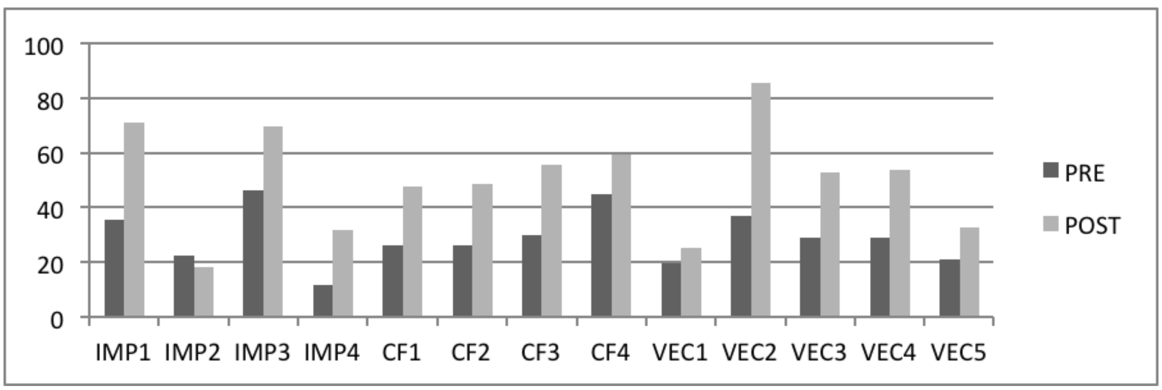


Figure 5.7. Percentage of students answering each question correctly on the pre- and posttest. IMP indicates impulse, CF indicates constant force, and VEC indicates vectors.

Item by item, students improved on all but the second impulse question. As a result, statistically significant student learning occurred across all three sub-sections of the test. For the sake of comparison with prior studies, item analysis was performed on each of the 13 graded test items. Item by item analysis can be accomplished through a McNemar Chi-Square test, but the solo sample was too small for that method to be reliable. A binomial logistic function allows for comparisons between groups to determine whether or not there were group differences for each test item using dichotomous data. Pre- and posttest items were inserted as Boolean codes into a table (0 = wrong, 1 = right) for each student for each item separately. A binomial logistic was run for posttest score and group on posttest score. No test items showed significant differences between groups, but five test items showed significant gains for both solo and group classes. The significance and Chi-Square values for each of these items are presented in Table 5.4.

Table 5.3

Chart of performance on pre and post by for impulse, constant force, and vectors, including effect size and significance of gains.

Test Section	Pretest Mean (SD)	Posttest Mean (SD)	Effect Size (Cohen's D)
Impulse (4)[1]			
Solo	1.36 (1.00)	2.32 (.78)	1.08*
Dyad	1.07 (1.03)	1.78 (1.19)	0.64*
Overall	1.16 (1.02)	1.93 (1.11)	0.72*
Constant Force (4)			
Solo	1.32 (1.17)	2.45 (1.06)	1.01*
Dyad	1.26 (1.09)	2.00 (1.05)	0.69*
Overall	1.28 (1.10)	2.13 (1.06)	0.79*
Vectors (5)			
Solo	1.68 (1.25)	2.55 (1.14)	0.73**
Dyad	1.72 (1.17)	2.46 (1.28)	0.60*
Overall	1.71 (1.19)	2.49 (1.24)	0.64*

[1]Number of items
*$p < .001$
**$p = .016$

Table 5.4

Individual items which had significant gains. There was no main effect for group in any case: all gains reflect the overall population.

Item # (type)	**Test Statistic**	**Significance**
Impulse #1	Wald $\chi^2 = 4.88$[1]	$p = .027$
Impulse #3	Wald $\chi^2 = 11.00$	$p = .001$
Constant Force #3	Wald $\chi^2 = 3.87$	$p = .049$
Vectors #2	Wald $\chi^2 = 19.59$	$p < .001$
Vectors #4	Wald $\chi^2 = 4.10$	$p = .043$

[1] The Wald chi-square statistic has one degree of freedom.

It is important to consider these items in light of other work that has been done on SURGE to add reliability to the findings and create more points of comparison. It is also important to look at the items with significant gains, or not, in the context of the supports and the kinds of thinking the supports brought out in the target students.

Both impulse item #1 and constant force item #3 were found to show significant improvements in the Taiwanese middle school students who were part of a previous study on SURGE, so these results support those findings (Clark et al., 2011). Item one involves students predicting the path of a rightward moving object after receiving a momentary tap upwards. This question, directly from the Force Concept Inventory (Hestenes et al., 1992), is literally represented in the game over and over while controlling the spaceship. In much the same way, constant force item #3 specifically asks about the path of an object after a constant force has been removed. Students experienced this phenomenon repeatedly in the game as well. Examples abound in the corpus of target student data where students describe their surprise when objects continued moving after they let go of the buttons. Recall from Chapter 4 in the initial description of each enactment: students overwhelmingly had an audible moment of discovery, or "Ah-ha" moment, when they realized how the game controls worked. Squire et al. (2004) theorized that students use a game's representations as "tools for action" when modifying their own mental models and the findings in this study further support that theory.

Vector item #2 involves adding two anti-parallel vectors to arrive at a sum. This action is very similar to the students moving their spaceship with a speed in one direction and then "undoing" that speed. A total of 10 of the 18 target students used the word "undo" to describe how they controlled the spaceship in one-dimension. Among the target students who got this item correct on the posttest, Annie said, "I know that I have to undo that sideways vector before pushing down. They have, um, inertia." Ione said,

> If I bounce off a wall after trying to slow down it just makes it worse because the direction I pushed to undo it became the direction it started moving in and it just gets out of control.

Though wordy, what Ione has intuited here is actually quite astute--she is trying to convey that when going too fast, she started to apply an opposite force but then hit the wall anyway. After hitting the wall, that "opposite force" she was initially applying has now become an additive force, providing more speed in the unintended direction. She used the collision mechanic as a tool for reflecting on her assumptions about the way force and motion interact to answer vector item two correctly. Enzo summed it up nicely by making a direct connection to normative physics:

> The thing with most games is with this if you press a button it consistently makes you go in that direction whereas with other games usually you can just tap a button and it will go a tiny bit. If there's no friction then you need an equal but opposite force in the other direction to counteract it.

Vector item #4 proved to be the most interesting question because I was not expecting significant gains at all. The reason for this is because the question aligns vectors in a way that the students do not see in the game. In SURGE, red x- and y-component velocity vectors are always presented at 90 degree angles along the x- and y-axis. In vector item #4, two component vectors are not orthogonal nor on the x- and y-axis, so the representation is rotated and obtuse compared to what students had grown accustomed to in SURGE. A careful look at the target student interviews indicates why this significant improvement may have occurred. Eleven of the eighteen target students described noticing that the yellow resultant vector was "in between" the red component vectors. This phenomenon suggests that students probably do not know how to add vectors exactly (after all, they did not perform significantly better on vector item #3 which asked for an indication of the length of the resultant as compared to the components), but that they have advanced their mental model of vectors to include the idea that the resultant vector resides between the component vectors, directly as a result of the game's representation of the vector cross.

Looking across questions which did not have significant gains is also useful, especially when considered in light of the first pattern described in Chapter 4, namely that scaffolds that are overlaid on top of play or that are not immediately useful for play are ineffective for students. The game includes an indication of the speed of the spaceship in a data window in the lower left hand corner of the screen. While some students did notice this, and indeed one student used it to his advantage (Elliot), most thought it was superfluous. As Allen, a very active gamer, remarked, "I saw it, but I never used it." Several questions on the pre-posttest asked about speed comparisons. As examples, impulse item #2 asked students to describe how an orthogonal impulse changes the speed of an object, and vector item #3 asked students to compare the length of the resultant to the length of isosceles components. These items could be tested in the game, but it would have to be done explicitly; there is never any reason for the player to actually use the speed value to play better. As Enzo explained:

> Maybe if you made like certain things that you had to go a certain velocity in but it wasn't just constant, or decrease speed, but like a number, like 12 meters per second. That would make me focus on the numbers in the bottom.

Enzo's statement fully aligns with the first pattern discovered in Chapter 4. Unless the speed indicator is useful for something, it will not be used for anything, despite the fact that it could have helped students both play the game better, and develop a more sophisticated understanding of velocity vectors and the way the resultant compares to the component vectors.

Summary

In this chapter I presented a wide-array of statistical analyses of the data. I began by looking for any group differences and found that not only did group placement not matter significantly for overall scores, it did not matter by assessment subsection either. A statistically significant effect was found for both gender and game experience with the dyad treatment and the overall enactment. These findings were discussed in light of the second pattern described in Chapter 4, where it was found that different levels of game experience interact with effective use of supports. For this reason, a correlation was attempted and confirmed, indicating that nearly 50% of the variance in student score gains could be attributed to target students' effective use of supports. This finding further validates the second pattern described in Chapter 4.

Item analysis was also explored, finding significant score gains on five of the 13 pre-posttest items. A closer examination of these items revealed that supports from the game and classroom directly aided students in reaching normative answers, but in the case of vector item #4, it is likely that the game nudged their mental model towards a better understanding of the resultant's placement in relationship to the components rather than actually helping them develop a quantitative understanding of the vector components and the resultant. Indeed, more quantitative-like questions on the pre-posttest, especially those questions asking for speed or vector size comparisons, did not show significant gains, and this was explained in the context of the first pattern described in Chapter 4. Unless scaffolds are immediately useful in play, or are provided outside of actual play time, students do not notice them or find them to be effective.

CHAPTER 6
DISCUSSION

The explorations of the research questions under study resulted in descriptions of the ways students navigated different supports while playing a digitally-based game in 8th grade science classes at an independent school. The intention of this work was not to look at supports independent of one another but rather to examine how each interacted with the others to improve student experiences. This work identified a positive correlation between the students' effective uses of supports and their gain score from pre- to posttest. This research also suggested that the timing of supports when students play a game for learning is paramount and that both prior amount and type of game experience have nuanced effects on support use and learning. When working in dyads, non-playing partners were more likely to notice and find effective the just-in-time scaffolding screens as well as the opening screens before each level. The timing of teacher scaffolding also proved important as teaching moments during play were not as effective as the interactions occurring outside of play (while the game is paused or while the partner is playing). Additionally, students with a more physics-based gaming background tended to start with more prior knowledge but they did not necessarily use the scaffolding offered by the game, depending on their level of proficiency with the game.

The Role of Game and Classroom Supports in Promoting Student Learning

One of the main research goals of this study was to uncover a relationship between students' uses of game and classroom supports and their learning in science, as measured by a pre-posttest. The correlations between support and pre- to posttest learning which were found do not indicate causation, but when considered alongside the proposed benefits of games in science education, it does appear that well-designed games act as “learning machines” encouraging players to engage in probing the world, making hypotheses, and re-probing the world with the new hypotheses in mind to arrive at new understandings of the game, and therefore science (Gee, 2007). Prior research on computer games and student learning found that the elements of motivation and fantasy present in games do not enhance student learning in and of themselves, but rather must be bound up in the learning goals of the game and classroom (Cordova & Lepper, 1996). In this study students who were more successful at effectively using supports from the game and the classroom performed better from pre- to posttest because the game was designed to reward physics understanding through play. For example, earning a gold medal required earning higher scores, which were achieved through efficient uses of impulses and forces, which could only be accomplished through an understanding of how to navigate velocity challenge regions and narrow corridors with proper alignment. In other words, endogenous and exogenous supports were bound up in one another. Work done on contextualizing instruction in project-based science (Rivet & Krajcik, 2008) provides another way to interpret this finding: contextualizing learning to be personally meaningful to the student improves the cognition of students who find authentic learning tasks to be personally meaningful to them. With a game, students are motivated not by personally meaningful tasks (though some were reminded of games which they have played in other contexts), but by a desire to, for example, beat impulse level six without collisions, or do well on a tricky level, or earn a better score than a friend: the locus of meaning is displaced, but that meaning is still effective in aiding the learning process. Though the correlations found cannot be directly attributed to the goal-orientation of the game and the endogenously contextualized game features, it does appear that these factors do more than merely motivate students to learn and actually support students in the process of organizing and interpreting their ideas as in a knowledge integration environment utilizing a digital tool (Linn, 2000).

Another goal of this study was to explore the myriad ways students navigate different types of supports in and around a digital game, which resulted in rich descriptions of play episodes and patterns. Students in the solo play condition noticed and used more game supports than teacher supports and, unsurprisingly, very little peer support: it appears that peer and teacher support were more helpful for students with less experience playing video games, but the nature of the partnership was most important in determining the effectiveness of peer support. These empirical findings support the conjecture of other researchers who suggested that a teacher's presence may help students when the game gets confusing or by pointing out things that student might miss (Squire et al., 2004).

Game Experience

Though not initially an explicit focus of this research, game experience emerged as a somewhat complex factor for support usage and learning in students. Though not conducted in a classroom, Heeter et al. (2003) found through empirical research on undergraduates that non-gamers faced disadvantages in game performance, and that, "to the extent that getting the intended impact from a serious game depends on playing well, non-gamers were mostly left behind" (p. 5). Their research also suggested that some non-gamers had more negative experiences with some games, which may be expected to interfere with the proposed learning benefits, though the main conclusion of the work identified player affect towards the game as the most important factor in game effectiveness. The students in this study had highly positive reactions to the game, so much so that differences could not be ascertained between the vast majority of students who had a positive reaction to the game (n=74), and those who had a neutral (n=1) or negative (n=1) reaction to the game. As a result, game experience had no statistical interaction with game enjoyment though the most experienced gamers, who were importantly all male students, started and finished statistically higher on the pre-post test assessment. This result contradicts earlier research on SURGE (Clark, 2011; D'Angelo, 2010) which found that game experience had no interaction with learning outcomes. These contradictory results could point to population differences: for example, many more students in this study had little or no prior-game experience than in prior studies, perhaps as a result of attending an independent school where students have a much higher than average homework load. No matter the cause, prior game experience is certainly a factor which needs more attention in future research not just for affective reasons as suggested by the National Research Council (2011), but for learning reasons as well as suggested by Heeter et al. (2011).

While Chapter 5 confirmed significant correlations between game experience and learning from pre- to posttest, the qualitative analysis in Chapter 4 put a face on those numbers. Qualitative evidence revealed that more game experience also had profound interactions with uses of supports. Students with a lot of game experience may be able to pick up games quickly without ever needing scaffolds from the game to perform at an acceptable level. This eventuality may cause them to miss the scaffolds which explicitly connect game events to science, like the introduction screens, or the just-in-time scaffolding which they may never see because they, for example, never have too many collisions. These findings support the idea that games must be challenging to be optimally useful: these more experienced students may find the game too easy, lacking the flow (Csikszentmihalyi, 1990) necessary for a truly challenging game experience, which interacts substantially with the way supports from the game are used or even noticed. When important content is bound up in and requires "messing up" to activate (such as the just-in-time feedback scaffolding screens which pop up after having two or more collisions), true challenge must be present for all players, even those with appreciably more experience. Alternately, different ways of presenting the information to avid gamers could be devised.

Further nuanced probing of the data revealed that the phenomenon may have more to do with type of game experience than general gaming experience. Students with more physics-based game experience (e.g., marble physics games, trajectory games, momentum and impulse controlled games) needed less help from the teacher and were more successful effectively navigating supports from the game than students who were experienced with word games, shooter games, and other non-physics games. These students may come to the learning environment with more intuitions about the ideas of impulse, forces, and vectors, and the game provides a context for them to attach those intuitions to actual content (Clark et al., 2011).

Another lens through which to view these findings is with a "hot" model of conceptual change (Pintrich et al., 1993). Initially, conceptual change was presented as a process requiring four stages: dissatisfaction with the currently held conception, intelligibility of the newly presented idea, plausibility of the new idea in context of the observed phenomenon, and fruitfulness of the use of the new conception in future circumstances (Strike et al., 1985). Pintrich et al. (1993) added a social cognitive perspective by including motivational beliefs of the student, like their goals and interest in the task at hand.

The vast majority of the sample enjoyed the game and found it appropriately challenging (averaging just about the challenge of a normal science class). With SURGE, game play approaches closely align with intended physics learning, but some opportunities were lost because students did not need certain game supports and features (i.e., the indication of the speed) to be successful in the game, or because students played well enough relatively quickly and did not need the game's help in improving play, though attending to those items may have helped the player gain more science content understanding. SURGE is very good at forcing students to adapt their approaches and increase their play sophistication in order to address tasks of increasing difficulty, but all game design elements must feel "fruitful" to the students in the play or they will go unused or even unnoticed. Unless students are made to feel dissatisfied by their game performance as a direct result of not using or understanding a game feature, it will remain a missed opportunity for learning. Indeed, "[p]rior knowledge plays a paradoxical role in conceptual change" (Strike et al., 1985, p. 191) by either impeding or facilitating change, and when prior game experience is considered prior knowledge, this research supports the idea that knowledge of games, and even certain types of games, changes the way students use supports (or not) to change their intuitions into scientific ideas.

Threats to Validity and Reliability

This work identified a positive correlation between students' effective use of game and classroom supports and pre- to posttest score gains, drawing supporting evidence from the game observation videos and stimulated recall interviews. It is important to consider these findings in the context of this study and the threats to validity inherent in the study's design.

First and foremost, true experimental design was not possible and as such statistical inferences must be considered in light of the fact that samples were not selected completely randomly, though classroom placement in 8th grade science level at the school is not "tracked". Any quasi-experimental design is less desirable from a statistical standpoint because of potential biases or differences between groups. For example, though science is not tracked at this school, math classes are tracked: this could potentially limit the possible class periods available to certain students, giving some classes a higher proportion of lower-math students than others. Statistical tests were incorporated to ensure these differences were not present, but true experimental design is the only way to eradicate these potential issues.

Secondly, the creation of the effective support score relied on coding non-standard data sources. That is, students in solo and dyad groups communicated in different amounts in the game observation videos (e.g., solo students rarely spoke at all), and target students in interviews naturally spoke more or less depending on the personality of the student and the willingness of the student to stop the videotape to comment, or not. Every effort was made to ensure a standard coding procedure between raters and, though high inter-rater reliability was established, decisions about which supports fell under which category were initially subjective in the creation of the coding metrics. The selection of clips for students to view in the interviews was standardized to a point, which may have resulted in missing a clip for which student comment may have changed the researcher's observation--indeed, any clip which was not showed to students relied more upon judgments made by the researcher in initially writing the observations and, as a result, in coding them. Again, carefully described coding procedures and adequate training time improved reliability of the coding.

Third, the small sample size of target students is another concern when considering reliability of the findings. Of course the 18 target students selected could potentially represent the overall 8th grade population at the school, but in such small numbers, differences in individuals are magnified, statistically speaking. Two possible outliers, Greg and Elliot, were discussed specifically; it is impossible to tell whether or not these observations were indeed a result of the circumstances described in Chapter 5, or if the results were actually representative of the correlation between effective support score and pre- to posttest gains.

Fourth, it is important to realize that the prevalence of different kinds of supports from the game is very game-dependent. SURGE in particular contained many features which directly marked critical features and moments of game feedback that refocused the player into a new approach or strategy. Other games may employ more of other types of support, and less of these types. These percentages were determined and assessed vis-à-vis effectiveness to arrive at the patterns described in Chapter 4, namely that scaffolds are best delivered outside of play or when they are directly useful for play in the moment, and that game experience, and type of game experience, have consequences for the usefulness of different kinds of support, both from the game and the classroom. Future researchers should not rely on these numbers and expect their chosen games will exhibit these same percentages: they are specific to this game and were used to describe only these findings.

Fifth, stimulated recall methodology is limited in the fact that, though students report what they thought at the time, their utterances may be affected by time elapsed between play and interview. Additionally, interviewed students all took the posttest before their interview to prevent tainting the posttest with the interview protocol: as a result, students may have had the posttest questions in mind when engaging in the interview. The interviews typically only incorporated 10 minutes of approximately 90 minutes of class time over two days, so I only captured the student thinking that occurred during that small sample of the overall enactment, potentially missing more effective cues to get at student thoughts during play.

Sixth, the framing of this study may have influenced student enjoyment in the positive direction. Indeed, enjoyment could be tied to the way a game is embedded in the overall instructional design. This work employed an instructional design that built from the best practices currently identified in the literature. The scripted introduction to vectors which preceded student play may have had some interaction on the pre- to posttest gains but the mini-lecture was very spare and included no direct connections to assessment items. We know from the literature that students with some modicum of background knowledge are prepared to be more successful with an intervention, so it would be disingenuous and unrealistic to suppose a short introduction to the material would not be part of using this game in the classroom.

Of course, it could be interesting to investigate how the order of the mini-lesson and play may affect outcomes.

Finally, the study occurred with an independent school population of students in Southern Connecticut. Much of the qualitative observations and data was made possible by focusing on only 6 classrooms (the entire 8th grade population) and 18 target students, allowing for detailed analysis of each individual target student and rich descriptions of play and the interactions therein. However, many different factors both within and without the classroom can have effects on student learning. The methodological choices guiding this study did not consider different instructional approaches at other schools, environmental factors at different schools, or the many personal factors individuating students from each other beyond those probed in the study (e.g., gender, game experience). Every effort was made to standardize the enactment of the research within each classroom, utilizing the researcher as the teacher, but students each had different science teachers normally and students may have had different orientations to science as a result of their normal classroom teacher; differences in classrooms as a result of teacher norms were not investigated as part of this work. All results should be considered in light of these threats to validity and reliability, but these findings have implications for those desiring to use games in their classroom, on the creators of games intended for classroom use, and on future research in games and learning.

Informing the Design of Classroom Learning Environments Involving Digital Learning Games

In their work describing the instructional design of classroom environments using the WISE knowledge-integration framework, Linn et al. (2000) delineate a four-pronged approach: make science accessible, make thinking visible, help learners learn from one another, and promote autonomous lifelong learning. These factors were considered when designing the enactment that constituted this research, though further needs emerged when working with a game specifically, as opposed to other digital interventions. When considering the use of a game in the classroom, the instructional design must carefully attend to the patterns described as part of this research. One of the primary design implications arising from the outcomes of this study is the criteria that teachers should only discuss direct connections to science outside of game time, or while the game is paused, as in this study, external support from teachers and peers was more effective when the player was between levels or not playing. The findings imply that Huizinga's (1938) idea of the "magic circle" extends to classroom use of games as well and that the timing of scaffolds is critically important.

Additionally, designing classroom environments centering around games must attend to prior gaming experience, and even the type of game experience. Students with a lot of physics game experience may have more prior intuitions that are normative (e.g., Allen), but others may not think of the connections to physics automatically, though they are otherwise successful in the game. Students like Jeremy and Nick played the game well without finding it necessary to develop more sophisticated understandings of velocity vectors, impulse, and force because they picked the game up very quickly. These kinds of students should be made to explicitly connect physics to the game in a formal way, perhaps even outside the game. Though SURGE-specific scaffolding problems were incorporated into the instructional design, Jeremy did not participate in completing the problems as he played the game while his more-knowledgeable partner Enzo completed the problems alone. Each student must be made responsible for making connections from the game to physics rather than allowing one student to dominate--of course, this is always the danger of group work so further checks and balances are desirable. Perhaps for homework between two days of play, the teacher could ask students to answer simple questions or do problems based on the game play as a formative assessment. Alternately, the teacher might rotate around the room and have students articulate their thoughts during play at a specific moment, such as going around an obtuse turn, or navigating a velocity challenge zone. Here the teacher may notice lack of understanding and remedy the issue so the student might re-contextualize his or her learning in play.

In the case where the student or students have very little gaming background, the teacher must be ready for intensive one-on-one work and even demonstrative training to help the student progress. Instances of frustration were observed across the sample, though only one student, Iris, who was not a target student, found the experience not enjoyable as a result of the frustration. SURGE requires a level of dexterity that might be too much for a small portion of students, and an alternate assignment or perhaps a more stepwise game or simulation might take the "in-the-moment" stress out of playing the game. Other games which do not require such precise mastery of controls may not have such a problem, but most students found the control of SURGE central to its gameness and eliminating that aspect of the game might reduce its effectiveness as a learning game in the way it is received by students.

In much the same way that classroom teachers are encouraged to differentiate instruction for different learning styles and abilities, the findings in this study support a recommendation that teachers try and pair up students with divergent gaming experience, though not too divergent. In other words, students with a lot of game experience should not be paired with students who have no game experience. Teachers should avoid pairings where students with very little applicable game experience work together, and consider that amicability is important to pair effectiveness. In the cases where students work alone, the findings in this study suggest that students with a lot of game experience should be held responsible in some way for connecting game content to physics content, either through paper-and-pencil work or through embedded or meta digital challenges which are provided by the teacher. For example, the teacher might encourage and reward experimentation in the way that Elliot approached his play sessions, looking for further structure in the scoring mechanics to arrive a deeper understandings of the game and the physics integral to the game. Students working alone with very little appropriate gaming experience will need more support from the teacher to progress in a way that does not feel frustrating, allowing the student to experience the desired flow of the game experience (Csikszentmihalyi, 1990).

The nature of the magic circle in the classroom is important to note for instructional design reasons. Placing a game in the sociocultural space of the classroom changes the way players interact with the game. Further exploration could examine students' uses of science content in the game when played outside of class (e.g., for homework) versus inside the classroom. Do students find game supports to be more effective while playing the game in school, where they have other science resources and the expectation of learning?

Recommendations for Educational Game Designers

The findings of this research suggest new considerations for game designers whose games are intended for classroom use. Game designers should avoid incorporating feedback on the play experience which requires the player to divert attention from the control of her or his character in order to use the feedback effectively. When scaffolding provided by the game is embedded in the action it is much more likely to register as fruitful for better play experiences. Additionally, game designers should consider incorporating different levels of challenge in their games allowing for students with more initial success to struggle with more demanding challenges.

It would also be desirable for designers to include classroom use manuals detailing the ways different levels interact with and change play, in order to remove some of that burden from the classroom teacher. Teachers need assurances that each student will feel appropriately challenged (not too little, not too much) but also that students at both the high and low end of play ability will experience all the game has to offer in terms of science content and connections to science processes.

Designers should make explicit in these manuals how all of the content and process features manifest themselves in the game, whether outside the direct play environment (i.e., narrative screens before and after play) or within the game play sessions. Teachers should be able to quickly glance at these content lists and draw students' attention to features/controls they might be missing, in the case of less experienced gamers, or take for granted, in the case of those with more experience. Other research has also suggested these types of manuals for education, but the current study recommends describing these features in a new way, vis-à-vis the timing and placement of supports within and without the game.

Recommendations for Future Research

Unlike in previous studies on SURGE (Clark et al., 2011; D'Angelo, 2010), main effects on learning gains were found to exist by gender differences and game experience. Additionally, it was observed that game experience and gender were highly correlated, with boys playing significantly more videogames outside of school than girls. Perhaps this difference from prior research studies is site-specific or a result of the independent school population, but further investigation of the differences between how very heavy and very light gamers approach playing games (and the types of games they play!) and use supports from the game and from their peers is desirable. Additionally, SURGE is a game which is concept-integrated and requires direct manipulation of velocity vectors in real time to achieve the game's goals. Other types of games should be examined in a similar way to determine if the genre of the game affects the use and effectiveness of game and classroom supports.

This study defined support as anything that improves the conditions for learning or leads someone to learn something directly: future studies might bifurcate the results of support in order to spotlight specific results of support engagement. For example, recruitment supports were not observed to elicit direct connections to physics in students but were rather used by students to improve the conditions for flow in play, reduce frustration, or increase motivation, making further play desirable. Future research might focus on just those moments where students make direct, concrete connections from their play experience to science concepts, though more prolonged engagement or much longer and involved interviews are probably required to collect the critical mass of empirical data to get at those connections.

Before even commencing with this research, finding a game with an appropriately sophisticated theoretical foundation, directly connecting the game design to current theories of learning in science, proved quite challenging. More games with the learning outcomes of SURGE should be both identified and, of course, developed with the same attention to theory twinned with desirable game design elements. Only then can theories be developed and tested across different types of learning games. The National Research Council (2011) recommends that "researchers should establish stronger theoretical underpinnings for the use of simulations and games by connecting research on simulations and games to the relevant theory and research on learning" (p. 122) and educational game designers would do well to attend to this research need as well.

Different dyads worked in different ways in the study, some cooperatively, some competitively within their dyad, and there were even some intragroup competitions. These different dynamics may have had similarly nuanced effects on learning gains and support interaction. Game and instructional designers must recognize the opportunities to get students to think about their play strategies. This study noted strategy switching but did not concentrate on those switches as the core of data collection. Further work should look at how group dynamics interact with the ways game and other supports facilitate students' sharing their ideas and strategic approaches to play.

Another research question was investigated as part of this study in order to examine the way use of supports from the game, peers, and the teacher in a classroom setting interacted with the aesthetic experience of playing a game as defined by Clark et al. (2009) and Gee (2003). Scaffolds from all sources were certainly observed to support the aesthetic experience of the game. For example, refocus supports facilitated the active learning process that Gee (2003, p. 88) poses players go through when engaging with a well-designed game:

> (a) probe the world
> (b) form hypothesis about what something in the game might mean in a usefully situated way
> (c) reprobe the world with that hypothesis in mind to see what happens
> (d) use feedback from the world to accept or rethink the original hypothesis

The analysis in Chapter 4 describes many instances of students arriving at the game with one intuition and then being challenged by the feedback from the game to change their hypotheses to create new intuitions and learning.

According to Clark et al. (2006, p. 26) digital games involve:

> (a) Digital models that allow users to make choices that affect the state of those models
> (b) an overarching set of explicit goals with accompanying systems for measuring progress
> (c) Subjective opportunities for play and engagement

Again, supports from the game facilitated the aesthetic experience of playing the game. Recruitment supports engaged students in a fun way, and recruitment and refocus supports also reinforced the overarching goals of the game while measuring progress.

These results were not presented as part of the overall analysis in the dissertation because the specific population was so overall positive about the gaming experience. While girls and non-gamers did perform less well on the posttest, their improvements were in line with boys and regular gamers, and they did not report a negative experience with the game. As previously mentioned, only one student, Iris, reported a truly negative experience with the game and she was not a target student. Therefore, richer descriptions of her experience could not be delineated and analyzed. So while this specific population nor the populations in Clark et al.'s (2011) work did not show the diversity of attitudes towards the game seen in other studies, that does not mean that the question is not worth investigating in future studies with different populations and, perhaps more importantly, different games. Heeter (2011) has shown qualitatively different effects on engagement and motivation among the vulnerable student populations as described, and such differences of experience should never be discounted when engaging in classroom research where forced gaming becomes the task.

Conclusion

Using games in education, especially with additional classroom supports, presents an array of affordances for both students and educators, but without a coordinated research base directly building off of prior work, the movement will be doomed to the fringe. Science and technology education researchers along with learning sciences researchers must build upon the strengths of each discipline's research base and describe a robust framework for using games in an optimized way for all grade levels, populations, and level of computer access.

This work represents just one single step forward in our understanding of games in science classroom learning. Well-designed educational games like SURGE are very effective at providing scaffolding *in situ* under the right circumstances in both solo and pair conditions. However, when in pairs students take some of the support burden off of the game by demonstrating techniques to each other, outright explaining game mechanics and the related physics to one another, and pointing out features to one another. The teacher can be especially helpful in mediating cases where partnerships are not fluid and one partner needs more support than the other to play effectively. The teacher can also helpful in solo situations by marking features that solo students may have missed and in demonstrating alternate play strategies that they do not have the benefit of seeing demonstrated by a peer. Learning outcomes were found to be correlated positively with more effective uses of support from the game, but also from all three levels of support combined, supporting the theory that games are learning machines (Gee, 2007), but that additional support improves their effectiveness in classrooms.

This work indicates that scaffolding for content learning may work best around games when it is delivered outside of the "magic circle" of play, unless the scaffold is directly salient and manipulable by the student in play. Additionally, all stakeholders should be cognizant of the role of prior game experience in the way scaffolds are both employed and digested by students: avid gamers, especially those who play physics-based or motion-based games, sometimes found the game easy enough that they did not need the added scaffolding in the game, thereby limiting their success on the posttest. The way students play can also affect scaffold accessibility: when in dyads, the non-playing partner notices pop-up scaffolding more readily than the player. Game designers can add multiple levels of difficulty, or include other checks into the game such that gamers of different ability and in varying play situations are necessarily obliged to use the physics information for something useful in play.

The findings in this study are hardly conclusive and require careful delineation: they are nuanced and of course localized to this population. Nevertheless, this work has contributed to the important task of determining how best to use games for learning in the classroom and makes strong cases for specific approaches to their use in schools. More work across different levels, populations, and classroom conditions is needed to create a robust body of literature around creating maximally effective learning environments for digital learning games. Scaling up the use of games in schools without this kind of work could result in the worst possible outcome: effective software collecting dust in the back of a closet, or a wonderfully rich web game going unplayed. But with concerted and coordinated efforts across multidisciplinary fronts, a robust theoretical and practical framework will emerge that will ensure the proper use of games in our schools.

Bibliography

AAPT. (2002). AAPT Statement on Physics First. College Park, MD: AAPT.

Abrams, S. S. (2010). The dynamics of video gaming: Influences affecting game play and learning. In P. Zemliansky & D. Wilcox (Eds.), *Design and Implementation of Educational Games* (pp. 78–92). Hershey, PA: Information Science Reference.

Barab, S, Dodge, T., Jackson, C., & Arici, A. (2003). *Technical report on Quest Atlantis: Volume I*. Bloomington, IN: Center for Research on Learning and Technology.

Barab, S, Thomas, M., Dodge, T., Carteaux, R., & Tuzun, H. (2005). Making learning fun: Quest Atlantis, a game without guns. *Educational Technology Research and Development*, *53*(1), 86–107.

Barab, Sasha, & Dede, C. (2007). Games and Immersive Participatory Simulations for Science Education: An Emerging Type of Curricula. *Journal of Science Education and Technology*, *16*(1), 1–3.

Bransford, J. D., & Schwartz, D. L. (1999). Chapter 3: Rethinking Transfer: A Simple Proposal With Multiple Implications. *Review of Research in Education*, *24*(1), 61–100.

Brown, A. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Science*, *2*(2), 141–178.

Caillois, R., Salen, K., & Zimmerman, E. (2006). The definition of play and The classification of games [1958] (pp. 122–155). Cambridge, MA: The MIT Press.

Chinn, C. A., & Mahotra, B. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*, *86*(2), 175–218.

Clark, D. B., Nelson, B. C., Chang, H.-Y., Martinez-Garza, M., Slack, K., & D'Angelo, C. M. (2011). Exploring Newtonian mechanics in a conceptually-integrated digital game: Comparison of learning and affective outcomes for students in Taiwan and the United States. *Computers & Education*, *57*(3), 2178–2195.

Clark, D., Nelson, B., Sengupta, P., & D'Angelo, C. (2009). *Rethinking science learning through digital games and simulations: Genres, examples, and evidence*. Washington, D.C.: National Research Council.

Collins, A. (1990). Toward a design science of education. New York: Center for Technology in Education.

Collins, A., Brown, J. S., & Holum, A. (1991). Cognitive apprenticeship: making things visible. *American Educator*. American Federation of Teachers.

Cordova, D. I., & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology*, *88*(4), 715–730.

Cox, A. J., Belloni, M., Dancy, M., & Christian, W. (2003). Teaching thermodynamics with Physlets in introductory physics. *Physics Education*, *38*(5), 433–440.

Creswell, J. W. (2007). *Qualitative inquiry & research design: Choosing among five approaches* (2nd ed., p. 393). Thousand Oaks: Sage Publications, Inc.

Cruickshank, D. R. (1980). Classroom games and simulations. *Theory Into Practice*, *19*(1), 75–80.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience* (Harper Per.).

DeBoer, G. (1991). Chapter 10: Process and product in science education. New York: Teachers College Press.

Dempsey, J. V, Haynes, L. L., Lucassen, B. A., & Casey, M. S. (2002). The instructional gaming literature: Implications and 99 sources. *Technical Report 96-1*. College of Education, University of South Alabama.

Dewey, J. (1910). Science as subject matter and as method. *Science & Education*, *4*(4), 121–127.

Dickey, M. D. (2005). Three-dimensional virtual worlds and distance learning: Two case studies of Active Worlds as a medium for distance education. *British Journal of Educational Technology*, *36*(3), 439–451.

Dillenbourg, P., Baker, M., Blaye, A., & Malley, C. O. (1996). The evolution of research on collaborative learning. In E. Spada & P. Reiman (Eds.), *Learning in Humans and Machines: Towards an interdisciplinary learning science* (pp. 189– 211). Oxford: Elsevier.

diSessa, A. A. (2000). *Changing minds: Computers, learning, and literacy*. Cambridge, Mass.: MIT Press.

diSessa, A. A., & Sawyer, R. K. (2002). A history of conceptual change research: Threads and fault lines (pp. 265–281). West Nyack, NY: Cambridge University Press.

diSessa, A. A., & Sherin, B. L. (1998). What changes in conceptual change? *International Journal of Science Education*, *20*(10), 1155–1191.

Duschl, R. A., Schweingruber, H. A., & Shouse, A. W. (Eds.). (2007). *Taking science to school: Learning and teaching science in grades K-8*. Washington, D.C.: The National Academics Press.

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

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

Facer, K., Joiner, R., Stanton, D., Reid, J., Hull, R., & Kirk, D. (2004). Savannah: Mobile gaming and learning? *Journal of Computer-Assisted Learning*, *20*, 399–409.

Facer, Keri. (2003). *Computer games and learning: Why do we think it's worth talking about computer games and learning in the same breath?* London.

Fisch, S. M. (2004). Making educational computer games "educational". *Computer*. Boulder, Colorado.

Flores, S., Kanim, S. E., & Kautz, C. H. (2004). Student use of vectors in introductory mechanics. *American Journal of Physics*, *72*(4), 460–468.

Gass, S. M., & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Mahwah, NJ: Lawrence Erlbaum Associates.

Gee, J. P. (2005). Learning by design: Good video games as learning machines. *E-Learning*, *2*(1), 5–16.

Gee, J. P. (2007). *What video games have to teach us about learning and literacy: Revised and updated edition* (p. 239). New York, NY: Palgrave Macmillan.

Habraken, C. L. (2004). Integrating into chemistry teaching today's student's visuospatial talents and skills, and the teaching of today's chemistry's graphical language. *Journal of Science Education and Technology*, *13*(1), 89–94.

Hammer, J., & Crosbie, W. (2006). Seven tensions between schools and games. San Jose, CA.

Heeter, C., Chu, C., Maniar, A., Winn, B., Mishra, P., Egidio, R., & Portwood-Stacer, L. (2003). Comparing 14 plus 2 forms of fun (and learning and gender issues) in commercial versus educational space exploration digital games. *International Digital Games Research Conference*. Utrecht: University of Utrecht.

Heeter, Carrie, Lee, Y.-H., Magerko, B., & Medler, B. (2011). Impacts of forced serious game play on vulnerable subgroups. *International Journal of Gaming and Computer-Mediated Simulations*, *3*(3), 34–53.

Hestenes, D. (1992). Modeling games in the Newtonian world. *American Journal of Physics*, *60*(8), 732–748.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, *30*(3), 141–158.

Holbert, N. R., & Wilensky, U. (2011). FormulaT Racing: Designing a Game for Kinematics Exploration and Computational Thinking. *Games, Learning, and Society Conferences*. Madison, WI: Morgridge Institute for Research.

Huizinga, J., Salen, K., & Zimmerman, E. (2006). Nature and significance of play as a cultural phenomenon [1938]. Massachusetts Institute of Technology.

Hung, C. Y. (2008). *Making sense of video games: An ethnographic case study on the meaning-making practices of asian adolescents*. Teachers College, Columbia University.

Judd, C. M., McClelland, G. H., & Ryan, C. S. (2009). *Data analysis: A model comparison approach* (2nd ed.). New York: Routledge.

Juul, J. (2003). The game, the player, the world: Looking for a heart of gameness. In M. Copier & J. Raessens (Eds.), *Level-Up: Digital games research conference* (pp. 30–45). Utrecht: University of Utrecht.

Kafai, Y. B. (2001). The educational potential of electronic games: From games-to-teach to games-to-learn. Chicago: Cultural Policy Center, University of Chicago.

Kali, Y., & Linn, M. C. (2007). Technology-enhanced support strategies for inquiry learning. *Handbook of research on educational communications and technology (3rd edition)* (pp. 445–490). Mahwah, NJ: Erlbaum.

Kelly, H. (2005). Games, cookies, and the future of education. *Issues in Science & Technology*, *21*(4), 33–40.

Kim, B., Park, H., & Baek, Y. (2009). Not just fun, but serious strategies: Using meta-cognitive strategies in game-based learning. *Computers & Education*, *52*(4), 800–810.

Kirriemuir, J., & McFarlane, A. (2004). Literature review in games and learning. *Futurelab*.

Klopfer, E., & Yoon, S. (2005). Developing games and simulations for today and tomorrow's tech savvy youth. *Tech Trends*, *49*(3), 33–41.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: University of Cambridge Press.

Lee, J., & Probert, J. (2010). Civilization III and whole-class play in high school social studies. *The Journal of Social Studies Research*, *34*(1), 1–28.

Lee, Y.-F., Guo, Y., & Ho, H.-J. (2008). Explore Effective Use of Computer Simulations for Physics Education. *Journal of Computers in Mahematics and Science Teaching*, *27*(4), 443–466.

Linn, M. C. (2000). Designing the Knowledge Integration Environment. *International Journal of Science Education*, *22*(8), 781–796.

Linn, M. C., Clark, D., & Slotta, J. D. (2003). WISE design for knowledge integration. *Science Education*, *87*(4), 517–538.

Linn, Marcia C., & Hsi, S. (2000). *Computer, teachers, peers: Science learning partners*. Mahwah, NJ: Lawrence Erlbaum Associates.

Lou, Y., Abrami, P. C., & D'Apollonia, S. (2001). Small Group and Individual Learning with Technology: A Meta-Analysis. *Review of Educational Research*, *71*(3), 449–521.

Malone, T. W., Lepper, M. R., Snow, R. E., & Farr, M. J. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. Hillside, NJ: Lawrence Erlbaum Associates.

Marx, G. (2002). Ten trends: Educating children for tomorrow's world. Retrieved from http://www.ncacasi.org/jsi/2002v3i1/ten_trends

Merriam, S. B. (1998). *Qualitative research and case study applications in education*. San Francisco: Jossey-Bass.

National Research Council. (2011). *Learning science through computer games and simulations*. (M. A. Honey & M. L. Hilton, Eds.). Washington, D.C.: The National Academies Press.

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academies Press.

O'Neil, H., Wainess, R., & Baker, E. (2005). Classification of learning outcomes: evidence from the computer games literature. *Curriculum Journal*, *16*(4), 455–474.

Pintrich, P. R., Marx, R. W., & Boyle, R. A. (1993). Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Review of Educational Research*, *63*(2), 167–199.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, *66*, 211–227.

Price, C. B. (2008). The usability of a commercial game physics engine to develop educational materials: An investigation. *Simulation & Gaming*, *39*(3), 319–337.

Reid, A. O. (1992). Computer management strategies for text data. In B. F. Crabtree & W. L. Miller (Eds.), *Doing Qualitative Research, Vol. 3*. London: Sage.

Rieber, L., Luke, N., & Smith, J. (1998). Project KID DESIGNER: Constructivism through play.

Rivet, A. E., & Krajcik, J. (2008). Contextualizing instruction: Leveraging students' prior knowledge and experiences to foster understanding of middle school science. *Journal of Research in Science Teaching*, *45*(1), 79–100.

Salen, K., & Zimmerman, E. (2003). *Rules of Play: Game Design Fundamentals*. Cambridge, MA: MIT Press.

Saloman, G., & Perkins, D. N. (1989). Rocky roads to transfer: Rethinking mechanisms of a neglected phenomenon. *Educational Psychologist*, *24*(2), 113–142.

Songer, N. (2007). Digital resources versus cognitive tools: A discussion of learning science with technology. In S. K. Abell & N. Lederman (Eds.), *Handbook of Research on Science Education* (pp. 471–491). Mahwah, N.J.: Lawrence Erlbaum Associates.

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

Squire, K., Barnett, M., Grant, J. M., & Higginbotham, T. (2004). Electromagnetism Supercharged!: Learning physics with digital simulation games. *Environment*, 513–520.

Squire, K., Giovanetto, L., Devane, B., & Durga, S. (2005). From users to designers: Building a self-organizing game-based learning environment. *Annual Convention of the American Educational Research Association*. Montreal.

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. (2003). Harnessing the power of games in education. *Insight*, *3*, 5–33.

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

Strike, K., Posner, G., West, L., & Pines, L. (1985). A conceptual change view of learning and understanding (pp. 211–231). Orlando, FL: Academic Press.

Sun, C.-T., Wang, D.-Y., & Chan, H.-L. (2011). How digital scaffolds in games direct problem-solving behaviors. *Computers & Education*, *57*(3), 2118–2125.

Swartout, W., & Van Lent, M. (2003). Making a game of system design. *Communications of the ACM*, *46*(7), 32–39.

Van de Akker, J., Fraser, B. J., & Tobin, K. G. (1998). The science curriculum: Between ideals and outcomes. Great Britain: Kluwer Academic Publishers.

Vygotsky, L. S. (1967). Play and its role in the mental development of the child [English Translation]. *Soviet Psychology*, *5*(3), 6–18.

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

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

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, *17*, 89–100.

Appendix A: Introduction and Framing for Students

Hi my name is Mr. Stewart. I taught at Hopkins from 2001-2007 and I still love this school more than I can say, which is why I asked Barbara Riley and your teachers if I could come back and do my dissertation research with the 8th grade science students. I first want to thank you for welcoming me and I wanted to quickly go over exactly what you guys can expect for the next couple of days.

A dissertation is a bit like a giant science experiment. Basically, I am interested in seeing how students use science games in the classroom to learn and what conditions make that learning most effective.

In short, you all will be playing a game developed at Arizona State University and Vanderbilt University called SURGE. The game is relatively simple because it's still in an early stage of development--please do not expect a game like Diner Dash or Super Mario Brothers! It uses some of the ideas you have already learned about force and motion, but also introduces some new material to you that you have not learned before: vectors. Before you play the game, though, I need to get some baseline reading of what you currently know about forces and motion, but also I need to ask some questions about vectors in case some of you have done some independent learning at home or just happen to know about them because your mom or dad is a physics teacher! This is what we call a posttest in research.

At your next class meeting, you will be playing the game alongside some problems that I would like for you to try and solve as you play. You should be aiming for the highest possible score you can get in the game, which means doing some experiments yourself to determine what makes a high score, what kind of things make a score go down, and how you can optimize your play to get the best score possible. This is why we play the game for a couple of days--I want to see how your play improves over time. Please feel free to ask me or your teacher questions at any time.

There will be video cameras placed around the classroom--please ignore them. I will be the only person to look at the video recordings and they will be completely safe and locked away while I am not viewing them. You will be off camera if you elected not to be videotaped as part of the research, and if you ever become uncomfortable with any aspect of the study, please come speak with me.

To pair condition only :

You all will be playing the game in pairs. You will need to decide how you want to split up duties in the game, but I would like for each person in a pair to get the same amount of time at the controls. It's very important to me that I can hear what you are thinking while you work together playing the game. Please speak with your partner as you play, or make suggestions to *your partner playing. Explain your ideas outloud, talk about your strategies for getting higher scores, but make sure that you speak in such a way that it does not bother the group teams.*

I am so excited to see you all playing the game and I look forward to the day when I can come share the results of my study with you. Please feel free to ask me any questions at this time before we begin!

After the pretest:

You've learned about the concept of speed from your teachers--basically how fast you go, or how far you move in a unit of time (like miles per hour or feet per second). A vector is something that has an amount and a direction associated with it--so whereas speed contains one piece of information, namely how fast something moves, a vector has two pieces of information attached to it. Velocity is a vector that is related to speed. Reporting a speed I could say, "I am going 5 meters per second."

[Write on board: Speed = 5 meters / second]

However, if you are reporting your velocity you would be more specific: "I am going 5 meters per second to the east."

[In another column on the board: Velocity = 5 meters / second, east]

Vectors are always represented by arrows in physics--the direction of the vector arrow indicates the direction of the vector, and the length of the vector arrow indicates how fast we are moving (longer arrows mean faster).

[Draw a 5 m/s arrow to the east, and draw another arrow to the west with that is 10 m/s (twice as long)].

The game you are about to play expands on these ideas by building on what you already learned about Newton's Laws. Please feel free to ask me [or your partner] questions as you play the game if you need help. I will be rotate around the room to check on each you periodically.

Appendix B: Stimulated Recall Interview

The stimulated recall interview aims to discover what students were thinking while they played the game (i.e., at a time before they took the posttest).

Adapted from Stimulated Recall Methodology in Second Language Research (Gass & Mackey, 2000, pp. 154-155).

What we're going to do now is watch the selected clips from the video I just took of your playing the game. I am interested in what you were thinking at the time you were playing the game. I can see what you were doing by looking at the video, but I don't know what you were thinking. So what I'd like you to do is tell me what you were thinking, what was in your mind at the time while you were playing the game.

I'm going to put the video camera on the table here and you can pause the video any time that you want. So if you want to tell me something about what you were thinking, you can push pause. If I have a question about what you were thinking, then I will push pause and ask you to talk about that part of the video. I will also ask other specific questions regarding your partner, the teacher, and the work you were doing alongside the game.

[Demonstrate stopping the video and asking a question for them. If the participant stops the video, listen to what he or she says. If you stop the video, ask something general, for example:]

I see you're laughing/looking confused/saying something there, what were you thinking then?

If their response is that they don't remember, do not pursue this because "fishing" for answers that were not immediately provided increases the likelihood that the answer will be based on what the person thinking now or some other memory or perception.

<u>The peer</u>

> Talk to me about what you were working on with your partner at that point?
>
> What did you think about when your partner said that?
>
> I noticed your partner [ ] when you said [ ], what were you thinking when they reacted that way?
>
> What were you trying to show your partner here?

<u>The teacher</u>

> What were you thinking after your teacher said that?
>
> Did the teacher help you understand anything better through that discussion?

The game

What were you thinking here/at this point/right then?

Can you tell me what you were thinking at that point [came to a diagonal motion map region, came to a decelerating motion map region, received a message from the game, beat a level with a silver/bronze medal]?

Why did you make that choice in the game at that point?

Are you having fun here?

Did you feel particularly challenged here?

Were you happy when you earned that (high score, medal)?

What were you thinking here (got a bronze medal or lower score)?

How did playing this game compare to playing a game at home?

Science

I noticed you said (physics/science term) here. What were you thinking about at that point?

I noticed you pointing at/looking at (some science feature in the game) . What were you thinking about at that point?

Did any particular game features stand out to you as science features? [Prompt specific game features if necessary: the arrows, the data in the corner of the screen, the "wake" trail behind Surge in the motion map regions, the screens that pop up when you have too many collisions or take too long, the screens between levels]

Is there anything else you would like to add about what you thought about while playing the game in general?

Appendix C: Introductory Survey

Your participation on this survey is 100% optional. Please listen to the directions of the teachers before filling out this questionnaire.

1. What is your name (first and last name please)?

2. If you play video games, where do you usually play them? If you do not play video games at all you may skip to question 5, but if you have ever played a video game please continue with the whole survey.

3. If you play video games, what do you play them on? Consoles (Wii, Playstation), computers, phones/iPods/iPads, handhelds (Nintendo DS)?

4. If you play video games, how many hours do you play video games a week? If your play change significantly from week to week, please go into more detail. For example, if you are only able to play games in the summer or on breaks, explain that and also the number of hours you play when you can play.

5. What is the last video game you played?

6. What is your favorite video game? What do you like about it?

7. What science classes did you have in middle school, and what science are you taking now, if any?

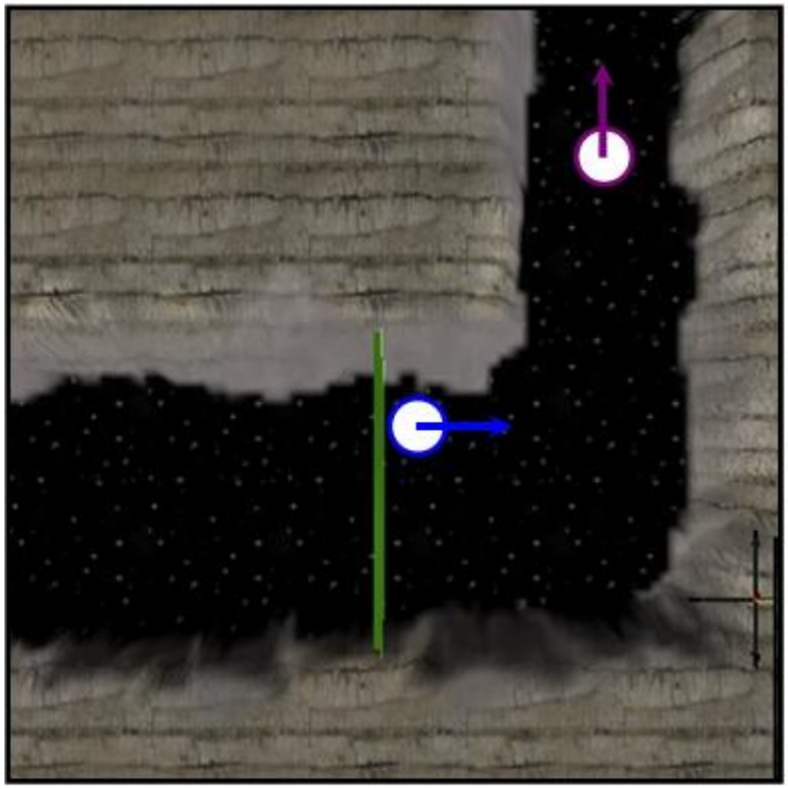
8. Do you like science? Please explain why or why not.

9. Would you be willing to participate in a 30 minutes audio recorded interview with me during one of your free periods after the second day we play the game? If so, what blocks are you free?

Appendix D: Game Data Tracking Sheet

Player				
Level (circle 1)	Impulse 1 2 3 4 5 6			
	Force 1 2 3 4 5 6 7			
	TRIALS			
Time				
Accel. Time				
Impulses				
Collisions				
Total score				

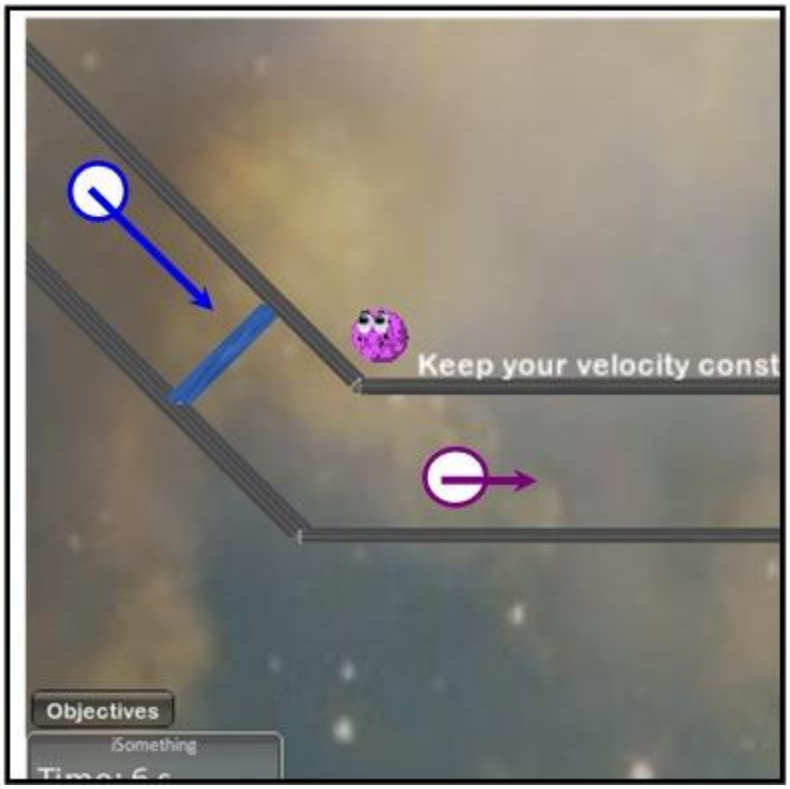
Appendix E: Sample SURGE Scaffolding



Example 1

Which impulses should be added to the blue sphere (velocity shown with blue arrow) to get it to where the purple sphere is (with velocity shown with purple arrow)?

The correct answer is two units to the left and two units up. You need to cancel out the left/right velocity vector as well as increase the up/down velocity vector.



Example 2

Which impulses should be added to the blue sphere (velocity shown with blue arrow) to get it to where the purple sphere is (with velocity shown with purple arrow)?

Appendix F: Play Reflections

Name: ______________________

Homework

How would you describe the rules of SURGE to a friend?

What are some strategies you used today to be successful in the game?

Rate your enjoyment of SURGE in class today compared to a typical classroom activity (circle the number that best describes your enjoyment):

1	2	3	4	5
Bad		Neutral		Great

Why did you rate your enjoyment the way you did?

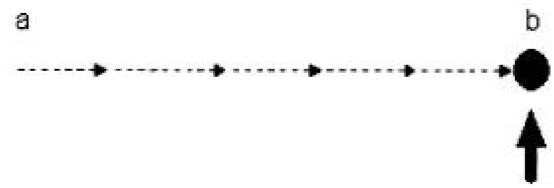
Rate the level of challenge you experienced while playing SURGE in class today compared to a typical classroom activity (circle the number that best describes your level of challenge):

1	2	3	4	5
Very Easy		Neutral		Very Challenging

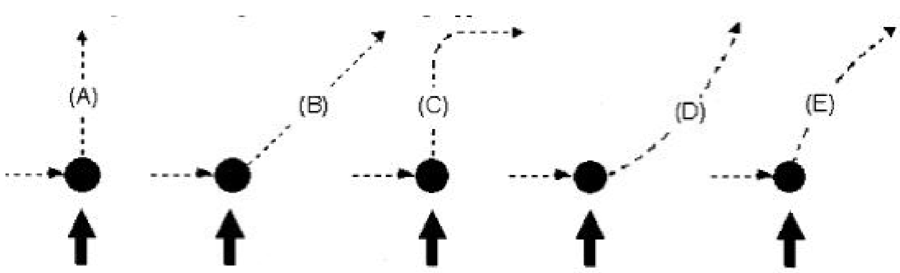
Appendix G: Pre/Post Assessment

USE THE STATEMENT AND FIGURE BELOW TO ANSWER THE NEXT THREE QUESTIONS (1 through 3).

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 **a** to location **b**. 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. {Curly brackets not seen by students}



1. {Impulse #1} Which path does the puck take after being tapped?



(A) Path "A"
(B) Path "B"
(C) Path "C"
(D) Path "D"
(E) Path "E"

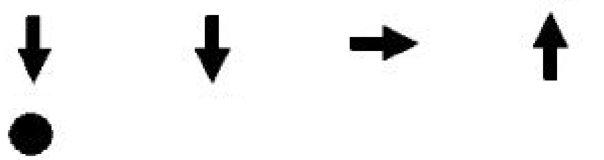
2. {Impulse #2} In the instant just after the puck is tapped, what is its speed?

(A) the <u>same</u> speed as before it got tapped.
(B) the speed given to it by the **tap**; the original speed doesn't matter.
(C) the <u>sum</u> of its original speed and the speed given to it by the tap.
(D) smaller than its original speed, and smaller than the speed given to it by the tap.
(E) greater than its original speed, and greater than the speed given to it by the tap, but <u>less than the sum</u> of these two speeds.

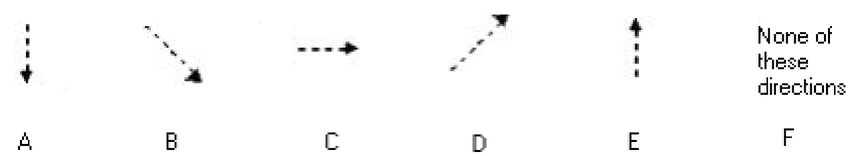
3. {Impulse #3) Look again at your answer to question 2. While the hockey puck is sliding on the smooth ice (no friction) in the rink <u>after</u> it's tapped, how is its speed changing?

(A) It isn't changing; the puck moves at a constant speed.
(B) The puck speeds up.
(C) The puck slows down.
(D) The puck speeds up for a while and then slows down.
(E) The puck moves at a constant speed for a while, and then it slows down.

4. {Impulse #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.



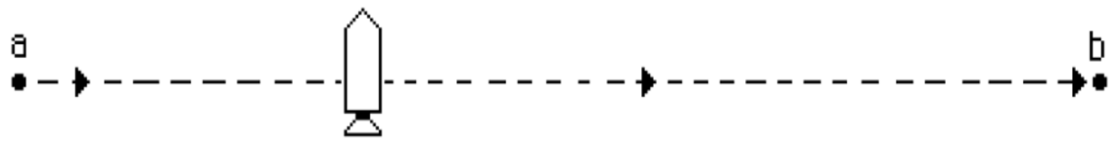
What direction does the puck end up travelling after the four quick hits?



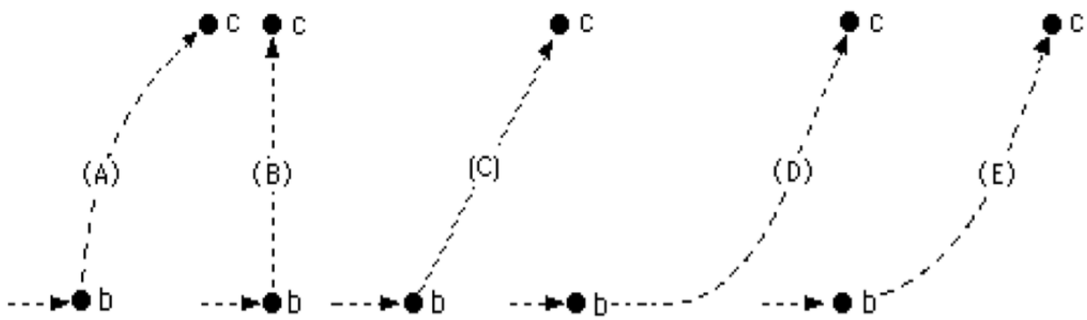
5. [Posttest only] Did anything in the game change how you think about questions 1 - 4 above? Please explain.

USE THE STATEMENT AND FIGURE BELOW TO ANSWER THE NEXT FOUR QUESTIONS (six through 9).

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 **a** to location **b**. No forces act on the ship during this time. At **b**, the captain turns on the ship's engine, producing a force (called a **thrust**) on the ship at a right angle to the line **ab** (toward the top of this page). The thrust stays constant until the ship reaches some location **c**.



6. {Constant Force #1} Which path below does the spaceship take between locations **b** and **c**?

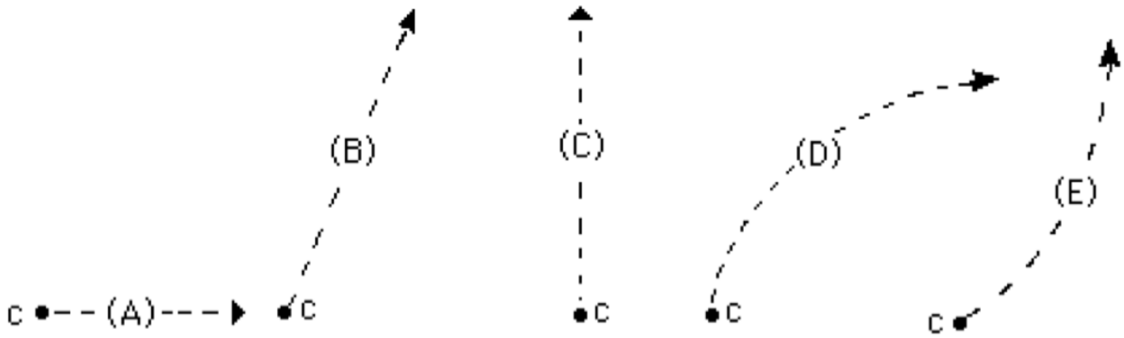


(A) Path "A"
(B) Path "B"
(C) Path "C"
(D) Path "D"
(E) Path "E"

7. {CF #2} As the spaceship moves from location **b** to **c**, its speed:

(A) is constant.
(B) is increasing.
(C) is decreasing.
(D) increases for a while and then stays constant.
(E) is constant for a while and then decreases.

8. {CF #3} At location **c** 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 **c**?

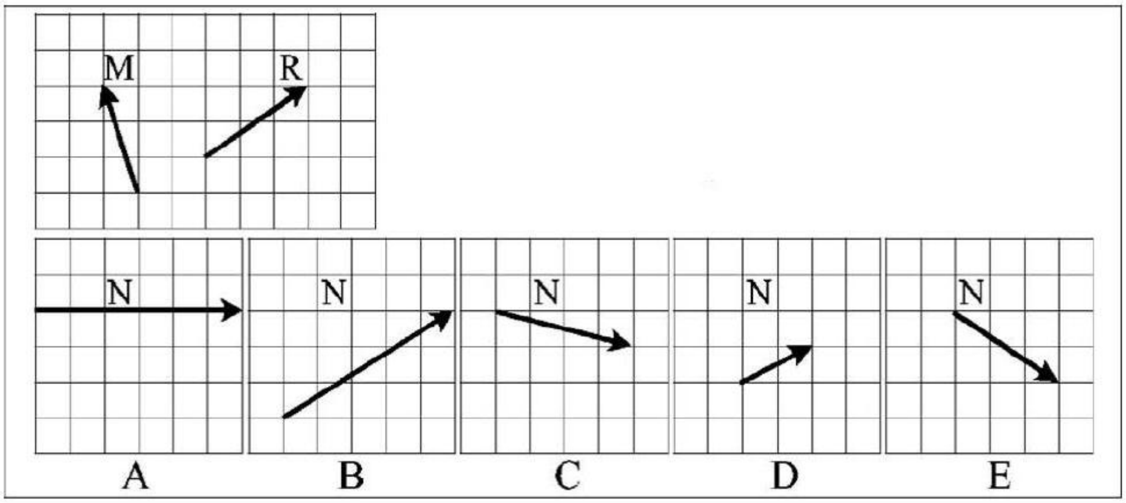


(A) Path "A"
(B) Path "B"
(C) Path "C"
(D) Path "D"
(E) Path "E"

9. {CF #4} Beyond location **c**, the spaceship's speed:
(A) is constant.
(B) is increasing.
(C) is decreasing.
(D) increases for a while and then stays constant.
(E) is constant for a while and then decreases.

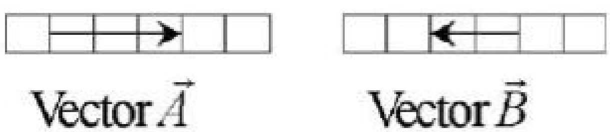
10. [Posttest only] Did anything in the game change how you think about questions six - 9 above? Please explain.

11. {Vectors [Vec] #1] Which is the vector N that when added to M produces R?

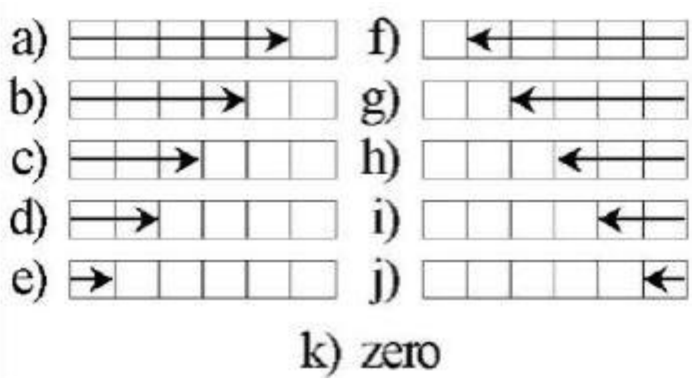


(A) Vector A
(B) Vector B
(C) Vector C
(D) Vector D
(E) Vector E

Use the picture below to answer the next question



12. {Vec #2} Given the two vectors *A* and *B* above, which vector below represents *A* + *B* ?



Use the picture below to answer the next two questions:

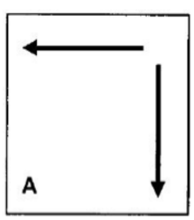
Shown above are two vectors, *A* and *B*, each of length six. The vectors each make a small angle α with the dashed horizontal line.

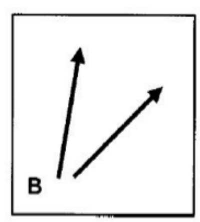
13. {Vec #3} Let *C* = *A* + *B*. Is the magnitude of *C* greater than, less than, or equal to six?

(A) Greater than six
(B) Less than six
(C) Equal to six

14. {Vec #4} What is the direction of *C*?
(A) Up
(B) Down
(C) To the left
(D) To the right
(E) Not enough information to know

15. {Vec #5} In the boxes below are two pairs of vectors, pair A and pair B. (All arrows have the same length.) Consider the magnitude of the resultant (the sum a pair) of each pair of vectors. Is the magnitude of the resultant of pair A larger than, smaller than, or equal to the magnitude of the resultant of pair B?





(A) Resultant of A is larger than resultant of B
(B) Resultant of A is smaller than resultant of B
(C) Resultant of A is equal to resultant of B

16. {On posttest only} Did anything in the game change how you think about questions 11 - 17 above?

Appendix H: Stimulated Recall Data Chunk

{curly brackets}: indicate the cuing event
normal type: indicates the interview content

Begin Chunk

*{Bentley begins playing **Constant Force Level 2**.*

He navigates the first turn carefully by canceling out his upward velocity and coming to a complete stop before pressing right and moving towards the next Fuzzy, which is southeast from him.

Lidia: "It's not so easy is it!?"
Bennet: "I'm going to do this faster than you becuase I'm kind of a risk taker."

Rather than canceling out his rightwards velocity first, he presses the down button and gasps because he starts moving in an unexpected direction (southwest rather than only south).

He desperately presses the left button exclaiming, "Oh shoot I need to slow down first!" just barely avoiding a collision with the wall before getting his velocity solely in the southern direction and safely saving the Fuzzy.

Lidia: "You saw me play it first so you know how the map is!

Bentley: Yeah, I do.

Bentley is now going in one direction at a time, for example coming to a stop in the horizontal direction before attempting to move in the vertical direction. Using this strategy, he attains the gold medal without any more close calls.}

Me: Did watching your partner first help you here?

Bentley: Not really becuase it's hard to learn which tactic to use watching someone else.

(Thinks for a moment).

Bentley: Here I went really fast and hoped I wouldn't hit the walls. Like, the red vectors are the two buttons that I was pushing at the time. They are representing like the engine of a ship with, can I?

(he points to the computer)

Me: Sure.

{He moves the video back to where he almost hit the wall going southeast and pauses it.}

Bentley: If you tap it in different directions it will be a smaller force this way and it will go at a slighter [sic] diagonal angle.

(He uses his hands to mime the ship moving down and to the right and indicates an upward force which would make it move more to the east than south.)

Bentley: And the yellow arrow responds to that force so you want the yellow arrow to point straight down here.

(at the point where he has paused it, canceling out the rightward motion would make the resultant straight down)

Me: What did you think the red lines were representing when you were playing here?

Bentley: They are the force from pushing the arrow buttons.

End Chunk

Appendix I: Segmenting and Coding of an Exemplar Data Chunk

Segment	*Codes*	*Rationale*
Bentley begins playing Constant Force Level 2. *He navigates the first turn carefully by canceling out his upward velocity and coming to a complete stop before pressing right and moving towards the next Fuzzy, which is southeast from him.* *Lidia: "It's not so easy is it!?"* *Bennet: "I'm going to do this faster than you becuase I'm kind of a risk taker."*	[P, R]	Lidia and Bentley share a lighthearted exchange about their different approaches, constituting an instance of **peer recruitment.** Bennett makes a game decision based on challenging his partner.
Rather than canceling out his rightwards velocity first, he presses the down button and gasps because he starts moving in an unexpected direction (southwest rather than only south). *He desperately presses the left button exclaiming, "Oh shoot I need to slow down first!" just barely avoiding a collision with the wall before getting his velocity solely in the southern direction and safely saving the Fuzzy.*	[G, REF]	Bentley changes his course of action due to the **game's direction maintenance (refocus)**, showing him that his current trajectory was not what he had anticipated, so he had to adapt.
Lidia: "You saw me play it first so you know how the map is!" *Bentley: "Yeah, I do."*	[P, D]	Bentley admits to learning the layout of the level from watching his **peer** Lidia **demonstrate** playing the level first.
Bentley is now exclusively going in one direction at a time, for example coming to a stop in the horizontal direction before attempting to move in the vertical direction. Using this strategy, he attains the gold medal without any more close calls.}	[G, REF]	Bentley incorporated feedback from the world, having his attention refocused and has made a change in his play--rather than just stopping in one directions, he is not stopping in all directions before making turns. Bentley **chose** an effective strategy based on his interactions with the game and used it for the remainder of the level because it was effective.
Me: Did watching your partner first help you here? Bentley: Not really becuase it's hard to learn which tactic to use watching someone else. (Thinks for a few moments).	[P, D, NE]	Bentley clarifies his statement from the video indicating that, though he did learn the layout of the level from watching Lidia, his tactical decisions were his own. Therefore, this **peer support** was labeled **not effective**.

Bentley: Here I went really fast and hoped I wouldn't hit the walls. Like, the red vectors were the two buttons that I was pushing at the time. They are representing like the engine of a ship with, can I? Me: Sure. {He moves the video back to where he almost hit the wall going southeast and pauses it.} Bentley: If you tap it in different directions it will be a smaller speed this way and it will go at a slighter [sic] diagonal angle. (He uses his hands to mime the ship moving down and to the right and indicates an upward force which would make it move more to the east than south.)	[G, M]	Bentley used the **game's** representation of x- and y- velocity (**marking critical features**) to make direct decisions about how he should adjust his spaceship's path. Additionally, his descriptions of velocity changes are scientifically normative.
Bentley: And the yellow arrow responds to that force so you want the yellow arrow to point straight down here. (at the point where he has paused it, canceling out the rightward motion would make the resultant straight down)	[G, M]	The yellow arrow in the **game marks the critical feature** of the total velocity, and Bennett articulates an understanding that the yellow line responds to forces applied by the arrow keys.
Me: What did you think the red lines were representing when you were playing here? Bentley: They are the force from pushing the arrow buttons.	[G, M, NE]	The red lines are a **game** support which **mark the critical feature** of x- and y-velocities. Bentley conflates force with velocity and still articulates that he thinks they represent force, hence the **not effective code**.