ABSTRACT

BUFFUM, PHILIP SHERIDAN. Design and Analysis of Virtual Learning Companions for Improving Equitable Collaboration in Game-based Learning. (Under the direction of Kristy Boyer and Tiffany Barnes.)

The social dimension of learning informs much of the current research and development of learning technologies. In some lines of research, the learning tool itself provides the social dimension, often in the form of an intelligent virtual agent. In other lines of research, a learning tool provides an environment in which multiple human learners can collaborate. Integrating these lines of research, this dissertation makes progress toward answering a question that demands increased attention: *How can designers of learning technologies leverage the social capabilities of intelligent virtual agents to support collaborative learning?* Specifically, I examine the effects of virtual learning companions on the discourse of pairs of middle school students collaborating in a game-based learning environment for computer science education, with a focus on equity.

To investigate this line of research, I crafted multiple individual episodes with virtual learning companions, designed for specific locations in the game where students could most benefit from their presence. Through their conversations with the virtual learning companions, I hypothesized that a wider range of students would benefit from productive learning experiences, and that students would collaborate more equitably. These hypotheses emerge out of prior research that has shown the benefits—particularly with respect to equity—that virtual learning companions can have for individual students. To investigate equitable collaboration among pairs of students, this work looks at the dialogue of student pairs as they collaboratively interacted in the learning environment. Following established methods for analyzing discourse, I have annotated transcripts of the student pairs by manually tagging each utterance.

These coded transcripts allow for the investigation of research questions concerning the impact of the virtual learning companions, including how the impact differs based on student characteristics such as gender and prior computing experience. They also lead to insights on the general nature of equitable collaboration within game-based learning experiences for middle school students. The quantitative and qualitative analyses included in this dissertation provide the groundwork for much future research in the fields of virtual agents, game-based learning, computer-supported collaborative learning, learning sciences and computer science education.

Design and Analysis of Virtual Learning Companions for Improving Equitable Collaboration
in Game-based Learning

by
Philip Sheridan Buffum

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Computer Science

Raleigh, North Carolina

2017

APPROVED BY:

_____   _____
Dr. Kristy Elizabeth Boyer   Dr. Tiffany Barnes
Committee Co-Chair   Committee Co-Chair

_____   _____
Dr. Eric N. Wiebe   Dr. James C. Lester

_____
Dr. Arnav Jhala

ProQuest Number: 10758838

ProQuest 10758838

DEDICATION

At age 13, when asked to name and describe a hero, I wrote about my aunt Jenny, a middle school teacher with more strength and grace than I could imagine ever possessing. I dedicate this work to her…

…and to my four grandparents, all of whom have often been in my thoughts as I completed graduate school.

BIOGRAPHY

Philip Sheridan Buffum was born in Providence, Rhode Island. He benefitted from having exemplary teachers throughout his K-12 education, and also from learning alongside great classmates. While he always enjoyed mathematics, his schools never offered any computer science classes. Instead, his teachers spurred on his love of languages and history, ultimately leading him to concentrate in Classics at Brown University. The privilege of studying the humanities there with remarkable professors has influenced how he analyzes problems. As for the specific types of problems that he seeks to analyze, he can credit the privilege of working with inspiring peers and mentors at Brown's Swearer Center for Public Service. Following his graduation in 2008, he joined the National College Advising Corps, serving as a College Advisor for students who would be the first in their families to pursue post-secondary education.

While serving as a College Advisor in an urban high school, Philip became interested in computer-based learning technologies and how they might serve students more equitably and effectively. He subsequently enrolled in his first computer science course, hoping to comprehend more fully the technical challenges that confront designers of learning technologies. Upon discovering the joy of computer science, he subsequently began taking more courses at the University of Rhode Island, eventually leading him to graduate school at North Carolina State University. At NC State, he has continued his focus on equity in education and technology, with a particular emphasis on computer science education.

ACKNOWLEDGMENTS

In studying how to bring more equitable collaboration to middle school classrooms, I find constant motivation in my own past and present experiences. Indeed, I have seen how a person can benefit from collaboration throughout his entire formal education (and beyond).

First of all, this dissertation is the product of much advice and insight from all the members of my committee. I carry immeasurable gratitude to my academic advisor, Kristy Boyer. She has both guided my research with her expertise in the research areas of this dissertation, but also encouraged me to investigate the questions that spark my curiosity and passion. Tiffany Barnes has given me careful feedback on both the computer science education research questions and the game-based learning research questions. James Lester has been a pillar of support throughout my graduate education, with insight on all areas of this research, particularly virtual agents. Eric Wiebe, with his expertise in STEM Education and human factors research, has often encouraged me to address critical questions that might not otherwise occur to me. I am also particularly grateful to Arnav Jhala for serving on my committee, both for his expertise in computational models of narrative and for his outside perspective on my research.

I furthermore owe a great debt to all my collaborators in the Center for Educational Informatics. In particular, Bradford Mott has advised me on both research and technical questions throughout my graduate studies, and his timely assistance has helped make my dissertation study possible. I have also had the great privilege of working on the Engage project for many years with Megan Frankosky, Wookhee Min and Kirby Culbertson, all of whom have

been abundantly generous in sharing their expertise and offering me their feedback. All my lab mates in the LearnDialogue Group have been constant and selfless sources of support. Fernando Rodríguez, Xiaolong Li, Joseph Wiggins, Allison Martínez-Arocho, Alexandria Vail, Jennifer Tsan, Xiaoxi Zheng and Mickey Vellukunnel were particularly instrumental in me finishing the work in this dissertation.

I also feel tremendous gratitude to the faculty and staff of NC State's Department of Computer Science and the Graduate School. Kathy Luca has always granted me more kindness and patience than I deserve. Douglas Reeves and George Rouskas have consistently been supportive and responsive graduate directors. From my first week in graduate school, Sarah Heckman has been a mentor, advisor and role-model. There have been many others at NC State (and other institutions) who have also contributed to my success, and I hope that I have sufficiently expressed my sincere appreciation to them along the way.

I could not have completed the work in this dissertation without, also, the support from many wonderful people in my personal life. First and foremost, I want to thank my wife, parents, siblings and extended family for the abundant love and patience that they have supplied me throughout my life, including my years in graduate school. In addition to my family, I have benefited from the friendship of many wonderful individuals. I dare not try naming them all, lest I leave someone out, but I recognize the following extraordinary people who have supported me as house-mates during my time in graduate school: Jim Witschey, Victoria Bailey, Joel Kerr, Danielle Cohn and Jordan Liew. A dog named Carly also provided me years of comfort and companionship, up through the first couple years of graduate school.

Finally, I humbly thank Alisha Kusmicki and her students at Carnage Middle School for their support in the classroom study reported in this dissertation. I have enjoyed many parts of this whole research process, but none more so than all the time that I got to spend with them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1   INTRODUCTION

Learning is a multi-faceted process that includes cognitive, affective and social dimensions. The social dimension of learning informs much of the current research and development of learning technologies. In some lines of research, the learning tool itself provides the social dimension, often in the form of an intelligent virtual agent (Baylor, 2011). An interaction with such an agent might provide a simulated experience that prepares a trainee for future social interactions with real humans (DeVault et al., 2014), or alternatively it might offer cognitive or affective support to help a student succeed in the immediate learning task (Woolf et al., 2010). In other lines of research, a learning tool provides an environment in which multiple human learners can collaborate (Kafai & Burke, 2013). Just as computer-supported collaboration tools have revolutionized the modern workplace (Dong, Ehrlich, Macy, & Muller, 2016), researchers are endeavoring to understand how educational tools and practices can support groups of two or more individuals as they learn (Radermacher, Walia, & Rummelt, 2012). Although we have learned much from the ongoing research on learning tools that serve as social agents, as well as the research on learning tools that support collaboration between students, little work has been done on integrating the two. (Please see Chapter 2 for a detailed review.) In my dissertation, I aim to make progress toward answering a question that demands increased attention: *How can designers of learning technologies leverage the social capabilities of intelligent virtual agents to support collaborative learning?* Specifically, this document will examine the effects of virtual learning companions on the discourse of pairs of students collaborating in a game-based learning environment for computer science education, with a focus on equity. The recognition of collaboration's important role in education motivates my research.

1.1 Overarching Context: Collaborative Learning

In an ever-more interconnected world, collaboration has emerged as a core twenty-first century skill. While the term "twenty-first century skill" has, at times, threatened to become overly ambiguous due to the variety of groups using it, research efforts to consolidate definitions for

the term note a consensus on the significance of collaboration (Dede, 2010). In the United States, the Partnership for Twenty-First Century Learning (P21) formed in 2002 as a coalition of business, education and government leaders who sought to reach common agreement on the skills needed for success in the new century; P21's resulting framework emphasizes collaboration as one of its four key Learning and Innovation Skills (Partnership for 21st Century Learning, 2015). The Organization for Economic Cooperation and Development (OECD) has taken an international look at the skills needed in the twenty-first century workforce across all OECD countries; it refers to collaboration as part of one of its three dimensions, to which a rich diversity of skills relate. The International Society for Technology in Education includes collaboration as one of seven skills for students, noting that students need to "use digital tools to broaden their perspectives and enrich their learning by collaborating with others and working effectively in teams" (ISTE, 2016). Foundational to all these frameworks is the sense that technological innovation has irrevocably changed the nature of our work and personal lives, leading to a future in which high-skilled personnel increasingly work in teams on cognitively sophisticated projects (Karoly & Panis, 2004).

Of course, we should note that society—and the education field in particular— recognized the importance of collaboration long before the turn of the twenty-first century. Starting in the early twentieth century, social constructivist theorists such as John Dewey (Dewey, 1910) and Lev Vygotsky (1978) began asserting the need to situate an individual student's learning within the context of social interaction. Toward the latter half of the twentieth century, the works of those early theorists enjoyed increasing influence in the general field of education research, as well as in the interdisciplinary research field of artificial intelligence and education. While defining the nature of intelligent tutoring systems, Kurt VanLehn noted how those constructivist views expanded the range of applications for intelligent tutors (VanLehn, 2006). Developing their theory of legitimate peripheral participation, the education researcher Jean Lave and computer scientist Etienne Wenger wrote that "as an aspect of social practice, learning involves the whole person; it implies not only a relation to specific activities, but a relation to social communities – it implies becoming a full participant, a member, a kind of person" (Lave & Wenger, 1991). In short, we can study

collaboration not just for its present-day prominence as a core twenty-first century skill, but also for its established centrality in various theories on how individuals learn.

Yet we still lack a clear understanding of how we can refine digital learning tools to improve collaboration. Efforts are underway to develop assessments of collaborative problem solving, in part by having an individual student interact with a virtual agent who operates as the student's collaborator (*PISA 2015 Draft Collaborative Problem Solving Framework*, 2015). Research on collaborative learning has also uncovered patterns of behaviors that indicate inequitable collaborations within pairs of students (Shah, Lewis, & Caires, 2014). In addition to being able to assess an individual student's collaboration skills or the equity within a pair of students' collaboration, we need to begin studying interventions for improving collaboration. Such a study might entail developing a tool for collaborative learning—or adopting an existing tool—and then adding a new feature specifically designed to foster better collaboration.[1]

1.2 Motivation: Virtual Learning Companions

Compared to supporting individual students, developers of intelligent learning environments confront unique challenges when designing to support pairs or groups of students. In the past decade, researchers such as Carolyn Rosé have begun to explore how to design intelligent tutoring systems for collaborative learning, often in the context of post-secondary education (Kumar & Rosé, 2011). To support groups of young students (e.g. in middle school), however, we will benefit from further inquiry into the potential of *virtual learning companions*, a type of intelligent virtual agent that has been shown to hold particular promise for young learners (Plant, Baylor, Doerr, & Rosenberg-Kima, 2009). In contrast to intelligent tutors, which generally act as experts in the subject domain, learning companions behave as social peers to a student, often providing affective scaffolding. Facilitating such a coequal relationship between student and virtual agent can produce increased learning gains for young students

---

[1] In regard to collaboration, this document will define "better" as "more equitable". To wit, the purpose of the intervention (described in Chapter 7) was to bring greater equity to the paired collaborations.

(Ogan, Finkelstein, Mayfield, et al., 2012) as well as beneficial affective outcomes, such as reduced frustration (Arroyo, Woolf, Cooper, Burleson, & Muldner, 2011). Research has established that learning companions can bring greater equity to intelligent learning environments by supporting diverse learners (Grawemeyer, Johnson, Brosnan, Ashwin, & Benton, 2012; Woolf et al., 2010), although this previous work has largely been done with individual students rather than groups.

For my dissertation, I specifically aimed to study how virtual learning companions might function as the feature for improving equitable collaboration in a system. I have conducted my research within the context of a *narrative-centered learning environment* designed for pairs of middle school students learning computer science. In this case, the narrative-centered learning environment is a game-based learning environment designed around an overarching narrative.[2] Taking advantage of the narrative element, I developed the virtual learning companions to integrate within the story of the game, thereby justifying their presence. I crafted four individual episodes with the virtual learning companions, designed for specific locations in the game where students could most benefit from their presence. The learning companions engage the students in conversations, during which the students have a menu of responses from which to choose. In each episode, the learning companions have three objectives: 1.) encourage students to collaborate productively, 2.) enhance the game's narrative by telling personal stories and 3.) mitigate students' potential frustration by empathizing with the struggles they might be encountering. Through their conversations with the learning companions, I hypothesized that a wider range of students will benefit from productive learning experiences, and a greater degree of equity will be seen in the collaborations of student pairs.

1.3 Research Questions and Hypotheses

At a high level, I propose to investigate how virtual learning companions can induce more equitable collaboration between pairs of students in a narrative-centered learning environment.

---

[2] Chapter 3 provides details on ENGAGE, the narrative-centered learning environment for this research.

To assess equity, I look at the dialogue of student pairs as they collaboratively interact in the learning environment. The data for this study comes in the form of audio and video recordings of students during Spring and Fall of 2016. Following established methods for analyzing discourse, I have annotated transcripts of the student pairs by manually tagging each utterance (Lewis & Shah, 2015; Shaenfield, 2010). I include the specific tagging scheme in the chapter on the study design, Chapter 8. The coded transcripts allow for the investigation of the following research questions (RQs):

- RQ1: To what extent does the addition of virtual learning companions affect collaborations between two middle school computer science learners, with respect to equitable contributions and as assessed through analyses of student dialogue?

- RQ2: To what extent are student characteristics of gender and prior computer science experience associated with the impact of the virtual learning companions?

The first question (RQ1) addresses that overarching purpose of the study, as motivated by the discussion above in Section 1.1. The second question (RQ2) covers complementary inquiries that deserve attention when our aim is to study equity. Specifically, we must look at how other variables—including but not limited to gender—interact with the control variable (*i.e.*, the presence of the virtual learning companions). As discussed in Chapter 6, gender equity is an area of concern, and a wealth of research has demonstrated that virtual learning companions may offer particular benefits for female students.

To address the study's research questions, I have tested several specific hypotheses using the data collected from the two conditions of the experiment.

H1.1 The learning companions will significantly reduce the occurrence of one student dominating the collaboration. The ratios of suggestion utterances to emotive utterances will have greater convergence between student partners in the Companion condition than in the Baseline Condition.

H1.2 The learning companions will increase the extent to which students seek their partner's input. There will be a greater number of questions in the Companion Condition than in the Baseline Condition, and the ratio of questions tagged *Query* to questions tagged *Rhetorical* will also be greater in the Companion Condition.

H1.3 For instances in which one student issues an utterance tagged *Guess*, there will be a greater probability that the partner responds with an *Assert* in the Companion Condition than in the Baseline Condition.

H1.4 For instances in which one student issues an emotive utterance tagged *Negative*, there will be a greater probability that the partner responds immediately with an emotive tagged *Positive* in the Companion Condition than in the Baseline Condition.

H2.1 The impact in the Companion Condition, as tested through hypotheses H1-H4, will be more pronounced with female students than male students.

H2.2 The impact in the Companion Condition, as tested through hypotheses H1-H4, will be more pronounced among students without prior computer science experience than those with experience.

1.4 Contributions

The main study for this dissertation work flows directly from research that I have been pursuing over the past five years. My research to date, which is foundational to my proposed work, includes seven peer-reviewed publications that I wrote as a first author, as well as multiple doctoral consortia and poster presentations. I have investigated how middle school students reason about algorithms for exploring data (Buffum et al., 2014), produced a knowledge assessment for middle school computational thinking (Buffum, Lobene, et al., 2015) and explored the benefits of conducting research in the context of in-school initiatives (Buffum et al., 2016b). Particularly relevant to the research questions for my proposed work, my published contributions to date also include work on studying collaboration and addressing gender equity (Buffum et al., 2016a; Buffum, Frankosky, et al., 2015). Relatedly, my research has added to the literature on how female students can particularly benefit from virtual learning companions (Buffum, Boyer, Wiebe, Mott, & Lester, 2015).

This dissertation builds on these contributions to produce greater insight into how to bring greater equity to collaborative learning experiences. Specifically, it makes the following novel contributions:

C1. *Dialogue corpus of collaborative learning within a game-based learning environment.*
In conducting this research, I created a dialogue corpus that consists of transcripts from

43 sessions of collaborative gameplay. Each transcript is annotated with dialogue act labels. Not including the transcribed dialogue, I collected over 150 hours of additional audio and video ripe for future analysis.

C2. *Dialogue act classification scheme.* I developed a custom dialogue act classification scheme for analyzing equitable collaboration within game-based learning. To develop the scheme, I referenced related work on student dialogue within pair programming (Shah et al., 2014) as well as long-standing work in computational linguistics (Stolcke et al., 2000).

C3. *Dialogue analysis of equity in pair programming, within the context of collaborative game-based learning.* Emerging lines of research are analyzing the dialogue of children collaboratively learning computer science, with a focus on equitable collaboration (Lewis & Shah, 2015). This dissertation adds to that research with the unique perspective of examining children collaboratively learning computer science within a game-based learning environment.

C4. *Empirical analysis of virtual learning companions' impact on collaborative learning.* Emerging lines of research are also starting to investigate how intelligent tutors can support collaborative learning (Kumar & Rosé, 2011). This dissertation makes a novel contribution by examining virtual learning companions supporting the collaborative learning of middle school students.

C5. *Design recommendations for virtual learning companions that support collaborative learning.* The work produces insights into how designers of intelligent learning environments can incorporate virtual learning companions into a system to support collaborative learning activities. It clarifies the challenges that such activities present, in contrast to learning activities in which individual students interact with the computer-based learning environment.

C6. *Replication of analyses relating to gender differences on the impact of virtual learning companions, with novel application to collaborative learning.* Existing research suggests that virtual learning companions might particularly benefit female students (Arroyo et al., 2011). This dissertation study investigates how that hypothesis holds for virtual learning companions who support pairs of students in a collaborative learning

activity. The results suggest that the particularly positive benefits for girls might not, in fact, extend to collaborative learning experiences.

C7. *Empirical analyses relating to prior experience differences on the impact of virtual learning companions.* Existing research has indicated that students' prior computer science experience may impact how well they collaborate (Katira et al., 2004). This dissertation adds to that research, and examines differences in the impact of the virtual learning companions based on students' prior computing experiences.

1.5 Organization

In addition to my work on developing and researching the virtual learning companions, this document describes other research that I have conducted with the ENGAGE[3] narrative-centered learning environment during my graduate studies. For a reader who wishes to quickly understand the dissertation study, I suggest reading Chapters 2 and 3, and then skipping ahead to read Chapters 6 through 11. Chapter 2 contains review of related work on learning companions, narrative-centered learning environments and middle school computer science education. Chapter 3 then provides an overview of the ENGAGE narrative-centered learning environment. The next two chapters, which the reader may wish to skip, describe early studies of the game's impact on computer science attitudes (Chapter 4) and the game's support of collaboration (Chapter 5). Chapter 6 details the first pilot study with the virtual learning companion, a study that laid the groundwork for my proposed research. Chapter 7 then explains how I built on that pilot study with an additional prototype study, ultimately leading to the design of the episodes that I have used for my data collection. Next, Chapter 8 outlines the study design for this dissertation. Chapter 9 details the quantitative analyses of the dissertation

---

[3] Regarding ENGAGE, I was part of a large team of students, staff, and faculty on the NSF-funded project. While the research contributions of this document are mine alone, I note that they build upon research and development contributions by many other individuals. Chapter 3 describes the game-based learning environment, which was built primarily by researchers other than myself. Chapters 4 and 5 describe classroom studies that were conducted as part of the broader ENGAGE project and were facilitated by many researchers (although the analyses of the data in those chapters are mine).

hypotheses, while Chapter 10 then provides further support in the form of qualitative analysis. Chapter 11 concludes with a summary and remarks on potential future work.

Chapter 2   BACKGROUND AND RELATED WORK

The research for this dissertation primarily relates to the field of virtual learning companions, yet the reader can also benefit from some research context on narrative-centered learning environments and middle school computer science education. As such, after an overview on the relevant literature on virtual learning companions, this chapter will then provide brief reviews on narrative-centered learning (Section 2.2) and middle school computer science education (Section 2.3).

2.1 Virtual Learning Companions

The idea of using virtual agents as an educational tool (*i.e.,* as a pedagogical agent) has become an ever-increasing focus of computer science research over the past two decades. While virtual agents have been employed as tutors "from the earliest days of computers" (VanLehn, 2011), Chou et al. describes how the field of intelligent tutoring systems (ITS) really opened up in the mid 1980s, ushering in a broader understanding of what a virtual agent might contribute to student learning (Chou, Chan, & Lin, 2003). Researchers began looking at how virtual agents could not only facilitate knowledge acquisition, but also motivate students to take advantage of the opportunity to gain knowledge. As Chou et al. detail, learning companion systems (LCS) became a popular new research subfield.

Chou defines a learning companion as a virtual agent whose purpose is to act as a peer for the human student; this open definition hints at the breadth of roles in which a learning companion can serve. Two other categories of pedagogical agent carry definitions that constrain them more narrowly to certain relationships with the student. *Intelligent tutors* generally operate as an authority figure; such an agent can take the form of knowledge expert, motivator, or mentor, all of which have demonstrated benefits (Baylor & Kim, 2003). *Teachable agents*, conversely, are designed to appear less knowledgeable than the student, requiring the student to learn for the sake of "teaching" the agent (Biswas, Leelawong, Schwartz, & Vye, 2005; Chase, Chin, Oppezzo, & Schwartz, 2009). The definitions for intelligent tutors and teachable agents place these two types of pedagogical agent in stark contrast to each other; learning companions, on the other hand, may share some overlap with

both. As such, my research on learning companions will consider the relevant literature on all three forms of pedagogical agents.

2.1.1 Affective Impact of Learning Companions

A key concept for understanding animated pedagogical agents' motivational success is that of the 'persona effect' (Lester et al., 1997). At a time when animated agent technology was significantly less advanced than it is today, Lester et al. found that lifelike pedagogical agents have a significantly positive effect on students. The researchers investigated how middle school students react to the presence of an animated agent within the Design-a-Plant interactive learning environment. Along with promising learning gains, it was found that the presence of the animated agent had a significant positive impact on students' perception of their learning experience. Lester termed this the 'persona effect', and concluded that it can motivate students to continue to use the interactive learning environment for longer periods of time. I note that some research has raised questions about whether the persona effect extends to older learners (such as college students) (Miksatko, Kipp, & Kipp, 2010), but I can find no research contradicting the evidence of the affective benefits for middle school students, and a meta-review has confirmed the particularly strong benefits of pedagogical agents for K-12 students (Schroeder, Adesope, & Gilbert, 2013).

In a more recent study with the Wayang Outpost intelligent tutoring system, Woolf et al. found that animated learning companions can have an especially positive impact on the motivation of low-performing students, including those with learning disabilities (Woolf et al., 2010). Although this study did not directly show that students with a learning companion achieved higher learning gains, it found that those students engaged more often in behavior that has been shown to lead to learning gains. The key was that the learning companions demonstrated a significant impact on student affect. In fact, Woolf goes so far as to say that "learning companions are essential for low-achieving students' affect." Specifically, the study indicated that low-achieving students who interacted with a learning companion came away with appreciably higher confidence compared to similar students without a learning companion.

In both the Design-a-Plant and Wayang Outpost studies, the learning companions operated as advisors and mentors to the human students; as mentioned above, learning

companions are not restricted to such traditional relationships with students. Ramirez Uresti and du Boulay provide a concise overview of the different roles that a learning companion might play in service of motivating students (Ramirez Uresti & Du Boulay, 2004). For instance, such an agent might behave as a role model, a collaborator, a competitor, an advisor, or a pupil of the student user. Ramirez Uresti and du Boulay took particular interest in researching that last role: learning companion as student of the human student. They explored this in their LEarning COmpanion system for binary Boolean Algebra (LECOBA), a text-based interactive learning environment consisting of three agents: a tutor, a learning companion and the human student. Building upon the related work of Hietala and Niemirepo (Hietala & Niemirepo, 1998), they tested how students' learning is affected by interacting with "weak" (less knowledgeable) learning companions, as contrasted to interacting with "strong" (more knowledgeable) learning companions. They also looked at the motivation of students to interact with these two different learning companions. In the short studies that they conducted, they found some intriguing trends in learning activities (confirming their hypothesis that weaker learning companions are the most beneficial for learning), although no significant differential learning gains.

Working with LECOBA around the turn of the millennium, Ramirez Uresti and du Boulay were building upon previous research that integrated the Learning by Teaching paradigm within LCSs. They cite several studies from the 1980s that showed students learning more (and better) when they are teaching other students. They also discuss some prominent examples from the 1990s of LCSs in which the learning companion acted as a pupil of the human student (Aimeur & Frasson, 1996; Goodman, Soller, Linton, & Gaimari, 1998; Hietala & Niemirepo, 1998). Those early systems had mixed results, but Ramirez Uresti and du Boulay's own research (showing weaker learning companions as the most beneficial for motivating students to learn) gave them hope that this research would bear fruit eventually. They attributed the negative results of earlier research to the designs of the respective systems. One contemporary system that they held in high esteem was that of the Teachable Agent group at Vanderbilt University.

The Vanderbilt researchers—a union of computer scientists and education psychologists—began in the 1990s to conduct the foundational research for what would later

evolve into teachable agents (Brophy, Biswas, Katzlberger, Bransford, & Schwartz, 1999). Their idea was to combine the beneficial learning activities of teaching (as seen in numerous human-human studies) and of programming (*e.g.*, Seymour Papert's pioneering research with Mindstorms (Papert, 1980)). They came upon this idea following their classroom research that demonstrated the importance of providing students with opportunities to develop and assess their own knowledge and to interact with their fellow learners. The Vanderbilt researchers decided that their virtual humans would provide a felicitous way to combine the benefits of teaching-to-learn with the benefits of programming-to-learn. In an initial study with a cartoon character named Billy (not a teachable agent, but merely a scripted prototype), students help the character perform in an environment that demanded knowledge of ecosystems and water quality. The students in the study showed impressive perseverance, and embraced the embedded assessments as useful opportunities to appraise their teaching skills rather than as tests of their knowledge (Brophy et al., 1999). The ensuing research into how to create the ideal teachable agent resulted in the first version of Betty's Brain, in which the agent shows her qualitative reasoning through a concept map created by the human student (Biswas et al., 2005). The concept map represented the teachable agent's knowledge rather than the referent domain (in this case, river ecosystems), as its purpose was to help the human student structure his or her thinking about the referent domain.

Biswas et al. noted three critical benefits of their system: providing a means for students to develop structured networks of knowledge, helping students develop metacognitive skills, and motivating students (Biswas et al., 2005). That third aspect – regarding the motivational benefits – holds particular relevance to my research and thus merits a closer look.

The Protégé Effect: Chase et al., 2009

Chase et al. (Chase et al., 2009) coined the term 'protégé effect' to describe the psychological benefits of teachable agents. One can think of a teachable agent as a hybrid in that it combines the properties of both agents and avatars (the two main classes of virtual characters in educational technologies). As discussed above, embodied pedagogical agents have been shown to produce the 'persona effect' (Lester et al., 1997), which has great motivational potential. Avatars, meanwhile, have been demonstrated to produce what Yee and Bailenson (Yee & Bailenson, 2007) term the 'proteus effect', referring to how users (in our case, students)

subconsciously take on the attributes of their avatars. For instance, if a user has a large, muscle-bound avatar, he or she might act more assertively in a virtual environment than if the user has a small and frail avatar. Chase posits that avatars are beneficial for learning because of this 'proteus effect', and also because they can motivate students to take risks (since any negative repercussions would not affect the student directly). As a hybrid agent/avatar, teachable agents can act autonomously from the human user, while also reflecting the interactions that it has had with that human user.

In an experimental study to investigate this 'protégé effect' in Betty's Brain, the researchers encouraged one set of students to think of the character as a teachable agent, while the other set thought of the character as strictly a personal avatar (Chase et al., 2009). The researchers assessed the two conditions by examining the computer-generated logging data, the quality of the resulting concept maps, and by a post-test that contained three levels of questions. Students in the teachable agent condition spent their time more wisely, producing a greater effort towards real learning (as opposed to gaming the system). The post-test showed especially good results for integration and application questions (which involve deeper understanding) and especially strong results for low-achieving students (similar to other studies of learning companions (Woolf et al., 2010)). Furthermore, the motivational differences between the two conditions were not due to the teachable agent students having "more fun"; indeed, the logging data indicate that students were spurning the more game-like activities for pure learning activities, such as reading. The researchers then conducted think-alouds to determine if students treat teachable agents as independent, sentient beings, and how students respond to failure when they have a teachable agent as mediator. The resulting corpus of statements was then divided into three major categories: mental attributions ("I don't understand" vs. "He knows it!", etc.), responsibility attributions ("Yeah! I did it!" vs. "Thanks a lot, Queenworld", etc.) and affective statements ("Cool!" vs. "I'm not a good teacher", etc.). Conforming to the researchers' hypothesis, students in the teachable agent condition were more likely to assign responsibility for a failure. Chase attributes this to the human students knowing that, while they are responsible for their teachable agents' performance, they do not have to accept all the blame for a given failure.

Crucially, students did not view their teachable agents as pure reflections of themselves (*i.e.*, as avatars), nor did they consider it a programming assignment – they saw the teachable agents as actual protégés. Chase et al. attribute the 'protégé effect' to three motivational factors: an ego-protective buffer (which encourages students to take risks), the adoption of a growth mindset (Dweck, 2006) and the fostering of a sense of responsibility within students. The findings motivate additional work on investigating how virtual learning companions can leverage these benefits to encourage productive learning behaviors. In particular, Chase et al. note that the short (single session) nature of their studies leaves unanswered the question of how to extend the motivational benefits of the "protégé effect" to a long-term learning environment. The researchers suggest that one way to sustain the motivational benefits long-term would be to integrate a teachable agent within a narrative-centered game (Chase et al., 2009). I am proposing to investigate this vein of research by integrating learning companions into the ENGAGE narrative-centered learning environment.

2.1.2 Recent Work on Learning Companions

Some recent work has started to investigate learning companions in long-term learning environments, producing evidence that a learning companion's effect on student motivation may increase with recurring interactions (Sjödén & Gulz, 2015). Indeed, the lengthier exposure to the learning companion can open up a more varied set of interactions between student and virtual agent, creating a richer, more motivating overall learning experience (Pareto, Haake, Lindström, Sjödén, & Gulz, 2012). In a four-week study of students interacting with a conversational agent to help with developing an online portfolio, students appreciated the social function of the agents, although they did not find the conversational agents helpful with the task itself (Doering, Veletsianos, & Yerasimou, 2008). For the most recent version of Betty's Brain, meanwhile, an open question remains regarding how to ensure that productive learning behaviors persist from one day to the following days over a period of several weeks (Segedy, Kinnebrew, & Biswas, 2015).

Much of the recent research on learning companions, however, has continued to produce important findings in shorter (*e.g.*, single session) studies. Such research has produced evidence that students as young as pre-school can benefit from teachable agents (Axelsson, Anderberg, & Haake, 2013) and conversational agents (Tewari & Canny, 2014). Other

research has sought to develop models for determining the behaviors of a learning companion. For instance, a recent study has investigated a model for adjusting an intelligent tutor's social role and relationship to the student so as to improve the student's learning experience (Pecune, Cafaro, Ochs, & Pelachaud, 2016). In a similar vein, other researchers have developed a model that analyzes the learning trace of autistic children to inform an intelligent tutor's affective feedback (Mondragon, Nkambou, & Poirier, 2016).

Another vein of research has focused on conversational agents as learning companions, with research suggesting that children interact similarly with animated characters as they do with other people in one-on-one conversations (Hyde, Kiesler, Hodgins, & Carter, 2014). Researchers in Sweden have added a social conversation feature to their teachable agent system, leading to more positive student experiences in terms of both engagement and learning (Gulz, Haake, & Silvervarg, 2011). The off-task natural language dialogue in that system supplements the on-task learning activities. In contrast, researchers at Carnegie Mellon have designed their teachable agent system, SimStudent, with on-task, natural language dialogue as the central mechanism for how students "teach" their virtual agent (Carlson, Keiser, Matsuda, Koedinger, & Rosé, 2012). In SimStudent, the teachable agent uses dialogue to get students to explain each step in solving math problems, although research suggests that adding more social dialogue could lead students to feel more socially aligned with the agent and thus more likely to reap the potential learning benefits (Ogan, Finkelstein, Mayfield, et al., 2012).

Finally, a particularly robust area of research on learning companions has explored how various characteristics of a virtual agent, such as its gender (Arroyo, Woolf, Royer, & Tai, 2009) and ethnicity (Iacobelli & Cassell, 2007), impact student learning outcomes.. More recent work has begun to explore furthermore the impact of a virtual learning companion's dialect (Finkelstein, Yarzebinski, Vaughn, & Ogan, 2013; Yarzebinski, Ogan, Rodrigo, & Matsuda, 2015), realism (Hyde et al., 2014) and personality (Bian et al., 2016). Research has also revealed that certain factors also specifically impact student motivation, including a virtual learning companion's role in the learning environment, its realism and its perceived level of expertise (Baylor, 2011; Y. Kim & Baylor, 2006a).

2.1.3 Learning Companions and Gender

As detailed in Section 1.2, one of the three research questions for this study concerns gender; it thus bears particular relevance that research on learning companions has featured gender as a focal point. As mentioned briefly above, some of the research on gender and learning companions concerns the gender of the learning companion itself. The gender of the learning companion has been found to have an impact on students' attitudes and learning (Arroyo et al., 2009), as well as the relative likelihood of the agent receiving counter-productive, abusive behavior from the student (Silvervarg, Raukola, Haake, & Gulz, 2012). For example, after interacting with a female pedagogical agent for engineering education, middle school students showed increased interest, utility and self-efficacy, as well as a decrease in stereotyping among the male students (Plant et al., 2009). As with studies of other agent characteristics such as those listed in the previous section, this research informs what kind of learning companion we incorporate into a digital learning environment to achieve a desired outcome.

Beyond the characteristics of the virtual agent, however, research on gender and learning companions has revealed that these pedagogical agents have especially positive affective results for female students. For example, learning companions have been shown to improve female students' confidence in mathematics and to reduce the gender gap in frustration in that field (Arroyo et al., 2011). They also have the potential to improve self-efficacy among female students (Y. Kim, Wei, Xu, Ko, & Ilieva, 2007). Female middle school students respond more positively to pedagogical agents that include affective support than to ones that only offer task-oriented support (Burleson & Picard, 2007). Relatedly, female students also show more openness to seeking and receiving help and spend more time reading a learning companion's messages (Arroyo, Burleson, Tai, Muldner, & Woolf, 2013). In short, there is strong evidence in support of tailoring the behavior of such agents based on the gender (as well as achievement level) of the learner (Arroyo et al., 2013; Burleson & Picard, 2007; Y. Kim & Baylor, 2006b).

On a related note, in addition to their gender-specific impact, learning companions have also been found to have especially great impacts for low-achieving students (Rader, Echelbarger, & Cassell, 2011; Woolf et al., 2010). As one study with the prominent intelligent tutoring system AutoTutor discovered, however, the demographic information of a student is

not the only important factor to consider: context is also hugely important. Even though the affective pedagogical agent was generally more effective with low-domain knowledge students than with high-domain knowledge students, that effectiveness relied on the timing of the affective support, with immediate or preemptive affective support proving detrimental to learning (D'Mello et al., 2010).

2.1.4 Learning Companions and Collaborative Learning

Given the research questions for this study (particularly RQ1), we also must take note of literature on learning companions that support collaboration between multiple human students. Research has demonstrated that socially capable virtual tutors can also support groups of students in collaborative learning situations (Kumar, Ai, Beuth, & Rosé, 2010). A conversational agent has been used to encourage students to engage in a transactive form of dialogue, in which students collaboratively build on each other's reasoning about the given domain (Tegos, Demetriadis, & Tsiatsos, 2016). A study of paired gameplay has demonstrated how a teachable agent can add beneficial forms of both collaboration and competition, producing a dynamic learning experience that leads to improved math comprehension (Pareto et al., 2012).

Apart from the work that explicitly addresses collaboration and learning companions, one can imagine another set of literature that can help guide us here: research on systems that involve multiple learning companions. By interacting with two virtual learning companions, students can see how collaborative relationships *should* work and consequently model their own behaviors appropriately (Cordar et al., 2015). Relatedly, multiple characters have been shown to better facilitate student learning than a single character (Baylor & Ebbers, 2003a, 2003b). Research has investigated two virtual characters competing in a game show on behalf of their students (Chase et al., 2009; Matsuda et al., 2012), and has shown benefits of different virtual character roles such as "teachable agent" (Biswas, Jeong, Kinnebrew, Sulcer, & Roscoe, 2010), emotionally adaptive character (Girard et al., 2013; Girard & Johnson, 2010) and tutor (Vanlehn et al., 2014). With two characters, we can achieve a give-and-take between agents that makes explanations more interactive and provides the opportunity for productive dialogue around differing opinions, personalities (Chowanda, Flintham, Blanchfield, & Valstar, 2016) and culture (Ogan, Aleven, Kim, & Jones, 2010). The two-character approach has been used

with great success in an interactive computer science museum exhibit (Swartout et al., 2010). With two characters, we can also design humorous dialogues that are engaging to children (Piwek, 2008).

2.1.5 Related Work on Virtual Agents in Non-Education Domains

Before concluding our review of the existing literature on learning companions, we should take note of related work on virtual agents in non-educational domains. For instance, the field of Affective Computing has produced ample affect synthesis research with virtual agents that are not always pedagogical in nature. Examples of such research include studies into the affective influence of an agent's gaze and facial expressions (Castellano et al., 2009) as well as social dialogue (Kumar et al., 2010). Virtual agents have also been found to effectively engage users for a wide range of tasks and contexts. BiLAT is a game-based simulation in which United States Army soldiers can practice negotiation skills that are essential for navigating unfamiliar, high-stakes situations (J. M. Kim, Hill Jr., et al., 2009). The field of healthcare also has benefited from promising research on how virtual agents can help train nurses (Cordar et al., 2015). Relatedly, virtual nurses have successfully assisted hospital patients, some of whom even prefer certain services to be provided by virtual agents rather than by a human doctor or nurse (Bickmore, Pfeifer, & Jack, 2009).

That last piece of research is part of Timothy Bickmore's work on *relational agents*. Relational agents are designed to interact with users repeatedly, building trust and rapport and ultimately forming real relationships. They have proven useful and effective in a range of health care applications, including the provision of social support to older adults, who are often at risk of loneliness and isolation (Ring, Barry, Totzke, & Bickmore, 2013), and the promotion of breastfeeding among new mothers, regardless of their political points of view (Shi, Bickmore, & Edwards, 2015). The requirements of affect synthesis for these agents are slightly different than for agents in short-term interventions; emphasis must be placed on enduring believability and evolution of affect. To wit, just as we would not appreciate a human friend interacting with us with the same exaggerated affect as a museum tour guide (Swartout et al., 2010), a user would not want to continue interactions with a relational agent who failed to evolve socially as the relationship deepened.

Bickmore and Picard provide a comprehensive overview of the challenges, as well as the potential applications of relational agents (Bickmore & Picard, 2005). In illustrating the hazard of "not doing it right," so to speak, they remind us of Clippit, the notorious Microsoft Office Assistant. Although this embodied virtual agent, which took the form of a cartoonish paperclip, may have received high marks in a short-term evaluation, he failed to endear himself to users in the long run:

> "One way to get insight into the problem is to consider an "equivalent" human-human interaction. Imagine an individual that shows up in your office uninvited, with no introduction, barging in when you are busy (perhaps while working on an important deadline). He offers useless advice while projecting the image of being helpful, and then proceeds to ignore your initially polite expressions of annoyance. This character persists in trying to help despite that you increase the clarity of your emotional expression (perhaps through facial expressions or explicit verbalizations). Finally you have to tell the character explicitly to leave, which he eventually does, but first he gives you a wink and a little dance. Would you want to see this character again? If this behavior were that of a human office assistant, then he would eventually be fired, or at least severely marginalized."

Thus we start to see the importance of detecting and adapting to affect for relational agents. Natural language processing, to facilitate conversational dialogue, also plays an integral role. Following a study of the effectiveness of a relational agent within their FitTrack System, Bickmore and Picard were surprised by feedback indicating that users were annoyed by the repetitiveness of the relational agent's dialogue. This negative feedback came despite the fact that the designers had painstakingly authored enough dialogue scripts to offer what they deemed significant variability, and the fact that the users interacted with the system over a duration of just four weeks. Open questions remain concerning how much variability is needed and how to sustain the variability over a much longer period (and ultimately to the point where the relational agent can sustain the relationship *indefinitely*).

Relatedly, a key component for sustaining a long-term relationship is memory. Lim et al. explain the need for including this capability within virtual agents that hope to maintain long-term relationships (Lim, Aylett, Ho, & Enz, 2009). Memories shape our personalities,

decision-making, opinions and intelligence. Lim et al. argue that including 'human-like" memory within relational agents can improve their abilities both to process information and to project a consistent personality that would better facilitate social relations with human users. Work remains to be done on defining how and when information from short-term memory is transferred to long-term memory, as well as clarifying the role of forgetting. Forgetting can be useful for improving efficiency, always a concern for computer scientists as well as cognitive scientists, but the question remains as to what precisely should be forgotten and when. Addressing this question is vital for the believability of virtual agents. Lim et al. note that it also involves ethical concerns: in some instances, a virtual companion should 'forget' sensitive information for the sake of its human user.

As the above research should make clear, creating a relational agent presents a number of complex challenges. The research for my dissertation will avoid some of those challenges by relying on the affordances of the ENGAGE narrative-centered learning environment (See Chapter 3 for a description of the system). Specifically, I will take advantage of ENGAGE being a narrative-centered learning environment. The overarching story can provide context and justification for the presence of the virtual learning companions. The next section will provide an overview of game-based learning in general and then specifically detail some of the research on game-based learning environments that incorporate narrative as a central component.

2.2 Narrative-Centered Learning

Game-based learning in general has been widely utilized for computer science education. Moreover, a growing body of evidence is emerging that suggests game-based learning environments hold great promise for middle school students in particular (Esper, Wood, Foster, Lerner, & Griswold, 2014; D. C. Webb, Repenning, & Koh, 2012). Recent syntheses of the game-based learning literature have found that games can indeed yield positive learning outcomes across a range of subjects and settings (Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012). A recent pair of meta-analyses have independently concluded that, in general, digital game technologies are often found to be more effective than traditional instructional methods in terms of cognitive outcomes, such as learning and retention (Clark, Tanner-Smith, & Killingsworth, 2013; Wouters, van Nimwegen, van Oostendorp, & van der Spek, 2013).

A burgeoning field of research has centered upon classifying players into types based on playing preferences in order to better predict the level of player satisfaction for a given game (Chris, Rebecca, & Lennart, 2011). This research area goes beyond game research to broader user experience research, which has revealed the influence of user's preferences for interactive styles (Hart, Sutcliffe, & De Angeli, 2013). In commercial games, researchers have mapped in-game behavior to real-world demographics, deriving insights into how to better support target demographics (Yee, Ducheneaut, Shiao, & Nelson, 2012). Ongoing research is also investigating how to bring these insights to game-based learning environments, including the examination of whether personalized approaches may better engage particular types of gamers than one-size-fits-all approaches (Orji, Mandryk, Vassileva, & Gerling, 2013).

As with the review of literature on learning companions in the previous section, a look at the research regarding gender here will help inform the research goals of my dissertation. In the United States, where male game developers historically dominate the game industry, developers of game-based learning environments recognize the critical importance of creating games that will appeal to females as well as males (Denner, Bean, & Werner, 2005). Indeed, a wealth of research over the past two decades has examined how to design games that will appeal to girls (Connolly et al., 2012; Jenson, de Castell, & Fisher, 2007; Kafai, 1995). Perhaps most relevant to my research with the ENGAGE narrative-centered learning environment, multiple studies have established that girls enjoy games with a narrative component (Cassell & Jenkins, 1998; Gorriz & Medina, 2000).

Narrative-centered learning environments, which situate learning within a story, provide contextual structure to the knowledge and understanding that students acquire (Mott, Callaway, Zettlemoyer, Lee, & Lester, 1999). Aiming to take advantage of the transformative power of narrative, these immersive environments afford multiple perspectives, situate learning and facilitate transfer (Dede, 2009). Indeed, the theory of *transformational play* takes us beyond any broad claims about the potential of games to make learning "fun", and provides more concrete guidance on how to design game-based learning environments that can positively impact students in ways unique to games (Barab, Gresalfi, & Ingram-Goble, 2010). In this model, the student takes on the role of a protagonist and, using newly acquired domain knowledge, transforms (a) the fictional context of the game environment, (b) her understanding

of the domain knowledge and (c) herself as someone who uses that domain knowledge to solve socially relevant problems.

The game-based learning community has expanded efforts to conduct empirical studies on narrative-centered learning environments over the past several years. For example, a series of studies with the River City game found that students demonstrate positive learning gains and increased inquiry behaviors (Ketelhut, Nelson, Clarke, & Dede, 2010). Quest Atlantis, a popular multi-user virtual environment has been the subject of several quasi-experimental studies, which revealed significant student learning gains (Hickey, Ingram-Goble, & Jameson, 2009), as well as substantial motivational benefits (Warren, Dondlinger, & Barab, 2008) compared to baseline conditions. Research on Crystal Island has suggested that the narrative-centered environments might equitably engage all students because of the multiple forms of interactivity that they can provide, such as interacting with narrative elements, physically interacting with objects in the game environment, and interacting with traditional learning tools such as paper-based worksheets (Rowe, Shores, Mott, & Lester, 2010). Research has also demonstrated narrative-centered learning environments to be useful testbeds for studying virtual agents (McQuiggan, Rowe, & Lester, 2008).

Studies such as these have begun critical progress toward establishing an empirical account of the effectiveness and design of narrative-centered learning environments. My work builds on this research as well as the literature in Section 2.1 on learning companions. As noted above, my research does so in the context ENGAGE, a narrative-centered learning environment for middle school computer science. The next section will provide some background to establish where my research fits into the existing work on middle school computer science education research.

2.3 Computer Science Education in Middle School

Over the past decade, the computing education research community has placed an increasing emphasis on computer science in middle school, culminating in the White House's announcement of the Computer Science for All initiative in January 2016. Curriculum frameworks and interventions have reached millions of students, contributing to significant momentum. We are beginning to see longitudinal studies of K-12 students' computing attitudes

and self-efficacy (Aritajati, Rosson, Pena, Cinque, & Segura, 2015), design-based research examining how well a block-based programming-based curriculum prepares students for later text-based programming (Grover, Pea, & Cooper, 2014), and empirical studies to investigate the factors that affect middle school students' learning of algorithmic thinking (Grover, Pea, & Cooper, 2016). These research projects join a landscape of increasingly diverse computer science interventions. For example, a recent project has combined jewelry design with 3D printing to teach students about technology and programming (Starrett, Doman, Garrison, & Sleigh, 2015), while another project has built an intervention that engages children, together with their grandparents, in classroom activities (Sauppé, Szafir, Huang, & Mutlu, 2015).

Because of the widely recognized need to build students' computational thinking skills, several computer-based learning environments have emerged. Middle school programs have utilized Scratch programming extensively (Burke & Kafai, 2012; Castro et al., 2016; Franklin et al., 2016) and emphasized reaching students with disabilities (Adams, 2010), urban youth (Maloney, Peppler, Kafai, Resnick, & Rusk, 2008) and underrepresented groups (Franklin, Conrad, Aldana, & Hough, 2011). Alice 3D has been used to integrate computing within the context of a wide variety of subjects such as math, science and language arts (Rodger, Qin, & Hayes, 2009), and to help students understand what their future careers in computing might look like (H.C. Webb & Rosson, 2011). A community of practice for middle school and high school teachers has also emerged around teaching introductory computing with Alice (Cooper, Rodger, Schep, Stalvey, & Dann, 2015).

Other languages such as Logo (Lewis, 2010) and more recently Kodu (Stolee & Fristoe, 2011) have also been used in middle school interventions. Emerging work is focusing on building a language and development environment, LaPlaya, tailored for early middle school and upper elementary school (Hill, Dwyer, Martinez, Harlow, & Franklin, 2015), as well as integrating computational thinking into middle school science with CTSiM (Basu, Dukeman, Kinnebrew, Biswas, & Sengupta, 2014). There is also growing momentum for using game design and game programming (D. C. Webb et al., 2012) and robotics (Larkins, Moore, Covington, & Rubbo, 2013) as ways to introduce computer science to middle school students. Efforts have also been made to formalize assessment of computational thinking at the middle school level (Buffum, Lobene, et al., 2015; Werner, Denner, & Campe, 2012).

2.3.1 In-School Initiatives

I am conducting my study as part of an in-school initiative. As explained in Chapter 4, conducting a study in this way has significant benefits. Yet the approach stands in contrast to the bulk of research on middle school outreach efforts: until recently, few research initiatives were conducted with students in school and during normal school hours. To attain a full understanding of the recent work on in-school computer science outreach initiatives, I performed a comprehensive literature review of the SIGCSE conferences from 2011 to 2015. Of the approximately 500 full papers presented at these conferences, I identified 113 papers focusing on K-12 education. Slightly more than half of these K-12 papers reported on out-of-school initiatives (*e.g.*, summer camps), and another quarter of them focused on in-school computer science electives for secondary school students, such as AP Computer Science Principles.

Performing studies on such a general population of students—rather than on a subset of students who self-select to attend computer science activities—can provide greater insight into how all students learn computer science, regardless of prior knowledge or interest. For example, in-school research studies have examined the programming abilities of primary students as related to their reading abilities (Seiter, 2015) and to their general academic performance in other subjects (Oliveira, Nicoletti, & del Val Cura, 2014), in both cases providing better understanding on how *all* students learn programming, and not just those who self-select computer science. Crucially, research has also looked at how to design in-school interventions with an emphasis on sustainability (Koh, Repenning, & Motter, 2013).

Reaching K-12 students with computer science content during school hours ultimately requires empowering K-12 teachers to teach that computer science content. It is important to note that the instructors for after-school and summer initiatives are often different than those who would teach an in-school course for middle school. Accordingly, a significant amount of research in this area has addressed teacher professional development. Perhaps the most popular approach for professional development has been to introduce K-12 teachers to computer science content through summer workshops. A three-day teacher workshop has been used to help middle school teachers incorporate Lego robotics into their classes (Kay, Moss, Engelman, & McKlin, 2014). One-week teacher workshops have used Stencyl, a game

development environment (Liu et al., 2014), App Inventor, an Android development environment (Liu et al., 2013), and Scratch and Alice, two popular drag-and-drop programming environments (Liu, Lin, Hasson, & Barnett, 2011). Other research has looked at introducing computer science content to student teachers through pre-service teacher education, such as through a four-week summer course (Bell, Frey, & Vasserman, 2014) or through a one-week module inserted into one of the required courses for education majors (Yadav, Zhou, Mayfield, Hambrusch, & Korb, 2011). There has also been research on providing in-service teachers with research experience in computer science labs at local universities (Tashakkori et al., 2014), and pairing undergraduate computer science majors with local middle school teachers, which can have benefits for all involved (Burns, Pollock, & Harvey, 2012). While these workshops have been shown to be effective at having an immediate impact on teachers' knowledge and attitudes, less research has been done to examine how likely teachers are to keep teaching the content year after year (Koh et al., 2013). Moreover, little research has been reported on how effectively these teacher benefits translate into measurable impacts on *students*, and particularly on students who are not predisposed to study computer science.

We might reasonably expect greater and more sustainable impacts if teachers have an established *curriculum* to accompany their professional development. One middle school curriculum uses a "braided teaching" approach by interleaving a number of computer science concepts throughout a course for middle school students (Pasternak & Vahrenhold, 2012). Another middle school curriculum focuses on improving computer science attitudes by highlighting the breadth of the field, an approach that should perhaps start at the primary school level (Carter, Blank, & Walz, 2012). Several recent efforts have created pre-secondary computer science curricula by leveraging existing tools also popular with out-of-school initiatives, such as Scratch (Schofield, Erlinger, & Dodds, 2014) and Alice (Rodger et al., 2012). These pre-existing tools offer great functionality. The sheer breadth of possibilities that an open-ended learning environment like Scratch affords, however, can make it difficult for K-12 teachers unfamiliar with computer science to ensure that student work is adhering to a curriculum's learning objectives, perhaps making it less likely that the teachers will fully adopt it (Levy & Ben-Ari, 2007; Ni, 2009). More work needs to be done to investigate whether

relying on these ubiquitous tools privilege students who have already used them in prior computer science activities.

Most initiatives involve some software system, but there has been comparatively little recent work that reports on developing software systems specifically for in-school initiatives. The research that has been reported has led to valuable insights. AgentSheets has been used to assess middle school students' computational thinking (Basawapatna, Koh, Repenning, Webb, & Marshall, 2011). Most importantly for the goal of broadening participation in computing, research on the AgentSheets system has explored ways to use computer science pedagogy to improve computer science attitudes among female students (D. C. Webb et al., 2012). Another large-scale effort, Bootstrap, uses its custom software tool WeScheme to integrate computing concepts with middle school algebra (Yoo, Schanzer, Krishnamurthi, & Fisler, 2011). This custom software enables Bootstrap to reinforce algebra content in ways that a pre-existing tool like Scratch would not, and this ability makes it particularly appealing to math teachers (Schanzer, Fisler, Krishnamurthi, & Felleisen, 2015). With both AgentSheets and Bootstrap, we see the great potential value of developing software systems tailored to the curriculum. Software development requires many resources, of course, so it is unsurprising that so few examples exist. Yet, just as strong curriculum development can be integral to supporting K-12 teachers who have little or no computer science backgrounds, a tailored software system can have a profound additional impact on supporting these teachers.

2.3.2 Equitable Collaboration

Collaborative learning has also been an area of focus in computing education research, most notably in the form of pair programming (Hanks, Fitzgerald, McCauley, Murphy, & Zander, 2011; Lewis, Titterton, & Clancy, 2012). Evidence suggests that collaborative learning provides many benefits for computer science learners, including improved performance and lower attrition among undergraduates (Lasserre, 2009). Pair programming has also been studied with younger learners, and compared to other forms of collaborative learning (Lewis, 2011). Ongoing efforts to develop formal assessments of computational thinking at this level have solidified claims that pair programming has great potential benefits for middle school students (Werner et al., 2012). When compared to non-collaborative learning environments, pair programming can have a particularly positive impact on girls' enjoyment and perception

of learning (Liebenberg, Mentz, & Breed, 2012) and can improve Latina students' perceptions of computer science and aid in developing their identities as computer scientists (Thiry & Hug, 2014). However, recent work highlights the potential negative impact of unbalanced collaboration, in which one partner dominates the learning task, leading to inequitable learning experiences (Shah et al., 2014). Understanding the nuances of how collaboration affects student learning still stands as a critical open research question.

An emerging approach to the study of collaboration in computer science education is game-based learning. The combined benefits of collaborative learning and educational games may lead to increased learning and student engagement in computer science courses (Anderson & Gegg-Harrison, 2013; Nickel & Barnes, 2010). Games that support multiple players may also lead to a more diverse and sustainable learning experience (Foster, Esper, & Griswold, 2013), and research suggests that game-based learning environments should include opportunities for students to collaborate (Isbister, Flanagan, & Hash, 2010). In fact, even when a system does not include opportunities to collaborate in its intended implementation, learners may gravitate towards collaborating with one another regardless of the designers' intentions (Ogan, Walker, et al., 2012). Chapter 5 describes how we built on this research to examine learning gains in a collaborative game-based learning environment that students play over a sustained period of time. The evidence indicates that this fusion of collaboration and game-based learning led to equitable learning gains, regardless of a student's gender or prior experience with similar gaming environments.

Chapter 3   THE ENGAGE GAME-BASED LEARNING ENVIRONMENT

As mentioned earlier, I conduct my research as part of the ENGAGE project,[4] a middle school initiative that has integrated the benefits of collaborative learning with the engaging nature of narrative-centered learning environments. The initiative aims to leverage the benefits of these two strategies to spark interest in computing and promote significant learning gains.

3.1 The System

In the game, students embark on an adventure to liberate an underwater research station from a mysterious villain who has seized it for evil purposes. Playing the role of a computer scientist dispatched to save the day, the students learn and apply fundamental computer science and computational thinking skills (Wing, 2006). As the students advance through the immersive, three-dimensional game world (Figure 3-1), they must employ their nascent computational thinking skills to overcome challenges (*e.g.*, advance across flooded rooms and other obstacles). With the curriculum's learning objectives derived from the AP CS Principles curriculum framework (Astrachan et al., 2011), some of these challenges require the students to write programs in a block-based programming interface, but the emphasis is on students deepening their conceptual understanding rather than only developing programs. Additionally,

---

[4] ENGAGE was funded by the National Science Foundation through Grant CNS-1138497. Any opinions, findings, conclusions, or recommendations expressed in this report are those of the participants, and do not necessarily represent the official views, opinions, or policy of the National Science Foundation. The Principle Investigator for the ENGAGE project was James Lester, with Co-PIs Kristy Boyer, Bradford Mott and Eric Wiebe. Kirby Culbertson was in charge of the artistic design of the game. Wookhee Min was the primary software developer. Megan Frankosky organized the study design and implementation. As a Graduate Research Assistant, my main contributions were in the design of the computational thinking learning activities in the game, and development of the knowledge assessment. I also helped run all studies, focus groups and teacher institutes, as described in Chapters 4-6. The analyses presented in Chapters 4-6 are my own. I also developed the virtual learning companions (described in Chapters 6-7) independent from the rest of the ENGAGE project.

the narrative of the game reinforces the idea of the students becoming computer scientists. By focusing the game's goals on student conceptual understanding and identity formation, the aim is to avoid overly privileging students with prior programming experiences. The project keenly hopes to empower all students to view computer science as a possible future subject of study.

Figure 3-1. Screenshot from ENGAGE, Level 2
In this level, students learn how to interpret binary numbers and about the binary representation of data

The game is divided into three levels. The first level introduces students to basic programming concepts such as sequencing and iteration. The students will later apply these concepts in more sophisticated programs in the subsequent levels. The second level focuses on binary numbers. Students learn how a binary system works, how to interpret binary numbers, and how computers use binary numbers to represent other forms of data, such as text and images. The third and final level focuses on exploration of data. Students write programs to make sense of data they receive from other characters in the game. In addition to learning about algorithms such as sorting and filtering data, students solve challenges that require the use of large data sets to solve globally relevant problems.
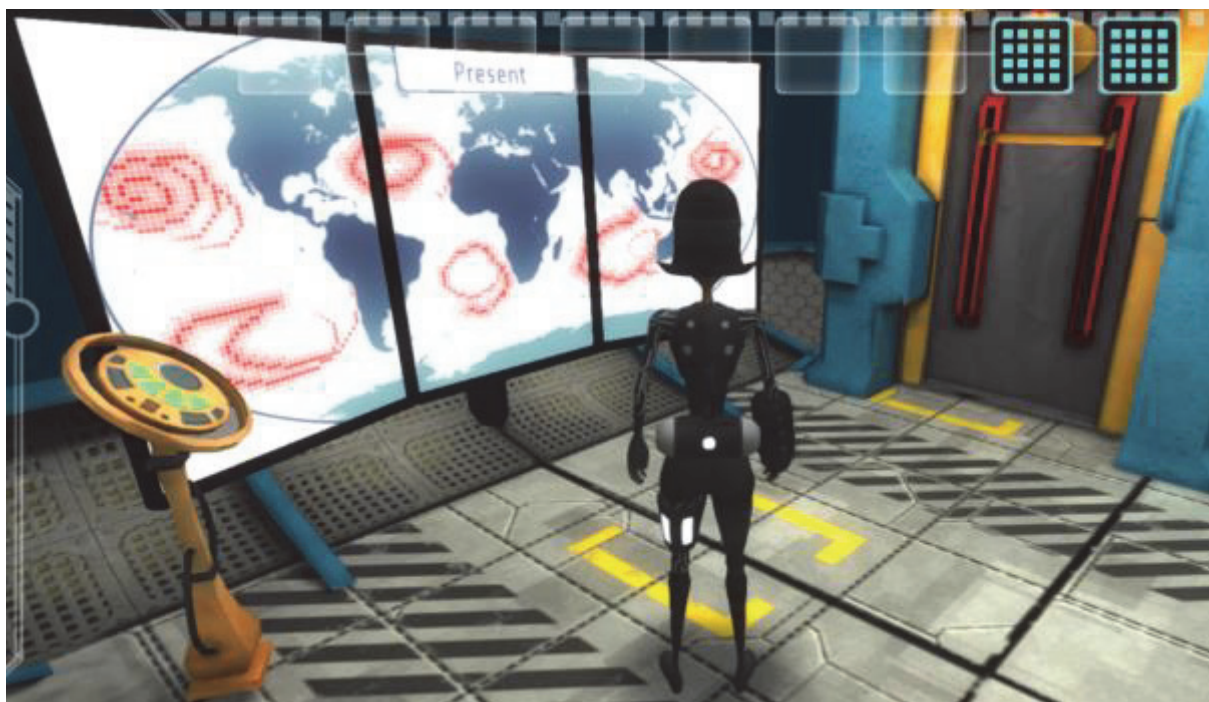
Figure 3-2. Screenshot from ENGAGE, Level 3
In this level students learn about using large data sets to solve globally relevant problems

The game was developed with the Unity 3D game engine. The game's most basic form of interactivity includes the student navigating her avatar around the three-dimensional environment using a keyboard and mouse. The computational thinking layer focuses on solving computing problems, most often writing programs in a block-based programming interface. After completing a program in the programming interface and then running it, students see the results affect the three-dimensional environment, such as moving an object from one location to another. When successful, the students can advance their avatar to the next challenge and consequently advance the non-branching narrative. ENGAGE's narrative (which was completed prior to my addition of the interactive virtual learning companion) is delivered through diegetic pop-up messages from non-player characters in the story.

The system collects game-trace data while the students interact within the game-based learning environment. The SQL database sequentially logs every action that the students take using the two types of controls: mouse and keyboard. Table 3-1 below contains an example gameplay trace.

Table 3-1. Example of a segment of gameplay trace

Group	Real Time	Game Time	Event Type	Attribute
42	9/17/14 8:43	523.5	Room Enter	DW1-EnterLiftRoom
42	9/17/14 8:44	537.86	Dialogue	Open
42	9/17/14 8:44	541.276	Dialogue	Advance
42	9/17/14 8:44	543.342	Dialogue	Close
42	9/17/14 8:44	550.172	Camera Control	Center camera
42	9/17/14 8:44	561.604	Camera Control	Center camera
42	9/17/14 8:44	575.213	Camera Control	Center camera
42	9/17/14 8:44	580.801	Pair Device	Dining Lift 1
42	9/17/14 8:44	584.013	Device	Execute
42	9/17/14 8:44	587.875	Flip tile	Kitchen_Binary_Panel_2
42	9/17/14 8:44	588.936	Flip tile	Kitchen_Binary_Panel_4
42	9/17/14 8:44	590.17	Flip tile	Kitchen_Binary_Panel_3
42	9/17/14 8:45	592.591	Device	Execute
42	9/17/14 8:45	596.387	Programming Interface	Open
42	9/17/14 8:45	598.868	Programming Interface	Close
42	9/17/14 8:45	601.445	Flip tile	Kitchen_Binary_Panel_1
42	9/17/14 8:45	602.494	Flip tile	Kitchen_Binary_Panel_5
42	9/17/14 8:45	604.444	Device	Execute
42	9/17/14 8:45	604.677	Step on lift device	TRUE
42	9/17/14 8:45	610.248	Camera Control	Center camera
42	9/17/14 8:45	647.948	Camera Control	Center camera
42	9/17/14 8:46	655.423	Pair Device	Dining Lift 2
42	9/17/14 8:46	658.248	Programming Interface	Open
42	9/17/14 8:46	661.244	Programming Interface	Close
42	9/17/14 8:46	665.437	Flip tile	Kitchen_Binary_Panel_2
42	9/17/14 8:46	666.655	Flip tile	Kitchen_Binary_Panel_1
42	9/17/14 8:46	667.591	Flip tile	Kitchen_Binary_Panel_1
42	9/17/14 8:46	668.575	Flip tile	Kitchen_Binary_Panel_3
42	9/17/14 8:46	670.448	Flip tile	Kitchen_Binary_Panel_4
42	9/17/14 8:46	673.282	Flip tile	Kitchen_Binary_Panel_1
42	9/17/14 8:46	674.887	Device	Execute
42	9/17/14 8:46	678.289	Device	Execute
42	9/17/14 8:46	681.229	Programming Interface	Open
42	9/17/14 8:46	685.704	Programming Interface	Close
42	9/17/14 8:46	687.318	Flip tile	Kitchen_Binary_Panel_4
42	9/17/14 8:46	687.833	Flip tile	Kitchen_Binary_Panel_3
42	9/17/14 8:46	688.308	Flip tile	Kitchen_Binary_Panel_2

3.2 Classroom Studies of ENGAGE

Our ENGAGE project team has iteratively held a series of classroom studies of the game and associated classroom activities in multiple middle schools with diverse student populations, leading to refinements of both the game and the way we deploy the game within a larger curricula. Over the years, ENGAGE has been deployed in a variety of formats. I will conduct my dissertation study on data collected within the context of ENGAGE being deployed as part of a quarterly oceanography elective for middle school students. As in most of our studies of ENGAGE, the students chose partners before the first day of gameplay and then played the game in pairs over the course of several weeks.

In addition to the data collection for my dissertation study, this document contains various descriptions of pilot studies (including some for the ENGAGE game and others for the virtual learning companions) and some analyses of data collected during previously full classroom studies. The reader might benefit from the following summary:

- Chapter 4 first analyzes data collected from the oceanography election at Carnage Middle School during the 2014-15 school year, and then contrasts the findings to an after-school ENGAGE pilot study from Spring 2015 at Centennial Campus Magnet Middle School (CCMMS).

- Chapter 5 first describes an in-school pilot study of ENGAGE at CCMMS in Spring 2014, and then analyzes data from the oceanography elective at both Carnage and Ligon middle schools during Fall 2014.

- Chapter 6 first analyzes data collected from the oceanography elective at Carnage Middle School during the 2014-15 school year. As the game was still being iteratively refined during this period, the chapter is able to contrast the results from early in the school year (Game-Original) against the results from later in the year (Game-Refined). This chapter then proceeds to analyze the data collected from a virtual learning companion pilot study conducted at CCMMS in December 2014.

- Chapter 7 describes the findings from a virtual learning companion pilot study conducted at CCMMS in April 2015.

- Chapter 8, my Research Design chapter, describes how I collected the data for this dissertation study during the oceanography elective at Carnage Middle School in two phases: Spring 2016 and Fall 2016.

For all deployments involving the oceanography elective, students take a battery of assessment instruments before starting the game. These instruments include a survey to measure a student's computer science attitudes (Wiebe, Williams, Yang, & Miller, 2003) and a knowledge assessment (see Chapter 5). At the end of each day of gameplay, each student completes a "Daily Mini-Engagement" survey that contains six items about the student's affective response to the gaming experience that day. When students complete each of the three levels, they individually complete an interim knowledge assessment that only contains items on material that they learned during the recently completed level. After students complete the entire game, they individually take the full knowledge assessment, as well as the Computer Science Attitudes survey and the full User Engagement Survey (O'Brien & Toms, 2010).

Chapter 4 ENGAGE CLASSROOM DEPLOYMENT: TOWARD "CS FOR ALL"

In planning for the full classroom deployment of ENGAGE, which began in Fall 2014, our team put careful consideration into how computer science remains a "niche" subject in the United States, studied by a generally small, non-diverse population of students. Many initiatives in the United States have sought to broaden participation in computing. Some such initiatives involve out-of-school activities such as summer camps (Aritajati et al., 2015). Other initiatives focus on formal, in-school coursework through the development of innovative pre-college curricula such as Exploring Computer Science (Goode & Margolis, 2011) and AP Computer Science Principles (Arpaci-Dusseau et al., 2013; Gray et al., 2015). Fundamental to all of these initiatives is the mission to engage students who are historically underrepresented in computer science, including female students.

Yet for both of these types of initiatives—out-of-school activities and in-school computer science courses for secondary school students—questions remain as to just how effective they are in broadening participation (McGill, Decker, & Settle, 2015). Both these types of initiatives are likely to involve self-selection, attracting a certain subset of students rather than a truly representative sample. In other words, even if a given initiative has a large percentage of underrepresented students (*e.g.*, female students) among its participants, these students likely either already have interest in studying computer science or have someone in their lives encouraging them to study computer science. Of course, this does not negate the value of these initiatives: nurturing a student's pre-existing interest in computing and empowering her with new skills may contribute greatly to her persisting in the field.

To fully address the goal of broadening participation, however, we must also reach out to students who do not have predispositions or existing influences to study computer science or to participate in extracurricular computing activities. This chapter provides evidence that the key is to develop *in-school* initiatives that enroll a broad population of students (*i.e.*, not students specifically seeking a computer science elective). Furthermore, it is crucial to create such in-school initiatives at the *pre-secondary level*, as students begin their career trajectory as

early as middle school (Lent, Brown, & Hackett, 1994). Indeed, researchers have looked specifically at how the underproduction and underrepresentation issues in undergraduate computer science departments may be traced back to lack of exposure as early as middle school (Shashaani, 1994; Heidi C. Webb, 2011).[5]

This chapter describes how the ENGAGE project became one such initiative, including how we integrated the computing content with an existing middle school science curriculum and trained middle school science teachers to teach this course in four diverse middle schools. The results show that the in-school implementation of ENGAGE improved the computer science attitudes of students with no prior computer science experience, in a way that the corresponding after-school implementation could not. Section 4.1 describes the development and in-school implementation of ENGAGE. This includes software development, curriculum development and teacher development. Section 4.2 then reports on the initiative's success at improving computer science attitudes of those students who were less predisposed to study computer science. Overall, this chapter highlights ENGAGE's potential as a pre-secondary, in-school initiative to reach students who might not otherwise consider computer science as a field of study, and to improve their attitudes toward computer science.

4.1 Design of In-School Initiative

4.1.1 Development Process

Our research and development team developed both the ENGAGE game-based learning environment and its accompanying computer science curriculum synchronously over two years. Although the game-based learning environment can be deployed within a wider curriculum (including non-gameplay lessons with complementary learning activities), we designed the game so that it could feasibly operate in a stand-alone fashion. Students with no prior computer science knowledge can play the game from beginning to end, developing their computational thinking skills with only the in-game activities scaffolding their learning. This

---

[5] For more research context, see the section on this in the Background and Related Work Chapter (Section 2.3.1).

provides more flexibility in how ENGAGE can be deployed, allowing for out-of-school and in-school implementations.

Crucially, designing the game-based learning environment in this way also eases the burden on the teacher for in-school implementations. We anticipated that our middle school teachers would have limited prior computer science knowledge, so we sought to create a software environment that would ease the burden of teaching an unfamiliar subject. We hoped this step would positively influence how likely the teachers would be to teach the course again and maintain "curriculum integrity" (Koh et al., 2013).

In developing the curriculum for in-school implementation, we planned for a 20-session course, with each session lasting an hour. We envisioned a schedule in which gameplay sessions would alternate with non-gameplay sessions. During a given gameplay session, we planned for students to receive an introduction to a certain computational concepts within the game-based learning environment. Then the next day's session would center on a classroom activity that would reinforce the concept. As we progressed with the development of the game, we thus generated ideas for activities that would extend students' understanding of the material, while not being integral to success in the game.

During game development, we delayed drafting lesson plans for the non-gameplay activities because we wanted to include the eventual classroom teachers as co-creators. Ultimately, with the support of district-level stakeholders, we integrated our computer science content with an existing quarterly science elective focused on oceanography. Our partner schools offer this course to students each quarter of the academic year during normal school hours, and it draws a diverse population of students. The entire course consists of 45 hour-long sessions. Our in-school implementation plan thus became to integrate three types of sessions: gameplay sessions, non-gameplay sessions that reinforced computational concepts, and non-gameplay sessions that focus on the pre-existing science content. We approached this task in close collaboration with classroom teachers, as detailed below in Section 4.1.2.

4.1.2 Teacher Institute

In Summer 2014, we held a six-day Teacher Institute, which included a total of 40 hours of workshop activities with middle school science teachers. We recruited four teachers from four different middle schools, each serving a diverse student population. In addition to attending

the Teacher Institute, these four science teachers all committed to teaching our integrated curriculum during one of their class periods for the 2014-15 school year. As described below, the teachers co-designed this integrated curriculum with the research team, supplementing the research team's *computer science education* and *game-based learning* expertise with the teachers' *science education* and *middle school teaching* expertise.

In preparation for the Teacher Institute, the research team created a teacher kit that included a project overview, a proposed classroom implementation plan, guidelines for survey administration and a summary of key locations in the game. These key locations consisted of 1) challenges that students might find particularly difficult and 2) areas in which the teacher might especially impact deeper student learning by asking targeted questions about the given topic. Additionally, the teacher kit contained placeholders for lesson plans that would be created collaboratively during the workshop sessions. We aimed to produce detailed lesson plans specifying the learning objectives, necessary materials and connection to the content in the game-based learning environment. We also planned for the Teacher Institute to produce an updated in-school classroom implementation plan, clarified by teacher input.

In designing the structure for the Teacher Institute of 2014, we built upon our experience conducting a series of workshops the previous summer (Buffum et al., 2014). We devoted the first day of this new Teacher Institute to an overview of the project. This included an introduction to the game-based learning environment, a summary of the computer science content and our vision for how we might integrate the computer science content with the existing science curriculum. On each subsequent day, we used the morning hours to provide the teachers a solid understanding of both the project and the content knowledge. Since these teachers did not enter with backgrounds in computer science, we needed first to teach them the computer science topics that the students would be learning in the curriculum. We then used afternoon sessions to give the teachers time to draft lesson plans that would incorporate this computer science content into the existing science curriculum.

By the end of the Teacher Institute, we had produced a detailed in-school implementation plan, endorsed by all teachers. As noted above, the entire curriculum was designed for an academic quarter lasting approximately two months. We expected that the implementation plan would remain consistent from quarter to quarter, but, anticipating the

need to refine the plan iteratively, we only solidified a calendar for the first academic quarter. The goal was to set dates for certain lessons so that members of the research team could coordinate with the teachers in preparing those class sessions. As described below, members of the research team worked closely with the teachers during the first quarter to troubleshoot issues with the game-based learning environment, to provide content knowledge support to the teachers and to gather observations on how to improve the classroom implementation.

4.1.3 Study Implementation

Due to school constraints and other factors, not all teachers taught this curriculum each quarter of the 2014-15 academic year. During the first quarter in which each teacher taught the course, members of the research team attended most gameplay sessions and all non-gameplay sessions that involved computer science content. Teachers had the responsibility of delivering all lessons but could request assistance from the research team member in the case of any uncertainty. For the most part, the teachers did not need to rely on in-class support from the research team. They reported satisfaction with the training that they received during the Teacher Institute, as well as appreciation for how the game-based learning environment supported instruction.

The classroom study saw over 200 students complete the course during the 2014-15 academic year. Students overwhelmingly responded positively to the experience on post-surveys, and classroom observations further established the game's success at engaging students. Furthermore, preliminary studies have revealed significant learning gains, as measured by ENGAGE's purpose-built knowledge assessment instrument (Buffum, Frankosky, et al., 2015; Buffum, Lobene, et al., 2015). The rest of this chapter, however, will focus on another dimension of the initiative's success: reaching a diverse student population, as measured by computer science attitudes.

4.2 Comparison of Students With and Without Prior Programming Experience

As discussed above, the desire to reach a diverse student population motivated our in-school implementation strategy. The fundamental goal of our project is to broaden participation in computing. While there exist many useful approaches to supporting this grand goal, we focus

on reaching *students who might not otherwise consider computing fields*, helping them build their computational thinking skills and positively impacting the attitudes that they hold towards computer science.

This requires us to go beyond merely looking at traditional demographic information to evaluate our success. In previous pilot studies, conducted as out-of-school initiatives, we succeeded in recruiting a large percentage of female students (Buffum et al., 2014). Yet the self-selection that is inherent to such out-of-school initiatives allows for the possibility that we may only have been reaching the female students who already felt predisposed to study computer science, or who had parents who encouraged them to study computer science. Furthermore, even if we "hid" computer science content within some out-of-school activity advertised as another discipline, which can be successful at reaching students who are not predisposed to computer science (Franklin et al., 2011), we still might face issues of access that in-school activities largely avoid.

Thus, while we want to ensure that our initiative appeals to student groups generally underrepresented in computer science (and we have, in fact, analyzed ways in which it does so (Buffum, Frankosky, et al., 2015)), we also want to measure the success of our initiative at specifically impacting students with less of a predisposition toward studying computer science. To begin examining this, we can compare the computer science attitudes of students based on their prior computing experiences. This chapter reports on the data of one our Raleigh, North Carolina partner schools, Carnage Middle School, which offered the course each quarter of the 2014-15 academic year. Carnage is an urban middle school serving a racially and ethnically diverse community, with over one third of its students receiving free or reduced lunch. It must be noted that Carnage does provide its students a relatively high degree of exposure to computing due to its STEM theme and its proximity to a technology hub.

Of the 84 total students who completed all surveys (including pre- and post-surveys on computer science attitudes), 31 were female students and 53 were male students. As for race or ethnicity, the in-school implementation included 12 African-American students, 19 Asian students, 6 Latino/a Students, 1 Middle Eastern student, 6 mixed/multiracial students, 19 South-Asian students and 21 white students. Meanwhile, on the survey item, "Have you ever

participated in any activities that involve computer science or computer programming?", 48 of the participants reported "yes" and 36 reported "no".

To compare the computer science attitudes of those with and without prior programming experience, we utilized an attitude survey that was originally validated for college students (Wiebe et al., 2003) and which was modified for middle school students. This computer science attitudes (CSA) survey includes three subscales: *confidence* in computer science skills, *usefulness* of computer science and *motivation* to study computer science. Perhaps unsurprisingly, the students with prior programming experience scored higher on this CSA survey overall and on all three subscales, as seen in Table 4-1.

One-way ANOVAs found all these differences to be significant with p–values less than .05. The confidence subscale saw an especially stark difference ($p < .001$). We expected to see such differences in computer science attitudes on this pre-survey. Students with prior programming experience may express more positive computer science attitudes because of prior programming experiences they have had. They may also have participated in those prior programming experiences due to their existing predispositions toward studying computer science.

Table 4-1. CS Attitudes Pre-survey: Students with/without Prior Programming Experience
From Computer Science Attitudes Survey (CSA)

	Prior Programming	**No Prior Programming**
CSA Overall	3.64 (SD = .62)	3.11 (SD = .65)
Confidence Subscale	3.88 (SD = .74)	3.08 (SD = .85)
Usefulness Subscale	3.59 (SD = .81)	3.23 (SD = .80)
Motivation Subscale	3.44 (SD = .75)	3.01 (SD = .64)

The reinforcing interplay of *participation* and *interest* in computer science activities seems to start early. With this in mind, our in-school implementation aims to have particularly strong positive impacts on middle school students who *do not* have prior exposure to computer programming. To measure the initiative's success in this goal, we can look at the results of the

CSA post-survey. As Table 4-2 shows, the pre-existing differences in computer science attitudes largely disappeared by the time students took this post-survey.

Table 4-2. CS Attitudes Post-survey: Students with/without Prior Programming Experience
From Computer Science Attitudes Survey (CSA)

	Prior Programming	**No Prior Programming**
CSA Overall	3.53 (SD = .9)	3.32 (SD = .86)
Confidence Subscale	3.6 (SD = 1.05)	3.41 (SD = .95)
Usefulness Subscale	3.54 (SD = .96)	3.27 (SD = 1.02)
Motivation Subscale	3.46 (SD = .94)	3.29 (SD = .82)

On the post-survey, neither the CSA overall nor any of the subscales revealed any significant differences between students with and without prior programming experience. The most dramatic shift occurred in the confidence subscale, where students with no prior programming experience increased their confidence to near the level of those with prior programming experience. A repeated-measures ANOVA found this increase from pre-survey to post-survey to be statistically significant ($F(1, 35) = 5.039$), $p < .05$). In many ways, this should be expected; following their participation in this initiative, all students on the post-survey now have some prior programming experience, so in some ways they all leave the initiative in the same category. If we had not situated this initiative in-school, however, we may never have reached those students with no prior programming experience. Now, students in this crucial demographic leave the initiative feeling empowered, and perhaps more motivated to study a subject that they had never before considered.

To further illustrate the potential value of in-school initiatives, we compare the in-school implementation to an after-school activity our research team conducted in Spring 2015. For this activity, we advertised for student participants at Centennial Campus Magnet Middle School (CCMMS), which is *not* one of the middle schools where our course is offered during the school day, but is in the same district. Compared to Carnage, CCMMS has even larger percentages of students from underrepresented groups and a slightly larger percentage with

free or reduced lunch (over 40%). Similar to Carnage, CCMMS provides its students a relatively high degree of exposure to computing due to its STEM focus, enhanced even further by its close proximity to an engineering university.

The participating students stayed after school each day for two weeks, during which they played through the entirety of the ENGAGE game-based learning environment. We accepted anyone who wished to participate, but we placed an emphasis on attracting female students in order to have a relatively even gender split. A total of 18 students ended up participating in this after-school activity: 10 male students and 8 female students. There were 11 white students, 3 African-American students, 2 Asian students and 2 Latino students. We see here an over-representation of white students from a school population in which about two-thirds of students are non-white, and this is consistent with study results of other out-of-school initiatives (McGill et al., 2015). Furthermore, all but one of the after-school participants reported "yes" to the survey item, "Have you ever participated in any activities that involve computer science or computer programming?"

In other words, although we attracted a participant pool that contained a fair representation of female students, this after-school activity failed to provide the diversity of participant pool that could help us understand how the game-based learning environment impacts students who have no pre-existing interest in computer science. Unsurprisingly, these students all reported positive computer science attitudes, with an overall 3.89 (SD = .49). Indeed, these after-school participants reported positive computer science attitudes even in comparison to the in-school students with prior programming experience (reporting an average of 3.64, as seen above in Table 4-1). These differences in computer science attitudes get at the root of the problem we face with the out-of-school study implementation. Students who participate in these types of activities may already have an above-average interest in computer science. For an initiative such as ours that aims to reach a broad population of students, these out-of-school study implementations may not provide valid and generalizable assessments of how well our initiative affects students' learning and attitudes.

4.3 Conclusion

This chapter highlights an area of need in K-12 computer science education research: in-school initiatives. Many current outreach initiatives involve some degree of self-selection. In the case of out-of-school initiatives, such as summer camps or after-school clubs, participants tend to already feel motivated to study computing. Further, these activities often place demands on the participants' families (*e.g.*, transportation to and from the activity) that may disadvantage students from underrepresented backgrounds. Innovative curricula for computer science electives at the secondary school level address this issue of access, but they still likely enroll students who enter with higher computer science attitudes than the overall population. Furthermore, some evidence suggests that secondary school may be too late for significantly impacting computer science attitudes. To fully support the goal of broadening participation in computing, we must create in-school initiatives at the pre-secondary level.

Merely looking at demographic information (*e.g.*, gender) does not provide enough information as to whether an initiative is reaching a truly broad population of students. We must reach out to students who would not otherwise consider computer science as a subject of study, in addition to nurturing the computing careers of underrepresented students who already have pre-existing interest in the field. Through research on initiatives with this goal, we can gain insight into how best to design learning activities that appeal to young students with little incoming computer science experience and low predisposition for computer science interest.

In future work, it is important to investigate how the various areas of development (teacher professional development, curriculum and purpose-built software) interact with one other to produce the most effective initiatives. A key issue is sustainability. While many projects have reported on the immediate results of an initiative, we need to follow up to see whether—and how—teachers are continuing in successive years. Finally, as the overarching goal is to recruit underrepresented students into computing, it will be critical to conduct longitudinal studies to measure the extent to which these in-school outreach initiatives empower students to study computer science at advanced levels.

Chapter 5 GENDER EQUITY AND COLLABORATIVE GAME-PLAY

As mentioned in the previous chapters, addressing the national need for a computationally skilled workforce requires that rigorous computer science learning must go hand in hand with increased participation of students from underrepresented groups (Guzdial, Ericson, Mcklin, & Engelman, 2012). Accordingly, the computer science education research community has identified the critical need to create a classroom climate that fosters student learning and retention for these diverse learners (Barker, O'Neill, & Kazim, 2014). At the K-12 level, many current initiatives in the United States seek to broaden participation in computing, including the development of innovative pre-college curricula such as Exploring Computer Science (Goode & Margolis, 2011) and the AP Computer Science Principles course (Astrachan et al., 2011). Fundamental to these initiatives is the mission to engage students who are historically underrepresented in computer science, and also to support learning in a measurable way. An increasingly central element that pervades these curricular innovations is *collaborative learning*, in which students work together to solve problems (Coleman & Lang, 2012; Falkner, Falkner, & Vivian, 2013).

To measure the ENGAGE's effectiveness at supporting collaborative learning, we have administered several instruments during classroom studies, including a knowledge assessment (Buffum, Lobene, et al., 2015). This chapter examines the results of this knowledge assessment, providing insight to the potential benefits of collaborative gameplay, particularly for girls. In addition to examining overall performance on this assessment instrument, the chapter examines one exemplar in detail to illustrate the findings. Like many of the assessment items, this exemplar focuses on a segment of block-based code, which the student must interpret using algorithmic thinking. In this particular item, the algorithmic thinking centers on the concept of broadcasting, a key concept found in many other K-12 computer science interventions that use block-based programming (Burke & Kafai, 2012; Franklin et al., 2013; Smith, Sutcliffe, & Sandvik, 2014).

This chapter aims to describe the ENGAGE project's strategy for improving gender diversity in computer science activities through collaborative learning in a game-based

environment, as well as results establishing the effectiveness of that strategy for supporting learning. Section 5.1 describes an early ENGAGE pilot study (Spring 2014) conducted with two conditions: paired gameplay and single-player gameplay. The results of that pilot study suggested a paired gameplay approach has merit for supporting learning, particularly for female students. We thus proceeded with the paired gameplay approach in the full classroom deployment of ENGAGE (Fall 2014), occurring over eight weeks in the context of the oceanography elective at Carnage Middle School. Section 5.2 describes the learning gains of the students in that Fall 2014 oceanography class, with particular attention paid to differences based on gender. Section 5.3 then discusses the implications of the findings. In sum, the results show the promise of integrating paired gameplay with game-based learning environments to support computer science learning at the middle school level.

Figure 5-1. Screenshot of the Tri-Level Room from Level 1
(Later chapters will discuss this room in more detail.)

5.1 Pilot Study of ENGAGE Game: Single-Player vs. Paired Gameplay

As part of our user-centered development, we conducted an exploratory pilot study in which middle school students played the introductory level of ENGAGE (Frankosky, Wiebe, Buffum, & Boyer, 2015). Figure 5-1 shows a screenshot of ENGAGE from the segment of the game that students played during this pilot study. The study took place over one week in the spring of 2014, with each student participating in two hour-long sessions on separate days. Students participated during the school day as part of a regularly scheduled class and were randomly assigned to play the game either individually or collaboratively in pairs.

We designed this pilot study to investigate the impact of collaborative paired gameplay on student outcomes and experiences. The analyses in this chapter have a particular focus on cognitive outcomes, focusing on students' use of computational thinking. We assessed computational thinking through field observations, survey responses and an early version of our knowledge assessment instrument. We also used the pilot study to refine this instrument, which ultimately became our primary tool for measuring the learning gains of students in full studies of ENGAGE (Buffum, Lobene, et al., 2015). For the full classroom deployment described in Section 5.2, we used the more refined version of the instrument to track students' learning throughout a multi-week intervention. In this current section, we discuss the findings of the pilot study that informed the design of that full classroom deployment.

Participants. For the pilot, we worked with two seventh grade middle school teachers and their classrooms at Centennial Campus Magnet Middle School in Raleigh, North Carolina. After consent and assent was obtained, 28 seventh grade students were randomly assigned to either the paired (N = 14) or unpaired condition (N = 14) and then played the ENGAGE game across the span of two separate gameplay days. The random assignment may have resulted in more male-female pairings than we have seen in other studies (in which students choose their partners): 3 of the 7 pairs for this pilot study were mixed gender. Of the 28 total students, 26 completed a post-survey on engagement (two students assigned to the paired condition arrived very late and did not have time to finish the game or post-survey).

By conducting the pilot study during the normal school day hours and asking all students in the two classes to participate, we expected to achieve a more representative subject pool than a self-selecting, after-school study might provide. This strategy proved successful, as the participants included 14 female students and 14 male students. The demographic composition was 7 African-American students, 7 white students, 6 Latino students, 2 multiracial students and 1 Asian student, (with 5 unknown). Using a survey item that asked, "Have you ever done any activities that involve computer science or computer programming?", we classified 11 students as having prior programming experience and 12 students as not having prior programming experience.[6] Using a survey item that asked, "How often a week do you play computer or video games?", we classified 12 students as frequent video game players (those who responded to the item with "daily" or "almost daily") and 14 students as less frequent video game players.

Task. Once the students had been randomly assigned to their workstations, we had them log onto laptop computers and each individually complete the early version of the knowledge assessment. The version administered for this study consisted of 6 multiple-choice questions on programming concepts we expect students to master in the specific segment of game played in this study and took about 5 minutes to complete. Students in the single player condition were allowed to begin the game immediately upon completion of this assessment. For students in the paired player condition, we waited until both students in a partnership had completed the assessment and then briefly gave them instructions on how paired gameplay can work before having them start the game.

When played by two people, the game allows each player to select an avatar to represent him or herself. Only one avatar is visible in the game environment at any given time, switching at predefined intervals. We encouraged students to alternate who controlled character movement based on which avatar was visible. In effect, we encouraged a style of gameplay similar to pair programming, in which students alternate between being the driver (at the

---

[6] Some of the descriptions that students gave of prior programming experience were ambiguous, making it difficult to classify all students.

keyboard) and being the navigator (advising the driver). While we did notice some alternative approaches to pair gameplay among students in other grades, the seventh graders in this study all seemed comfortable adhering to this style.

As noted above, the pilot study was designed for two hour-long sessions on separate days. No student in either condition finished the game during the first session, so they all had to stop mid-game and resume two days later. During the second session, students completed the segment of game used in the study. Upon completion, they then took the knowledge assessment again, followed by a survey on their game experience, including items addressing specific game strategies that they may have used while playing.

Results. In this Spring 2014 pilot study of ENGAGE, we captured a rich, multifaceted corpus of data, including survey data, field observations and learning gains derived from the knowledge assessment. This chapter focuses on using the data to assess computational thinking. The following subsections include survey data that indicate gender differences were found in the computational thinking strategies used by students during this segment of gameplay, observational data that reveal some of the benefits of paired gameplay (along with some caveats), and knowledge assessment data that highlight areas of concern for collaborative gameplay interventions. In sum, the results show that female students may initially have been disadvantaged due to less prior game experience, but that a collaborative gameplay approach has benefits that could help mitigate this inequity if deployed with careful forethought.

5.1.1 Student Self-Reported Strategy Results

After completing the gameplay session, students completed a survey that included several items concerning their reaction to the game, as well as items on their prior relevant computing experiences and their demographic information. In analyzing the data, we looked for differences in game strategies used by students playing the game. In our observations, certain strategies seemed especially beneficial to students as they used their computational thinking skills to progress through the game environment. Collaboration might lead to wider use of these useful tactics, as students share their "best practices." Without collaboration, students with less gaming experience seem to be at a disadvantage, as results show them less likely to take advantage of these strategies.

One survey item in particular illustrates this challenge, an item we refer to as *Test Platform*. This item asked student participants to respond to the following question on a 5-point Likert scale: "How often within the game did you test the program for the moving platform without being on the platform?" This item refers to certain locations in the game where students needed to program a moving platform device, which (if programmed correctly) can transport the player's avatar to a desired location. If the avatar is standing on the moving platform and the platform crashes into an object (*i.e.*, the program was flawed), the avatar will fall off and be forced to repeat some prior gameplay. Figure 5-2 shows a screenshot of this happening to the player's avatar. Chapter 6 goes into more detail on the frustration that falling in the water can cause students, as well as our various approaches to addressing the issue. Notably, however, students can potentially avoid the negative consequence of crashes altogether by employing a strategy that I will refer to as Test Platform. In the Test Platform strategy, the player runs the program for the moving platform *before getting on it*. In this case, the player's avatar does not risk falling off the platform if the platform crashes. For this pilot study, we did not advise students to use this strategy; rather, students discovered it independently.

The results show that female students used the Test Platform strategy less often than male students. Whereas male students responded to the 5-point Likert scale item on this strategy with a 3.54 (SD = 1.450), female students responded with an average 2.21 (SD = 1.251). A one-way ANOVA found this to be statistically significant ($F(1, 26) = 6.482$, $p < .05$). Although we can view the Test Platform strategy as an example of computational thinking, it is interesting to note that no significant differences were found between students who reported having had previous programming experience and those students who reported none.[7]

However, while prior programming experience may not have had an influence on how often students used the Test Platform strategy, prior *gaming* experience did. Students who reported playing video games "daily" or "almost daily" responded to the item on this strategy with a 3.58 (SD = 1.621), while those with less frequent video game experience responded

---

[7] A similar proportion of male and female students reported having previous programming experience.

with an average 2.27 (SD = 1.100): a statistically significant difference (F(1, 26) = 6.302, $p <$ .05). It must be noted that female students reported a lower amount of weekly video game experience, responding with an average 3.21 on a 5-point Likert scale, compared to an average for male students of 4.0. A one-way ANOVA found this to be statistically significant (F(1,26) = 5.667), $p <$ .05).

Figure 5-2. Screenshot of the potential result of a student's programming error
When an error in a program causes the moving platform to crash, the player's avatar will fall off (if it is currently riding the platform).

Taken together, these preliminary results suggest that students' prior experiences may inform the actions that they take within a game-based learning environment. These prior experiences may help students quickly discover effective game-based learning strategies. If playing alone, an inexperienced student might eventually discover these strategies over time. However, that student might be more likely to do so if collaborating with other students with diverse prior experiences.

5.1.2 Observations of Single-Player and Paired Gameplay

The field observations support this claim that students can gain significant benefits when playing collaboratively with a partner. Overall, observations of students' interactions within the game indicated that students from both conditions had an enjoyable experience. Students' postgame comments echoed this sentiment. From a paired team member: "I think that this game was awesome and that I had fun playing with my partner", and from a single player: "I had lots of fun playing it by myself". The classroom teachers, who were in attendance for the entire duration and played the game themselves, commented that the students seemed particularly focused compared to a typical day.

Throughout the overall gaming experience, observations revealed advantages for paired gameplay versus single gameplay. For example, a given student might use the partner as a "sounding board", or the partner might provide suggestions for what to try next or reasoning about what was happening within the game. One student put it succinctly: "I enjoyed working with my partner because he helped me when I was trying to figure the game out." Additionally, in cases where one member of a pair was having great difficulty with character navigation (likely due in part to lack of gaming experience), the partner could take over during those times when character navigation was particularly tricky. Finally, students playing in pairs had the potential to receive superior affective support from their partners. If one student started feeling frustrated or discouraged, the partner could revive the frustrated student's spirits through the social collaboration.

We also paid close attention to potential disadvantages, however. While the single player students did not have the advantage of a peer who could provide encouragement and support, they also did not have anyone criticizing their actions. While most instances were in jest, there was some element of cross-partner frustration. Because the pairing of students was random, it is unsurprising to see some partnerships led to more successful interactions than other partnerships. In later classroom implementations of ENGAGE (such as the one described below in Section 5.2), we allowed students to choose partners rather than use random pairing.

5.1.3 Knowledge Assessment Results

Considering the benefits mentioned above of collaboration, we hoped to see superior learning gains from those students who play the game in pairs. At the time, we were still developing

our knowledge assessment (Buffum, Lobene, et al., 2015), which students took both before and after playing the game. Because the instrument had not been fully validated at this point, we must interpret the score results (which were similar for both conditions) conservatively. An examination of the results from individual questions, however, can provide specific insight into how well the students mastered certain concepts that we hoped they would learn. In particular, we were interested in how well the students mastered the concept of broadcasting, since field observations revealed that many students found the introduction of this concept in the game to be particularly challenging.

Many computer science interventions in K-12 have included broadcasting as a key concept (Burke & Kafai, 2012; Franklin et al., 2013; Smith et al., 2014), prompted in part by its prominence in block-based programming languages such as Scratch (Maloney, Resnick, Rusk, Silverman, & Eastmond, 2010). For example, the Exploring Computer Science curriculum includes an entire session on broadcasting (Goode & Margolis, 2011). Roughly defined, "broadcasting" is a specific type of event in which one device (or, in Scratch's terms, "sprite") sends a message to another device that is listening for some message. Our game features this concept towards the end of the game segment that students played in the pilot study. The virtual room in which broadcasting is the primary component requires the student to broadcast the message "open" to the exit door. In solving this challenge, we expect students to first review the read-only program for the door, revealing that it is waiting for the "open" message. Students must then use a pressure pad device that can broadcast messages, modifying the pressure pad's program to broadcast the correct message. Finally, students must run the programs for both the door and the pressure pad, at which point the door will open and they can exit the room. Later challenges in the complete version of the game reinforce this concept with further complexity, but students in the pilot study did not confront those additional challenges in the segment of the game available to them.
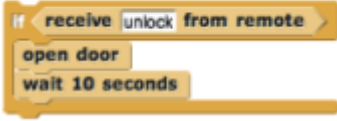
Figure 5-3. Illustrative example: Question 4 on the knowledge assessment
*Administered on both pretest and posttest. This item primarily addresses the concept of **broadcasting**.*

The knowledge assessment item that addressed this concept is shown in Figure 5-3. On the posttest, every student in the single player condition answered this item correctly, yet only 8 out of 14 students in the paired player condition answered it correctly. Although based on a small sample, this difference may illustrate one of the potential concerns of a paired gameplay approach: a stronger partner can advance the pair through a challenge without the weaker partner understanding how the challenge got solved. This issue has been seen in other K-12 computer science studies that emphasize collaborative learning (Shah et al., 2014). Of the seven partnerships, this problem may have manifested in four, as evidenced by one of the partners answering the broadcast item correctly and the other answering it incorrectly. Addressing this drawback is a major open question for designing game-based learning environments that support paired gameplay. As the next section describes, however, this negative consequence may decline over time in longer-term collaborative gameplay interventions.

5.2 ENGAGE Classroom Deployment: Learning Gains Among Students Playing in Pairs

Having revised the narrative-centered learning environment and the knowledge assessment instrument following the results of the above pilot study, we then conducted a full classroom deployment of ENGAGE in the Fall of 2014 at Carnage and Ligon middle schools, two urban schools in Raleigh, North Carolina. In contrast to the pilot study, which we had conducted in a controlled environment, this classroom deployment was integrated into a quarterly science elective (focusing on oceanography). At each school, a cohort of students attended the elective five days a week during their regular school day. One of the school's full-time teachers taught the elective, with members of our research team attending the gameplay session to provide support and record field observations. Over the course of the quarter (approximately two months), several class sessions a week were given for students to interact in the game-based learning environment.

Each student chose a partner on the first day and then collaboratively played the game with that same partner throughout the quarter. This paired gameplay model was motivated by the results from the pilot study, and also by logistical concerns. Because of limited technology in the computer labs of the two schools (a common issue in under-resourced schools), it would have been infeasible to have every student play the game individually on separate computers. The paired gameplay model thus allows deployment of the game-based learning environment with half as many working computers as there are students in the class.

Participants. This section reports on the 48 students who played the game in pairs during Quarter 1 of the elective (and gave consent for their data to be used). Of these students, 26 were male and 22 were female. The demographic composition was 21 White students, 13 Asian students, 8 African-American students, 2 Latino students, 1 Middle Eastern student and 3 other. On the survey item asking about prior computer programming experience, we classified 11 students as having prior programming experience and 29 students as not having prior programming experience. On the survey item asking how often they play computer or video games, 21 students responded "every day" or "almost every day", while the remaining 27 students responded "occasionally" or "almost never."

Task. During the first week of Quarter 1, before the initial introduction of the game-based learning environment, students completed the refined version of the knowledge assessment instrument.[8] We used this as a pretest to measure their incoming computational thinking skills. The full test consisted of 23 items that covered the entire gameplay, which is distributed over three distinct game levels. However, since only the first two game levels were available for this classroom deployment (Level Three was still in development), we will report only on the 15 items aligned with the first two game levels.

Students then played the game in pairs during the class sessions scheduled for gameplay (roughly every other day was reserved for gameplay, with the interceding days full of complementary science activities). When a pair finished Level One of ENGAGE (which occurred after three to five gameplay sessions, on average), both partners individually completed an interim posttest. This interim posttest included the items on the knowledge assessment that we expected students to learn while playing Level One.

Upon completion of that test, the pair would then resume gameplay in Level Two. This level, which is both longer and demands more complex computational thinking, took students approximately 7 to 10 additional gameplay sessions. After a pair finished this level, they completed the full knowledge assessment instrument as a posttest. For the purpose of this chapter, we break down the knowledge assessment into "Level One content" and "Level Two content", depending on where in the game we expected students to learn the concept targeted by an individual assessment item. There are four items that assess concepts introduced in Level One and 11 items assessing concepts from Level Two.

Results. The computational thinking knowledge assessment addresses our need to evaluate how well the game-based learning environment serves all students. By administering it as a pretest, we were able to assess the extent to which students already had these targeted computational thinking skills. We expected that some students would enter with more

---

[8] The validation process for this assessment instrument is still ongoing.

knowledge than others. Indeed, even at the middle school level, students may have widely varying exposure to computer science. Moreover, we hypothesized that students traditionally underrepresented in computer science would score lower on the pretest than their peers. This chapter is focused on underrepresentation based on gender, as well as whether there are differences based on students' prior programming or gaming experiences. Table 5-1 illustrates the differences we found along these three metrics. Overall, students scored an average of .458 (SD = .219) on the pretest (a perfect score would be 1.0), but significant differences were found between female students and male students.

Table 5-1. Comparison of Computational Thinking Knowledge on Pretest

	Performance on Pretest	**Significance**
Gender	*Female*: .383 (SD = .188) *Male*: .517 (SD = .227)	$p < .05$
Prior Programming Experience?	*No*: .42 (SD = .152) *Yes*: .63 (SD = .243)	$p < .01$
Frequent Video Game Experience?	*No*: .385 (SD = .187) *Yes*: .558 (SD = .225)	$p < .01$
Overall	.458 (SD = .219)	

To better understand the difference in pretest scores between male and female students, we ran a One-Way ANOVA in SPSS and the results showed a statistically significant difference ($F(1,43) = 4.486$, $p < .05$). The disparity was even greater between students based on their prior experience with programming ($F(1,39) = 10.456$, $p < .01$) and video games ($F(1,43) = 7.952$, $p < .01$). Thus, just as we had found during the pilot study that frequent video game experience correlated with more frequent use of the beneficial Test Platform strategy, our assessment instrument revealed a similar disparity in pre-knowledge of computational thinking concepts.

Having established that underrepresented students did indeed enter with less knowledge than their peers, we next compared pretest scores with posttest scores to examine learning progressions of students. As mentioned above, we also administered an interim posttest after students completed Level One of the game to measure the extent to which student had mastered those concepts early on. Table 5-2 shows the average scores for female and male students on each of these tests, displayed by level.

Table 5-2. Gender Comparison of Computational Thinking Knowledge Across Time

	Female (n = 19)	Male (n = 21)
Level 1		
Pretest (Level One content)	.51 (SD = .282)	.58 (SD = .266)
Interim Posttest	.434 (SD = .261)	.691 (SD = .315)
Posttest (Level One content)	.645 (SD = .268)	.726 (SD = .315)
Level 2		
Pretest (Level Two content)	.36 (SD = .183)	.53 (SD = .261)
Posttest (Level Two content)	.63 (SD = .246)	.70 (SD = .212)

The results indicate that, while female students demonstrated less knowledge early on, they made great gains as they progressed in the game. A one-way ANOVA found the difference between genders on the Interim Posttest for Level One to be statistically significant ($F(1,39) = 7.735$, $p < .01$). The difference between genders on the pretest of Level Two content was also statistically significant ($F(1,39) = 5.193$, $p < .05$). Thus, at this early stage of the gameplay, we do not see the gender gap closing. Indeed, the normalized learning gain here is disheartening. We calculated learning gain as (Post – Pre)/(1 – Pre). When using the Interim posttest, this calculation showed the male students as having a higher learning gain (.387 for males, compared to .110 for females). A one-way ANOVA found this to be statistically significant ($F(1, 39) = 5.684$, $p < .05$).

Yet the longer students interacted in the game-based learning environment, the less these differences manifested themselves. On the posttest, no statistical differences were found between male and female students, as female students' greater learning gains leveled the playing field. Indeed, female students mastered the Level One concepts as they saw them and applied them more often during Level Two, and this did not take away from their learning of the Level Two concepts. Here we see the importance of persistence. We hypothesize that the collaborative nature of the gameplay better enabled this persistence.

5.2.1 Illustrative Example: Broadcasting

To better understand these results, it is useful to examine an example of Level One content. As described above in the description of the Knowledge Assessment Data in Section 5.1, the pilot study revealed that the Level One content most challenging to students involved the concept of broadcasting. We assessed students' understanding of this concept in Question 4 on the knowledge assessment. As seen in Table 5-3, 60 percent of students answered this item correctly on the Pretest, indicating that the overall population of students for this classroom deployment came in with relatively high prior knowledge of this concept, perhaps due to having been previously exposed to block-based programming. Similar to the overall pretest scores for the Level One content (as seen in Table 5-2), there was a slight gender difference for this particular item, with only 55% of female students answering this pretest item correctly, compared to 64% of male students.

Students in this classroom deployment took the interim posttest immediately after completing the same early segment of gameplay that students in the pilot study had played. As described above, this segment of gameplay included a programming challenge in which broadcasting is the primary component. The pilot study revealed that all students who played this segment individually mastered the broadcasting content, as evidenced by correctly answering the corresponding item on the posttest. Only 57% of those who played in pairs did so. In this full classroom deployment with paired gameplay, a similar 62% of students mastered this concept by the time they reached the interim posttest (which corresponds to the posttest in the pilot study).

After taking the interim posttest, students resumed playing the game. The subsequent game content included many more complex challenges that require a base understanding of broadcasting in order to solve. For example, in learning about how binary numbers can represent letters, students must navigate their in-game avatar to step on a sequence of binary numbers that get interpreted to form a textual message, which is then broadcast to an exit door to open it. A chief concern with using a paired gameplay approach is that a weaker partner may never master basic content, leading to the stronger partner dominating the learning experience. However, as shown in Table 5-3, students of both genders improved on the basic content as they advanced through the more complex challenges of the game. By the posttest, 80% of students correctly answered this item.

Table 5-3. Illustrative Example: Gender Comparison on Broadcasting Item
Percentage of Students who Correctly Answered Question 4 (See Figure 5-3)

	Female (n = 20)	Male (n = 25)	Total (n = 45)
Question 4 (Level 1 Item)			
Pretest	55%	64%	60%
Interim Posttest	50%	72%	62%
Posttest	75%	84%	80%

Table 5-3 illustrates that the gender disparity on the broadcasting test item widened from pretest to interim posttest. Whereas the population of male students does show some improvement on this content by the time they take the interim posttest, the rate of correct responses from female students stagnated.[9] However, while female students did not on average exhibit improvement on the concept of broadcasting early on (as evidenced by only 50% correctly answering the item on the interim posttest), 75% were able to correctly answer this item on the posttest. Section 5.3 discusses these results.

---

[9] These gender differences were not statistically significant.

5.2.2 Examination of Pretest Knowledge Differences between Partners

As briefly mentioned above, a major concern for this paired gameplay approach arises when considering pairs of students in which Student A has more prior knowledge than Student B. The potential exists for Student A to dominate the interaction, solving problems rapidly and leaving little opportunity for Student B to explore and learn. With this concern in mind, we sought to examine if and how differences in pretest knowledge between two partners correlated to learning. Unsurprisingly, we found a range of differences between partners. In only three pairs did both partners score exactly the same on the pretest, while the greatest difference between the pretest scores of two partners was .467. To aid this investigation, we defined a new variable, *Difference_Pretest* as (*students' pretest score – partner's pretest score*). Students who scored lower on the pretest than their partners thus have a negative *Difference_Pretest*, while those who scored higher than their partners have a positive *Difference_Pretest*.

A test for correlation between *Difference_Pretest* and learning gains found no statistically significant differences. Students therefore did not seem to be disadvantaged if their partners had more prior knowledge. In fact, they may have benefitted from having a stronger partner. Classifying each student as either *Negative Difference_Pretest* or *Nonnegative Difference_Pretest*, we found that *Negative Difference_Pretest* students achieved a superior learning gain of .247 (SD = .191) compared to their peers' .113 (SD = .113). A One-Way ANOVA found this result to be statistically significant ($F(1,39) = 4.714$, $p < .05$). It should be noted that a ceiling effect may limit the learning gains of some students in the *Nonnegative Difference_Pretest* category. Regardless, the positive learning gains of the students with less prior knowledge further support the paired gameplay approach, as it addresses the potential concern that such students will be left behind if their partners dominate the learning experience.

5.3 Discussion

Game-based learning environments, in practice, call upon students to master two forms of competency: competency in gameplay as well as competency in the given subject matter. For a well-designed game, students might not even consciously consider the effort they are devoting to mastering gameplay mechanics. Regardless of their prior game experience, a well-

designed game will in principle engage them enough so that the challenge of learning the game mechanics do not unduly hinder their learning experience. Yet, to some degree, students who lack prior game experience will inevitably face greater cognitive load when starting a new game (Frankosky et al., 2015). This is a critical issue for designers of games for computer science education, as female students may be less likely to have prior experience with analogous games.

In ENGAGE, the broadcasting example illustrates how this issue can manifest itself. The early segment of gameplay available to the students in the pilot study includes a challenge that explicitly teaches the concept of broadcasting. As described above, this challenge requires the student to broadcast the message "open" from a pressure pad to the exit door. To accomplish this, the student must also use a crane device to move a box onto the pressure pad. For students who have prior experience with similar games, the mere sight of the crane device, box and pressure pad informs their next actions. They immediately know to use the crane device to move the box onto the pressure pad, without even reading the read-only program for the exit door (the ultimate objective being to open that exit door). Having accomplished that subtask, they can then turn their full attention to interpreting the read-only program of the door. Students without prior gaming experience lack this advantage. They have to put more cognitive effort into the gameplay mechanics, taking some of their resources away from the computational thinking.

A paired gameplay approach can perhaps address this challenge, with potential benefits for the novice regardless of the partner's experience with analogous games. When novices have an experienced partner, they can learn from that partner's prior gaming experience. If their partner is a fellow novice, they can learn from watching their partner explore the gameplay in ways they might not otherwise have considered. Additionally, when both partners have low gaming expertise, neither is likely to jump ahead to the computational thinking while the other remains confused about some gaming mechanic. This benefit that collaboration brings to mastering the gameplay mechanics comes in addition to the potential benefits that paired students can gain in learning the subject domain.

On the other hand, negative outcomes might arise depending on the partnership, as in any collaborative experience. Having a partner with more game experience can potentially

result in a novice not participating (nor cognitively engaging) as much in the learning experience. In this case, even if the two have similar levels of competency in the subject domain, the novice game player might defer to the experienced game player. Yet pairing two novice game players together might result in increased frustration if neither student is able to master some game mechanic, leading both students to disengage with the computational thinking aspect of the learning experience. Bringing a collaborative approach to game-based learning environments for computer science thus presents a paradox familiar to researchers who study pair programming: how do we create optimal student pairings? As with many issues in assessing collaborative learning (Shah et al., 2014), this becomes a complex issue ripe for future research.

Limitations. The full classroom deployment (described in Section 5.2) was not an experiment that contrasted a single-player condition to the paired gameplay. The results therefore do not identify the extent to which collaboration impacted the success of the intervention. In fact, a series of pilot studies (beyond the scope of this work) have shown that all students—female and male—can benefit from playing ENGAGE individually. Rather than establishing the superiority of any one approach to another, the results reported here lay the groundwork for further study of collaboration in game-based learning environments for computer science education.

5.4 Conclusion

The studies reported here highlight the potential of paired gameplay for improving the gender equity of game-based learning environments. From observing students who played the educational game in pairs and those who played it individually, we noted several beneficial aspects of pair gameplay, as well as one or two caveats. When students play in pairs, they can provide each other various types of support, although the quality of this support of course depends somewhat on the two individuals. While the learning benefits of collaboration (as measured by a knowledge assessment) might not manifest in the initial session of gameplay, we saw collaborative gameplay lead to equitable learning gains as students continued playing the game over time. Indeed, whereas female students (and students with less gaming experience) used certain key computational thinking strategies less often than their peers

during the introductory level of the game, we observed that collaboration led to a sharing of best practices as time went on. Through this "diffusion of innovation" (Murphy-Hill & Murphy, 2011) students achieved significant learning gains regardless of their gender or their prior gaming experiences.

Future work should investigate how the combination of educational games and collaboration affects students of other underrepresented groups. Although the classroom deployment had a diverse pool of student participants, this work has not yet examined differences based on race or ethnicity, for example. It is important to do so to improve our understanding of how to create game-based learning environments that are equitable for all learners. Additionally, we came away from this early analysis of ENGAGE knowing that future work should also investigate collaboration in games at a finer granularity by looking at game-trace data and multimodal data. This will provide deeper insight into the nature of collaboration and what collaborative strategies lead to equitable learning gains for both partners in a paired gameplay scenario. What is more, the analyses in this chapter motivated future work on exploring how pedagogical agents can be integrated into human-human collaboration within virtual learning environments in order to even more fully support a diverse ranger of learners with different needs.

Chapter 6   GENDER EQUITY AND VIRTUAL LEARNING COMPANIONS: A PILOT STUDY

Game-based learning environments aim to engage learners through play, and their significant promise lies in their potential to engage a broad population of students (Isbister et al., 2010; Sabourin & Lester, 2014). As described in greater detail in Chapter 2, research has established that they can support student learning (Johnson, 2010; J. M. Kim, Hill, et al., 2009; Rowe, Shores, Mott, & Lester, 2011), particularly when students interact with a given system over multiple sessions (Jackson & McNamara, 2013). Much ongoing research in this field now focuses on how to improve game-based learning environments. How, for instance, can designers align the subject matter content with core gameplay (Hall, Wyeth, & Johnson, 2014)? How can designers change the incentive structures within a game-based learning environment to promote a growth mindset (O'Rourke, Haimovitz, Ballweber, Dweck, & Popović, 2014)? What is the appropriate level of challenge for producing optimal learning experiences (Lomas, Patel, Forlizzi, & Koedinger, 2013)? How can we design a game-based learning environment so that students' performance within it translates into a measure of learning?

For this last research question, a key challenge emerges: game scores are influenced by individual differences in students (Harteveld & Sutherland, 2015). Relatedly, students who have low expectations upon beginning to interact with a game-based learning environment tend to learn less (Snow, Jackson, Varner, & McNamara, 2013). Designers of game-based learning environments therefore must take special care to support students with a great diversity of characteristics or preferences. As mentioned in previous chapters, this challenge is particularly relevant for creating game-based learning environments that support female students as well as male students. Female students have been shown in some studies to suffer a disadvantage in game-based learning environments (Cassell & Jenkins, 1998; De Jean, Upitis, Koch, & Young, 1999), perhaps due to their greater familiarity with gameplay mechanics (Rowe et al., 2011). Although some more recent studies have demonstrated that female and male students can achieve similar learning gains in certain game-based learning

environments (Lester et al., 2014; Papastergiou, 2009), little research has been done on how to deliver equitably positive impacts on female students' attitudes and motivation toward the subject of study.

In this chapter, I examine this issue using the metric of student *frustration*, which has been shown to have significant impact on learning and on students' interest in continuing to learn a subject matter (Baker, D'Mello, Rodrigo, & Graesser, 2010; McQuiggan, Lee, & Lester, 2007). My early analysis of ENGAGE gameplay and post-surveys uncovered a gender gap in the game-based learning environment: female middle school students exhibited more frustration from the beginning of gameplay. This gender gap emerged despite persistent effort by the R&D team (with extensive focus grouping and iterative design and development) to design a game that would particularly appeal to female students. We first addressed this problem by refining the game environment and character controls to remove elements that female students found frustrating. This approach succeeded at reducing frustration overall, but failed to close the gender gap.

We subsequently focused instead on *enhancing* the game elements that female students found *engaging*. Based on the finding that female students exhibited significantly more interest in interacting with narrative elements of the game, we designed an interactive virtual learning companion to further enhance the narrative. As described in Chapter 2, a virtual learning companion is a socially-oriented virtual agent that plays a non-authoritative role within an interactive learning environment (Chou et al., 2003). The results show that female students' strong interest in the narrative elements of the game translated to high levels of engagement with the virtual learning companion. Most importantly, the gender gap in frustration completely disappeared. The results highlight the importance of enhancing game-based learning environments with varied forms of interactivity, and specifically demonstrate the potential of interactive virtual learning companions to support gender equity in game-based learning environments.

6.1 Preliminary Analysis of the Narrative-Centered Learning Environment

We worked with three diverse middle schools to provide game-based computer science education experiences during the 2014-15 academic year. This chapter reports on the classroom deployment at one of these schools, Carnage Middle School, in which we consider the 67 students who played the game and consented for their data to be analyzed: 42 male students and 25 female students. Students played the game over multiple days. This chapter focuses on the initial two sessions of gameplay, when students first acclimate themselves to the game-based learning environment and form their initial impressions of the learning experience.

Figure 6-1. Examples of narrative statements in ENGAGE
The game provides narrative statements throughout the playing experience in the form of pop-up messages from characters in the game's overarching narrative.

The overall game experience positively engaged students during these two days, as measured by post-surveys and qualitative observation. When asked to agree/disagree with the statement, "The gaming experience was fun" on a 5-point Likert scale, students responded with an average of 4.313 (SD = .7221) for Session 1 and 4.0 (SD = .9045) for Session 2. Initial studies provided promising results about the game's effectiveness for supporting students who enter with varying prior experiences and abilities (Frankosky et al., 2015).

Yet, gender differences were found with respect to frustration (Buffum, Boyer, et al., 2015). Female students reported greater levels of frustration than male students on the daily post-survey for both Session 1 and Session 2 (see Figure 6-2 below). Field observations suggest

that the frustration during these initial sessions stemmed mostly from certain types of game mechanics. Specifically, the students who expressed High Frustration[10] during Session 1 seemed to particularly struggle with character controls. We consider this type of frustration to be a form of extraneous cognitive load (Paas, Renkl, & Sweller, 2003), as it does not relate directly to the learning tasks. Subsequent levels of the game featured challenges of increasing complexity, and we consequently observed a general increase in frustration there. For those more complex challenges, we believe that that rise in frustration stems primarily from the increased cognitive interactivity and thus may be a more beneficial form of frustration (Baker et al., 2010).

6.2 Iterative Refinement and a Persisting Gender Gap

As the 2014-15 academic year progressed, we iteratively refined the character controls and the game-based learning environment. For example, safety railings were added to areas where students—particularly female students—frequently fell into water pits, such as the Tri-Level Room (refer back to Figures 5-1 and 5-2 in Section 5.1). These refinements led to an updated version of the game being deployed in the middle of the academic year. I will refer to this new version as *Game-Refined*. As described below, these refinements succeeded in making the game experience more successful for students overall. However, the refinements failed to close the gender gap in frustration compared to the prior version, which I will refer to as *Game-Original*.

In *Game-Original,* students fell in the water an average of 8.1 times during the first two sessions of gameplay. On average, female students fell in 11.688 times (SD = 7.5518, n = 16), substantially more than male students, who fell in 6.275 times (SD = 6.3842, n = 32). A one-way ANOVA found this gender difference to be statistically significant ($F(1, 46) = 6.535$, $p < .05$). Perhaps not surprisingly, students who reported High Frustration on the post-survey for one or both of these sessions also suffered more falls, 10.8 (SD = 8.1453, n = 24), than their

---

[10] Throughout this chapter, I report on frustration as assessed by student responses to the 5-point Likert-scale post-survey item, "I felt frustrated while playing the game." I operationally define "High Frustration" as a response of 4 ("agree") or 5 ("strongly agree") on this item.

less frustrated peers, who fell in 5.5 times (SD = 4.9255, n = 24). A one-way ANOVA found this also to be statistically significant (F(1, 46) = 7.417, p < .01).

In *Game-Refined*, students did indeed fall in fewer times. Compared to the 8.1 times that students fell in the water for *Game-Original*, students now only fell in the water an average of 1.8 times. Moreover, the female students who played *Game-Refined* actually suffered fewer falls in the water (1.3) than their male peers (2.3). Overall, students who played *Game-Refined* also reported less frustration on the Session 2 post-survey (2.737 on a 5-point Likert scale) than students who had played Game-Original (3.391). A one-way ANOVA found this difference to be statistically significant (F(1, 63) = 4.287, *p* < .05). These results at first glance seem promising.

Yet closer inspection of the results revealed that we had not actually bridged the gender gap in reported frustration; we had merely prevented it from widening. Female students who played *Game-Original* reported a significantly greater amount of frustration than male students after Session 1, and then the frustration for all students rose in Session 2 (during which students typically encountered the Tri-Level Room, which features many water obstacles[11]). The left graph in Figure 6-2 illustrates the persistence of the gender gap for *Game-Original*. The right graph shows that, although frustration did not rise as precipitously for *Game-Refined*, the gender gap remained.
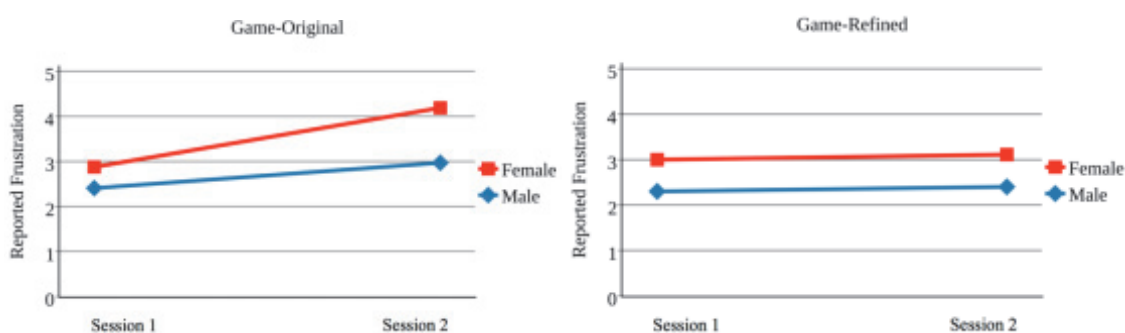
Figure 6-2. Male and female frustration on post-surveys for Session 1 and Session 2

---

[11] For an illustration of this room, refer back to Figure 5-1 in Chapter 5.

(5-point Likert scale)

In all versions of ENGAGE without the addition of virtual learning companions, we see a gender gap in both the first two sessions of gameplay. Aggregated together, female students reported an average Frustration of 2.91 in Session 1, compared to an average of 2.04 by their male classmates. We ran a One-Way ANOVA, and the results showed a statistically significant difference by gender for Session 1 ($F(1, 48) = 7.53$, $p < .01$). Similarly in Session 2, female students reported an average Frustration of 3.48, compared to 2.63 for male students ($F(1, 48) = 5.96$, $p < .05$).

6.3 The Appeal of Narrative

Ultimately, we seek to create game-based learning environments that are equally effective for all students. The failure of the iterative refinements to reduce the gender gap perhaps points to the inadequacy of merely focusing on the hindrances that female students faced. In contrast, an alternative would be to enhance some aspect of the game that particularly appeals to female students. With this in mind, the narrative component of the game drew our attention.

In designing the game, we hypothesized that the narrative would prove particularly effective at engaging female students. The results noted above (greater frustration for female students) indicate that the initial version of the narrative elements may have been insufficient. This is not to say, however, that these narrative elements lack promise. During classroom visits, we observed the benefits of these narrative elements, particularly for female students, who paid much closer attention to reading the narrative than their male counterparts. Not surprisingly, some students ignored reading the narrative elements altogether. However, these students were often those who most quickly mastered the character controls, perhaps from prior experience with similar games. Based on these classroom observations, I hypothesized that the narrative elements were particularly engaging for students who, in Session 1, reported frustration stemming mostly from character controls.

To explore this hypothesis, I examined how long students spent reading the narrative pop-up messages. In designing the game, we divided the narrative elements into short messages no longer than 15-25 words each, with the intention that students would spend a few seconds reading each one. When a student has finished reading a message, she clicks to hide it and then

resumes gameplay. As Figure 6-3 illustrates, female students spent more time (average 3.85 seconds) on the narrative than male students (average 2.84 seconds). A one-way ANOVA found this gender difference to be statistically significant ($F(1, 65) = 6.410$, $p < .05$).
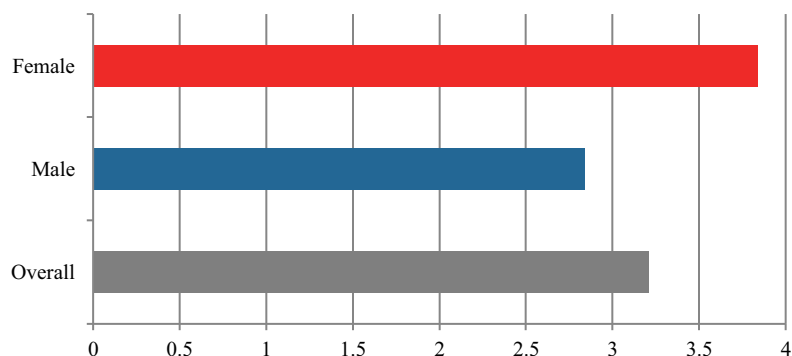
Figure 6-3. Average time (in seconds) students spent reading narrative statements

Based on our classroom observations, I hypothesized that this gender difference was not directly tied to the gender difference in reported frustration. That is, we observed no instances of narrative statements overtly frustrating students, nor did we see evidence of frustrated students avoiding the rest of the game experience by spending increased time on reading the narrative statements. Rather, we repeatedly observed certain students (particularly females) who seemed to be especially engaged with the narrative early in the gameplay, and who then later experienced frustration with other gameplay mechanics, particularly character controls. This phenomenon may be due to students having differing game-style preferences: some students may prefer an emphasis on narrative, while others prefer an emphasis on physical interactivity.[12]

6.4 Development of Virtual Learning Companion

Having observed these player differences, I decided to develop an interactive virtual learning companion to augment the non-interactive game narrative mechanism. I designed the virtual

---

[12] I ran several analyses to determine if there was any relationship between *prior experience playing computer games* and *reported frustration*, but found no significant results. The survey item for gameplay experience did not attempt to differentiate different styles of computer games that students typically play or prefer.

learning companion's interactive dialogue to deliver supplemental narrative that provides further insight into the game's story. The virtual learning companion's character was also designed to provide empathetic support to the student. The existing unaltered narrative also remained in the game.

Figure 6-4. Adriana the virtual learning companion

I used Unity 3D to develop a prototype of the animated virtual learning companion, a six-year-old girl named Adriana (Figure 6-4). In crafting Adriana's utterances, I aimed to design empathetic dialogue that would parallel the behaviors and resulting emotions that students might experience at this stage of the gameplay. The work was furthermore guided by research on the value of virtual agents telling autobiographical stories (Bickmore, Schulman, & Yin, 2009; Gulz et al., 2011). The learning companion's age was selected to leverage some of the established motivational benefits of situating virtual learning companions as non-authority figures (Chase et al., 2009; Woolf et al., 2010). Adriana tells the learner about her older sister, an expert computer scientist whom the player will meet later in the game. She also tells the learner about how she herself has visited the underwater research station and how she fell in the water many times. Adriana reassures the player that falling in the water in these

challenging rooms is not abnormal, and that moving around the station will become easier with experience. Students respond to Adriana with typed, natural language replies. The interaction with Adriana is designed to last approximately five minutes. Figure 6-5 shows an excerpt of interaction with Adriana.

Adriana: "Um, falling in the water is not fun, especially when it happens over and over. Guess how many times I fell in. Do you know?"

Student: "I dunno, maybe 2 or 3 times?"

Adriana: "A million times. That's what [my sister] said – I can't count very well, but she said I fell in a million times. She says I'm clumsy. Do you think I'm clumsy?"

Student: "Maybe a little."

Adriana: "Haha! Anyways, don't feel bad if you fall in the water, too. It's only really a problem in the part of the station where you are. Other than that, how do you like it?"

Student: "It is very nice."

Figure 6-5. Excerpt of student interaction with the learning companion
Note: Student inputs for this early prototype were free-response natural language responses. Later versions of the virtual learning companions have menu-based responses (See Chapter 7 for more details)

6.5 Participants and Methods

After developing a prototype of the Adriana virtual learning companion, I conducted a study to test its effectiveness in reducing the frustration that students, particularly female students, experienced during early gameplay. Two seventh grade classes at Centennial Campus Magnet Middle School (CCMMS) in Raleigh, North Carolina played ENGAGE with the addition of the interactive virtual learning companion. CCMMS serves a diverse student population, with over 40% receiving free or reduced lunch (a metric in the United States public schools for low income families). Both classes were in the subject area of social studies and were taught by the same teacher during different periods of the school day.

A total of 39 students were enrolled in these two classes. Because the study spanned multiple days, some students who participated in Session 1 were absent during Session 2, so these students' data were excluded from this analysis. Three additional students' data were excluded due to a malfunction during the virtual learning companion interaction, which I discovered in the game trace data. I thus report results from the 25 students for whom we have full data and signed consent forms. Of those 25 students (ages 12-13), 12 were female and 13 were male. [13] There were 11 white students, 7 African-American students, 2 Latino/a students, 1 Asian student and three mixed or other students. Three of the students (two male and one female) were absent the first day, so I exclude their data from some (but not all) of the analyses below. I include their data only for the survey questions about the virtual learning companion, with whom they did interact.

Each participant attended two hour-long sessions, which were held on back-to-back days. In Session 1, students individually played the game up to a pre-defined stopping point. All students were able to reach this stopping point within fifty minutes of gameplay. Immediately upon reaching the stopping point, each student completed the Session 1 post-survey.

At the start of Session 2 on the second day, students first interacted with the Adriana virtual learning companion. As with the rest of the gameplay, each student interacted with her individually. I tested two versions of Adriana, one with menu-based inputs and one with typed, natural language inputs, and participants were evenly distributed across the two versions. These interactions lasted an average of 5 minutes, with the shortest lasting just under one minute and the longest lasting almost 10 minutes. The conversations were largely predefined, so each student received a similar progression of textual utterances from Adriana; as such, I aggregate all students. As students finished their interactions with Adriana, they then proceeded with gameplay in the learning environment. Students continued interacting with the game-based learning environment until five minutes before the end of the session (resulting in

---

[13] Eleven of these 25 students interacted with a monologue version of Adriana. I have aggregated the monologue and dialogue data after finding consistent survey results across the two conditions.

approximately 40-50 minutes of gameplay for Session 2). Students then completed the Session 2 post-survey. In addition to engagement questions from the Session 1 post-survey, this post-survey also included four items about the student's perception of Adriana.

6.6 Results

By introducing the virtual learning companion, I hypothesized that students who were most engaged by the narrative would particularly benefit from the enhanced emphasis on narrative, mitigating their frustration. The data support this hypothesis. Only two of the 13 female students (15%) (and none of the male students) reported High Frustration. This result contrasts with the 76% of female students who experienced High Frustration after playing *Game-Original* and the 55% of female students who experienced High Frustration after playing *Game-Refined*. Unlike in the classroom deployments with *Game-Original* and *Game-Refined*, the gender gap in reported frustration closed following the introduction of Adriana. The results in Figure 6-6, which can be contrasted with those in Figure 6-2 above, illustrate the success here. The gap in frustration was eliminated for girls after interacting with the virtual learning companion.
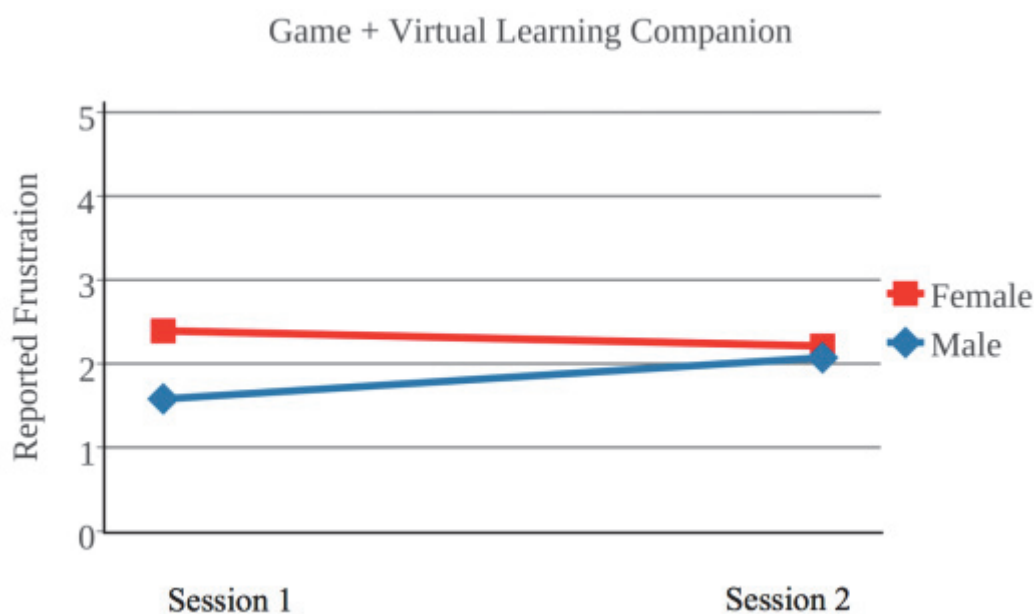
Figure 6-6. Male and female frustration on post-surveys after addition of the learning companion

(5-point Likert scale)

As with the classroom deployment of ENGAGE without a virtual learning companion, a gender gap was evident after Session 1 (before students interacted with the learning companion). Female students reported at this stage an average Frustration of 2.39 on a 5-point Likert Scale, compared to 1.58 for their male classmates. A One-Way ANOVA found this to be statistically significant ($F(1,23) = 6.585$, $p < .05$). However, this gender gap was no longer observed after students interacted with Adriana and then played more of the game (including the part that had previously been found particularly frustrating for students). On the post-survey for Session 2, female students reported Frustration at an average of 2.21, statistically equivalent to the 2.07 average for male students ($F(1, 26) = 0.09$, $p = .77$).

In order to better understand this success with female students, I examined the post-survey items that asked students about their interactions with Adriana. There were four such items in which students were asked to agree/disagree with statements on a 5-point Likert scale. Table 6-1 displays the results for these items. Female students generally responded with higher ratings on all of these items compared to their male classmates. One-way ANOVAs revealed statistically significant gender differences on all of these items at the $p < .05$ level. These results reinforce other findings on gender and animated virtual learning companions (Arroyo et al., 2011) that have shown female students respond favorably to this type of virtual agent.

Table 6-1. Survey responses on Adriana
(All female and male responses are significantly different, p<0.05.)

	Female ($n = 12$)	**Male ($n = 13$)**
I enjoyed interacting with Adriana.	4.08 (SD = 0.67)	3.15 (SD = 1.07)
I would enjoy interacting with Adriana again in the future.	4.08 (SD = 0.67)	3.23 (SD = 1.17)
Interacting with Adriana helped me to enjoy playing the game.	4.17 (SD = 0.83)	3.23 (SD = 1.24)
Interacting with Adriana helped me to feel less frustrated while playing the game.	3.67 (SD = 0.78)	2.77 (SD = 0.93)

The game data reveal that female students' especially positive reaction to Adriana corresponds with greater engagement with the narrative. Like the earlier studies of the game-based learning environment (without the virtual learning companion), the results show a statistically significant difference in the amount of time female students spent reading the narrative statements within the game compared to male students (F(1, 21) = 5.691), $p < .05$). In addition to spending more time on those narrative statements, female students also spent more time conversing with Adriana (Figure 6-7). I normalized the conversation times by subtracting the mean to account for the two different versions of Adriana (the one with menu-based inputs resulted in shorter interactions than the one with typed, natural language inputs). A one-way ANOVA on the normalized times then found a statistically significant difference (F(1, 21) = 8.421, $p < .01$).
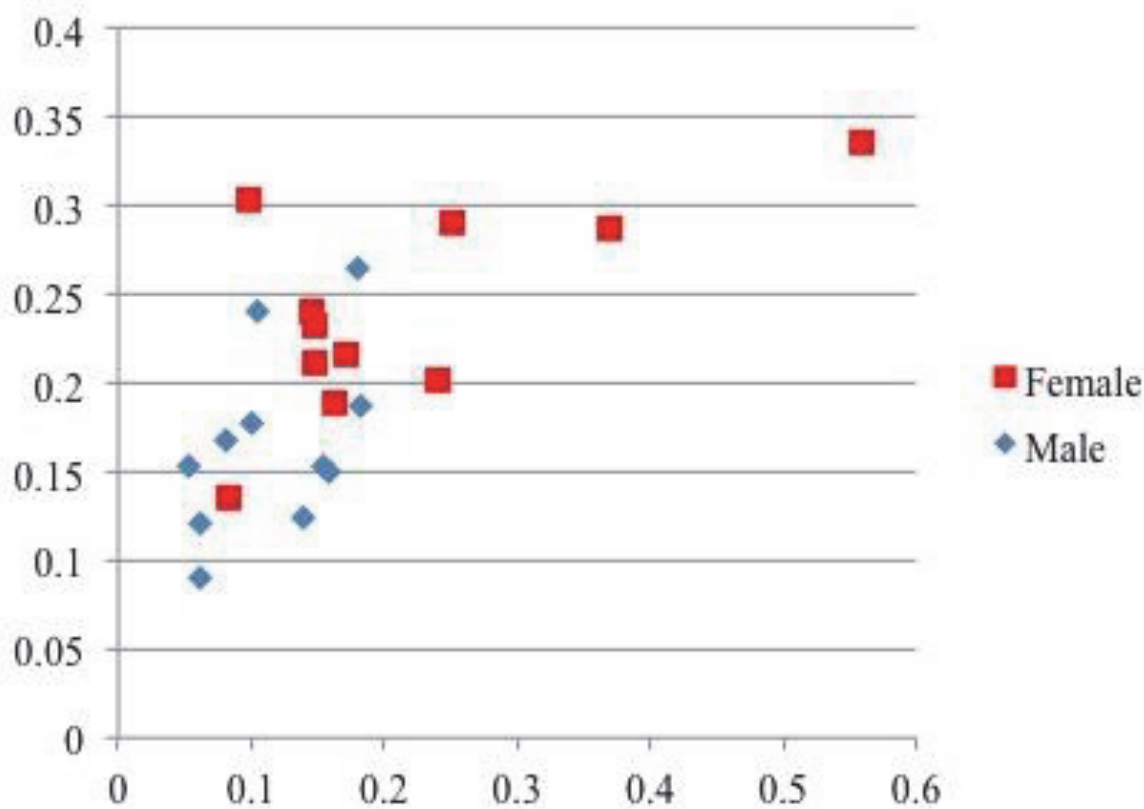
Figure 6-7. Scatter plot of *time on narrative in game* X *time interacting with the learning companion*

x-axis: length of time students spent on narrative elements in game
y-axis: length of time interacting with Adriana

Among those students who interacted with Adriana via typed, natural language inputs, the inputs of female students further illustrate their higher level of engagement. For the 12 dialogue turns at the core of the conversation, students could provide minimal responses (empty or a single word), or they could provide longer responses. Students who "bought into" the narrative, however, often wrote longer responses. To examine this phenomenon, I analyzed the core of the conversation, omitting the initial statements when Adriana asks the student to introduce him/herself, and the ending statements, when Adriana poses an open-ended question. I then classified student responses into one of three categories: *extended*, *basic (multiple words)* and *basic (one word)*. Extended responses involve a sentiment that goes beyond a simple "yes"/"no"/"I don't know" type of answer to Adriana's question. For example, an extended response might provide additional autobiographical facts about the student (*e.g.*, "no, but i do have a younger sibliling") or affective expressions (*e.g.*, "Yeah it was hard"). Basic responses merely answer the question, either using multiple words (*e.g.*, "Yes i did") or a single word (*e.g.*, "Yes").

Table 6-2 contrasts a subset of the conversations that a female student and a male student had with Adriana. The durations of these conversations were both near the medians for their respective genders. The male student tersely answers the question with an affirmative or negative response, occasionally using different terms other than "yes" and "no." The female student, however, intersperses standard one-word answers with more robust responses in which she freely shares personal feelings with the virtual learning companion. Both of these students reported that they enjoyed interacting with Adriana. The gender phenomenon holds true for nearly all the conversations we examined. After accounting for data loss due to system error (2 female and 1 male), there were conversations of 13 students: 5 female, 8 male. Female students averaged more extended responses (5.60 responses) than male students (1.88). A one-way ANOVA found this to be statistically significant ($F(1, 11) = 6.899$, $p < .05$). Relatedly, male students used more basic (one-word) responses (9.13) on average than female students (4.80). A one-way ANOVA found this also to be statistically significant ($F(1, 11) = 5.773$, $p < .05$).

Table 6-2. Subsets of interaction (1 female student example, 1 male student example)
Female students tended to use more words and spend more time conversing with the learning companion.

Female Student Inputs	Adriana Outputs	Male Student Inputs
	Um, okay. Hi...Well, my name is Adriana and I'm 6. I like talking to people and being silly. So I'm glad to meet you. Did you just recently get to the station?	
yea i need to look around		Yes
	Well, Delphina has lived there for over two years. She thinks I'm annoying. But I'm only 6! I think it's kind of my job to be annoying. Do you have any older siblings?	
yes		Yes
	Oh, cool! Yeah...You'll just love Delphina. You have to meet her. I think she's probably down on a lower level of the station than where you are now, though. Um...do you know where you are, exactly?	
no		No
	Um...I think maybe I might be able to guess where you are. Have you had to go on any moving platform thingy to cross water?	
yes a while ago but it took to long		Yes
	Yeah, I know where you are! I've been there when I visited Delphina at the station. I think you might see a couple of those thingies if you keep going further. Delphina's gotten really good at writing programs for them. Are you good at writing programs yet?	
ummm i just learned so not so good		Yeah
	You'll get good. I guess it takes some time to get used to it, but then it gets kind of fun. Well...Delphina says everyone in the station learns how to do some programming because it's needed for the ocean science they do. Do you think that's true?	
well its her op[in]ion but i guess		Yes
	Yeah, I guess...Actually, what I really want to be is a doctor and save people's lives, but Delphina says learning computer science will be good for that, too. Do you like helping people?	
yes i want to help people just not be a doctor		I guess
	Umm...okay...Do you want to hear a story?	
sure		Sure

Figure 6-8 depicts the proportion of response types for each of the 13 students, with each bar representing an individual student. All but one of the female students employed one-word responses for *no more than* 50% of their conversations, while all but one of the male students employed one-word responses for *over* 50% of their conversations. These two "exceptions" to the gender pattern are noteworthy, as they provide further support for the hypothesis that the Adriana conversations support students who are engaged by the narrative element (a construct which is not fully explained by gender). The female student was in the bottom quartile of time spent on narrative statements in the game environment. She also spent very little time in the Adriana conversation. The male student, meanwhile, was in the highest quartile for narrative engagement, and he was particularly engaged with the Adriana conversation compared to his male peers. Both of these students reported that they enjoyed interacting with Adriana. The exceptions to the gender patterns highlight the fact that a demographic division such as gender cannot capture the multi-dimensional individual differences that inevitably affect user interactions. As I discuss later, this observation points toward important future work.
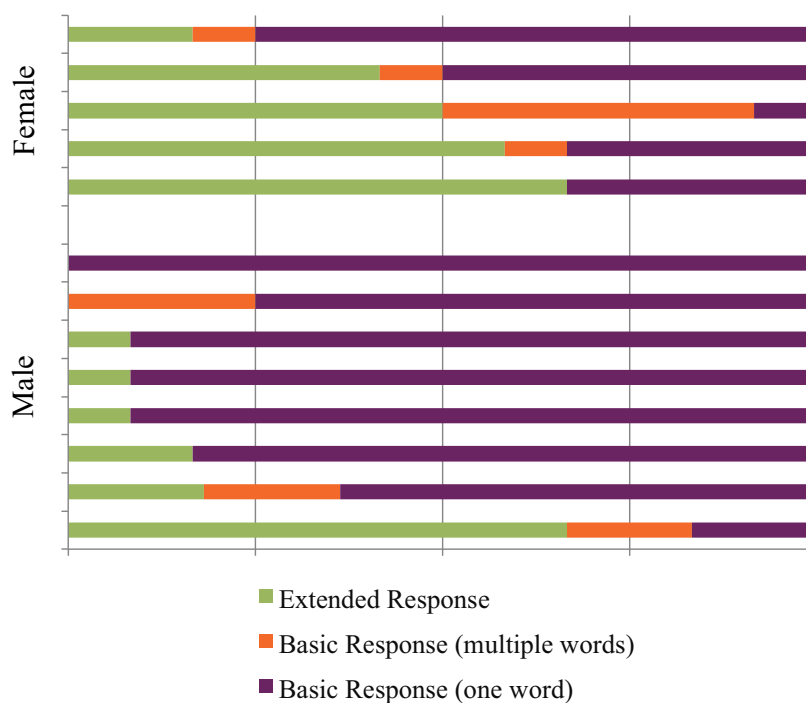
Figure 6-8. Gender differences in proportions of responses that are extended vs. basic
Each bar represents one student's interactions.

6.7 Discussion

The promise of game-based learning has captivated many who look at the popularity of commercial games among youth and envision schools harnessing that heightened engagement to enhance learning (Gee, 2003). This rising interest generates a pressing question: how do we create game-based learning environments that equitably serve all students? This chapter has demonstrated how *enhancing* specific aspects that most appeal to certain student subgroups can have a greater measurable impact on improving equity. As designers of digital learning environments, we must view students as individuals with varying preferences and strengths. This is not to say that we should ignore opportunities to remove unnecessarily frustrating game elements. As demonstrated by the success of *Game-Refined* (compared to *Game-Original*), doing so can reduce the frustration for all students and thus remains a valuable element of iterative refinement. Merely removing negative aspects, however, does a disservice to students who would benefit from different forms of interactivity.

6.7.1 Limitations

Although our initial series of studies with the game-based learning environment pointed to crucial design considerations for gender equity, it is important to note several limitations. First, the three versions of the game (*Game-Original, Game-Refined* and the version with the virtual learning companion) were not deployed simultaneously as part of a controlled experiment. Rather, the classroom deployment and pilot study were part of design-based research with iterative system development. Particularly noteworthy, two versions of the virtual learning companion were deployed, as part of an original study design to compare monologue with dialogue for the agent. However, because no significant differences emerged in self-reports or learning for students who used those two input methods, I aggregated the two in this analysis into the learning-companion condition. At a higher level, the results obtained here were in the context of a particular set of urban middle schools in the United States, and their generalizability to a broader population of students has not yet been established.

6.8 Conclusion

Creating equitable learning experiences for all students is an imperative that presents many challenges for designers of digital learning environments. For game-based learning, there is

evidence that female students can be particularly engaged with narrative features of a game. In this chapter, I have presented results on the ways in which virtual learning companions can enhance the narrative of game-based learning environments and achieve equity by supporting a wider array of student needs. More broadly, the results suggest that identifying types of interactivity that are especially beneficial for learner sub-groups holds great promise for supporting diverse learners.

This work points the way toward important future work in virtual learning companions to enhance equity in game-based learning. For example, there is great promise in building and investigating the benefits of learning companions that build rapport with students over multiple days or weeks. Research on relational agents (Bickmore & Picard, 2005) may guide this future line of inquiry. Additionally, future work should investigate the balance between enhancing narrative interactivity within a game and ensuring that the narrative reinforces, rather than distracts from, the learning objectives (Wouters et al., 2013). In addition to investigating this balance, we need to model learners at a finer-grained level than coarse demographics, though demographics are clearly an important part. Every student's learning is influenced by multi-dimensional individual differences that should be modeled in combination as adaptive game-based learning environments move forward. Finally, there is great promise in supporting collaboration within game-based learning (Kumar et al., 2010), and virtual learning companions may be well suited to foster collaboration skills. By exploring these and other phenomena in the context of game-based learning environments, the research community can identify ways to design game-based learning environments that equitably serve all students.

Chapter 7   ITERATIVE DEVELOPMENT OF THE VIRTUAL LEARNING COMPANIONS

The promising findings in Chapter 6 motivated further investigation into the design of virtual learning companions for ENGAGE. While the results of the pilot study suggested that Adriana improved the gender equity of the system, the study of that initial prototype set aside two key considerations. Firstly, students only interacted with Adriana once. As a game designed for 10-15 sessions of gameplay, ENGAGE ideally might have a virtual learning companion that interacts with students several times over the span of multiple sessions. Secondly, students interacted with Adriana individually. Considering the collaborative nature of the ENGAGE game, an ideal virtual learning companion would converse with pairs of students. This chapter describes the iterative development that followed from the study of the prototype virtual learning companion.

7.1   Prototyping with Multiple Sessions

I ran an after-school prototyping study in Spring 2015 to explore how to extend the virtual learning companion to support repeated interactions over multiple sessions. Creating such a virtual agent invariably presents a more complex challenge than a single-session virtual agent, but the narrative element of ENGAGE helps with this task. Timothy Bickmore's work on relational agents—virtual agents that maintain long-term relationships with users (Bickmore & Cassell, 2001)—can guide us. In particular, research has established the role that narrative can play in sustaining a user's engagement with a virtual agent over multiple sessions, guided by work in the field of interactive narrative (Bickmore, Schulman, et al., 2009).

Building on that research on relational agents, I created a series of three prototypes inspired by interactive narrative research, specifically the social interaction system Comme il Faut (CiF) (McCoy et al., 2014). CiF provides the basis for Prom Week, a game in which players manipulate the relationships between various virtual agents, thereby generating a playable story. Unlike the "god's eye" view of Prom Week, the prototype interactions for ENGAGE afforded the player a first-person, "face-to-face"
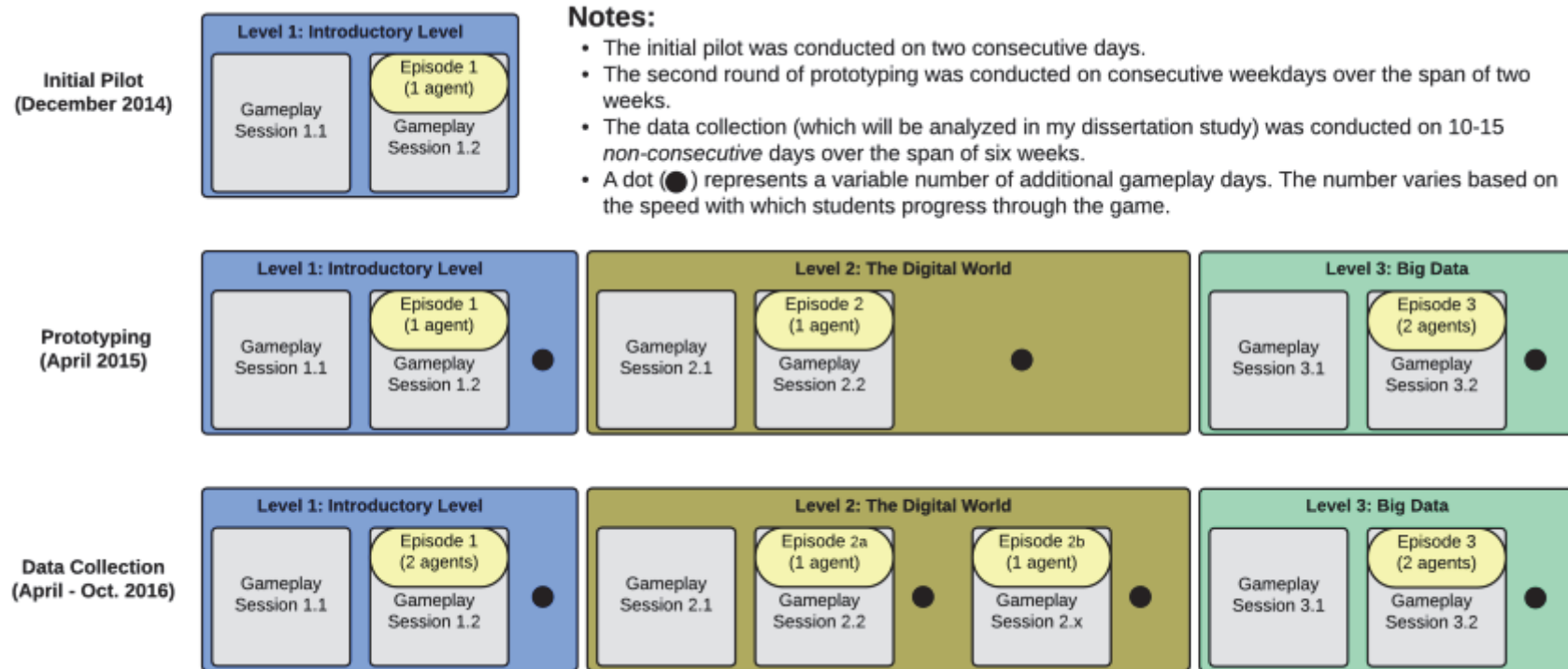
Figure 7-1. Comparison of study designs for the three data collections involving learning companions

interaction with a virtual learning companion. The interactions affected the learning companion's relationships with the player and with other characters in the story. In this way, I followed the lead of Anne Sullivan and colleagues (Sullivan, Grow, Mateas, & Wardrip-Fruin, 2012), who showed how CiF can be adapted to games in which the user takes on a role in the story (resulting in a new model that the authors named CiF-RPG). The complete CiF architecture includes several types of *relationships* that can develop between characters, several *traits* that a character might possess, several *statuses* that a character might have at any given moment, and a *cultural knowledgebase* that describes how characters can relate to various concepts and objects within the story world's cultural context. My system for these prototypes only included a subset of this architecture, focusing on one type of relationship and one or two other key elements (a status, for example) in each prototype interaction (see the prototype descriptions below for more specific details).

7.1.1 Prototype Descriptions

As in the work reported in Chapter 6, I developed these prototype episodes in Unity 3D and used many of the same art assets and scripts. Adriana's age (six years old) and role in the narrative (younger sister of Delphina, the station's computer scientist) also remained the same.

Episode 1. The first episode contains a simple interaction with the virtual agent Adriana. In this interaction, which students experienced on the first day of gameplay, Adriana narrates a linear story to the player, and the player simply clicks "next" to advance the narration. I used this episode as a baseline to gain an understanding of how users would react to the increased narrative interactivity of the later prototypes. Although the lack of choice might seem to make this episode overly simple, studies have shown that the mere presence of an animated virtual agent can have a strong effect on how middle school students feel about a game-based learning environment (Lester et al., 1997). I therefore considered it necessary to include this baseline interaction. (I do not include a system diagram for this prototype because of its simplicity).

Episode 2. I designed the second episode to occur at a specific location later in the ENGAGE gameplay (for players to encounter on approximately the third day, depending on how quickly they advanced). In this interaction, the player has a goal: to gain Adriana's trust so that

Adriana will divulge a secret piece of information. As with Episode 1, the interaction still comes in the form of a conversation that the player has with Adriana, but now the player has three options for each dialogue turn, with each option representing a different type of social move (*e.g.*, demand information, harass Adriana, flatter Adriana, etc.). Depending on the player's social move, Adriana's trust levels may increase or decrease for the player or other characters in the game. Following the CiF model (in a limited form), this prototype focuses on one type of relationship (*x-trusts-y*) that characters can have, and also includes one status (*happy*). Adriana's status is fixed at a high level of happiness, but the trusting relationship she has towards the player and towards another character (Carl Bishop, the case handler who deployed the students on their underwater mission) varies based on the player's actions. Figure 7-2 shows the system diagram for the conversation flow in Episode 2.
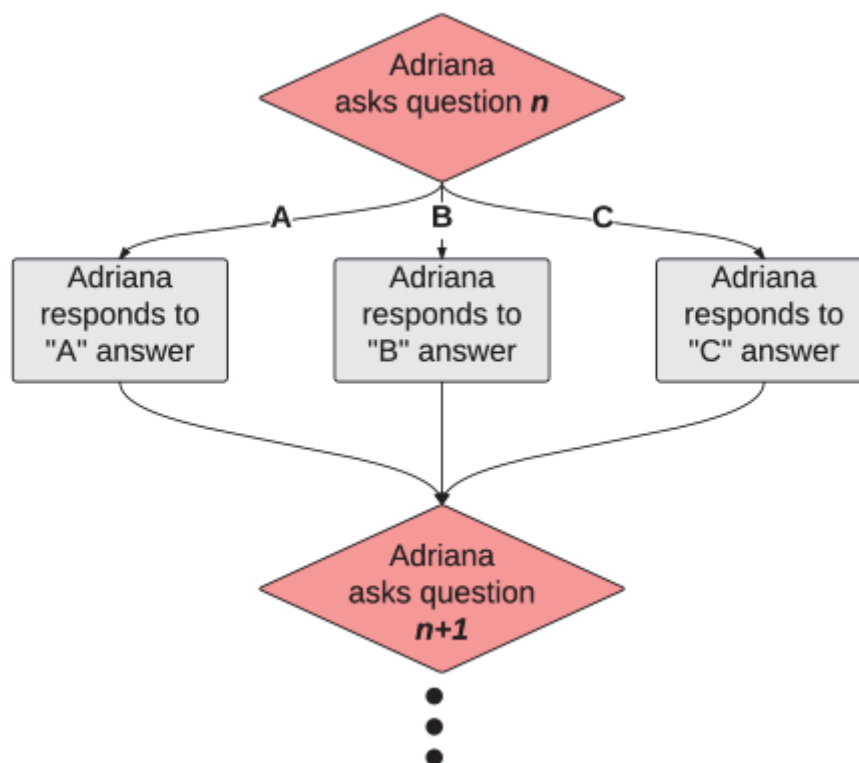
Figure 7-2. System diagram for Prototype Episode 2

Episode 3. Episode 3 adds a second animated virtual agent, to see how students react to interactivity with multiple characters. I designed this third prototype to occur at the end of gameplay. As with Episode 2, the user has three options for advancing a conversation, and the user's choice affects the relationships between characters. In this episode, however, the user can switch back and forth between interacting with Adriana and interacting with a new character named Johnny. The goal in this interaction is for the player to improve the relationship (*x-likes-y*) between Adriana and Johnny, and to improve Johnny's perception of a certain concept from the cultural knowledgebase (*computer science*). Figure 7-3 shows the system diagram for Episode 3 and Figure 7-4 shows a screenshot of the two virtual learning companions.
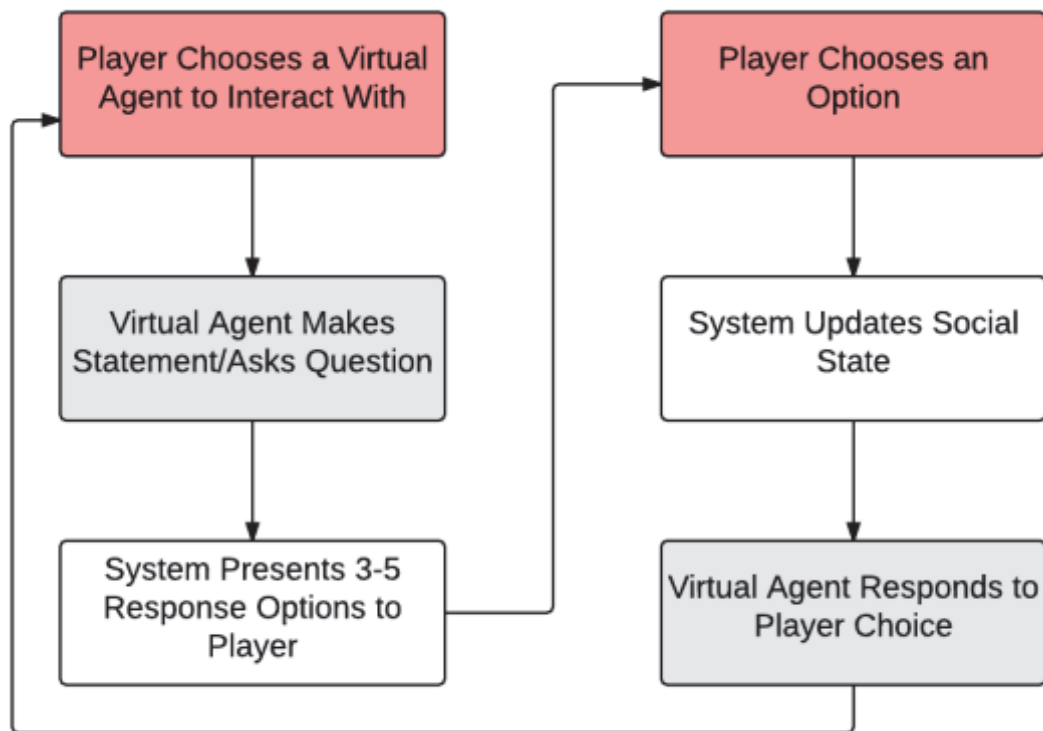
Figure 7-3. System diagram for Prototype Episode 3

Figure 7-4. Screenshot of Adriana & Johnny in Prototype Episode 3

7.1.2 Student Survey Responses to the Prototypes

I conducted the prototyping study during an after-school program at Centennial Campus Magnet Middle School in Raleigh, North Carolina. A total of 18 middle school students ended up participating in this after-school activity: 10 male students and 8 female students. Of students who reported a race or ethnicity, there were 11 white students, 3 African-American students, 2 Asian students and 2 Latino students. The students played the ENGAGE game over the course of a couple weeks, with each day's session lasting approximately two hours. Unlike in most classroom studies of ENGAGE, in which pairs of students play the game in two-player mode, the students in this prototyping study played ENGAGE in single-player mode. A short engagement survey was administered at the end of each day's session. For the sessions in which students encountered with the three episodes, this short engagement survey was supplemented with questions specifically asking about the students' reaction to the episode.

I was primarily interested in how the varying levels of interactivity would affect participant's reaction to Adriana (and Johnny). To measure this, the post-survey included five relevant questions that participants responded to on a 5-point Likert scale. Students took this

post-survey at the end of each session in which they experienced one of the episodes. Figure 7-5 visualizes the responses to these five questions.
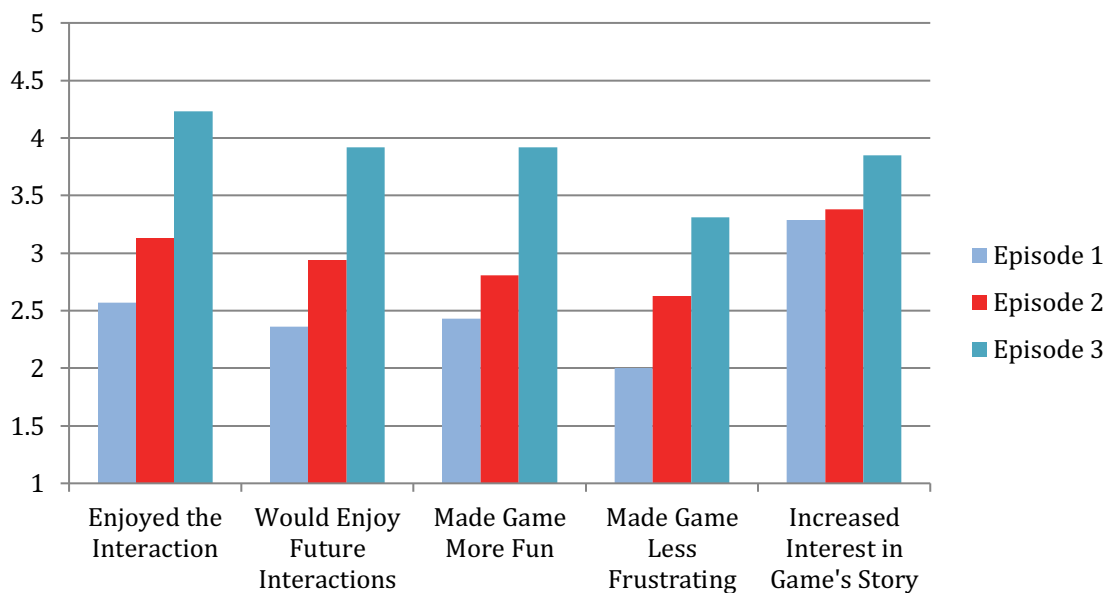
Figure 7-5. Student survey responses about the virtual learning companions
(Based on a 5-point Likert scale)

As the comparison bar graph shows, the participants generally responded more favorably to the second episode than they did to the first, and then responded even more favorably to Episode 3. The sample size is small, so we will refrain from making any definitive conclusions. However, these preliminary results show that the approach has promise. Although the authoring burden for all three episodes was roughly the same (the conversations in each episode were all non-branching), the increase in perceived narrative interactivity greatly improved participants' reaction to the virtual agents. Going forward, I hoped to build upon the success of Episode 3 by including more opportunities for users to interact with multiple characters and see their interactions affect the relationships between those characters. The final survey students took included four additional questions regarding students' perception of the narrative in the game. Figure 7-6 shows the results of these questions, which support the approach.

Final Survey
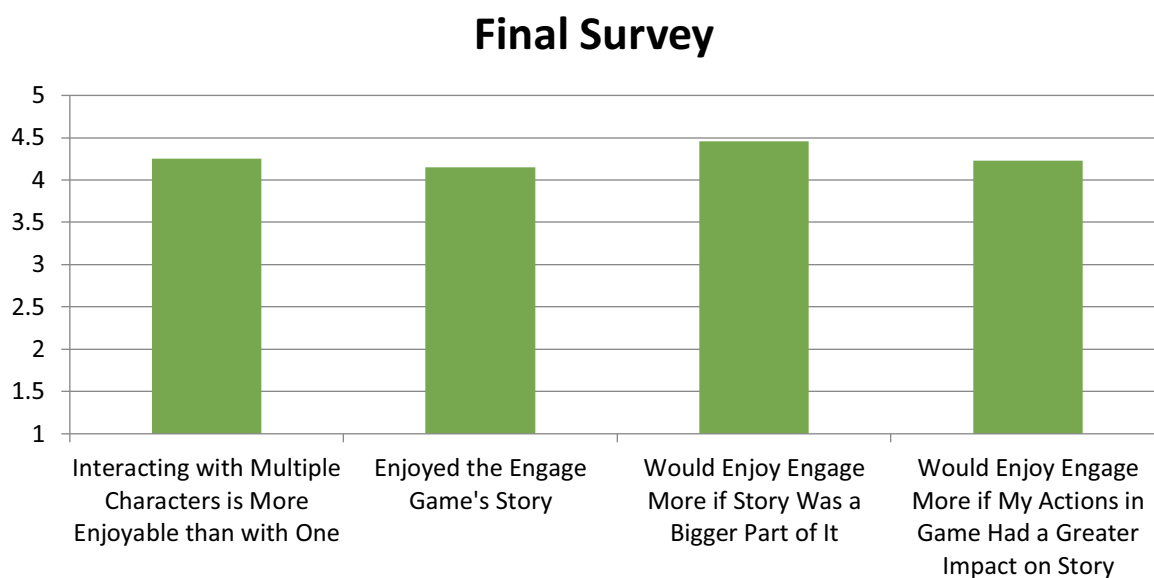
Figure 7-6. Survey responses support providing more narrative interactivity
(Based on a 5-point Likert scale)

7.1.3 Student Free-Response Feedback

The small number of students in this pilot study dissuades further quantitative analyses, such as the gender comparisons of Chapter 6. Instead, we can get insight from students' post-game feedback. After students completed the entire ENGAGE game (including all three episodes with the virtual learning companions), they individually completed a questionnaire. The questionnaire included several questions about the virtual learning companions, and students responded with free-form text. Two considerations emerged as particularly significant. Firstly, many students recommended increasing Adriana's age (up from six years old). Secondly, students' receptivity to the different conversation topics (*e.g.*, station backstory, importance of computing, empathizing with frustration) varied based on the students' prior experience. The following paragraphs examine these two issues in greater detail.

Regarding the age of the virtual learning companions, students made several arguments for increasing the age. Students who had prior experience with programming and gaming expressed the greatest conviction in increasing the age to at least ten years old. Emily, a white female student, argued for this by saying, "I think that her age should be older because she tells

you what to do and in real life I would not talk to a six year old about a game." Zac, a white male student, also did not like that she was so young because "her age makes me have to be nicer towards her because younger people usually have more feelings than older people." Jay, a South Asian male student, explained that "if she was older she would probably have understood more responsibilities" and be able to provide more useful help and advice. Finally, a white male student named Josh said, "she's too young to be taken seriously, [it affects] how I interact with her because I would trust a more mature person like another scientist at the station."

As for students without prior experience programming and gaming, most also recommended increasing the age, although perhaps by a more modest amount. Two Latina female students, Mary and Jessica, both made arguments for increasing the age based on making the narrative more realistic. Mary explained, "I think that you should make her a little bit older like, 8 years old because, how can she memorize the place at 6 years old." Jessica added, "I feel as though Adriana should be eight because it doesn't seem that common a 6-year old girl would know that much information on computer science. It affects how we speak to her because she seems much more defensive about the secrets she knows." Hilary, a white female student, noted that "her age was fine, since she was not so young that she couldn't speak, but not too old that you felt you could be rude to her, so as to not get any info. I think she would be better as an eight year old." Only one participant, a white female student named Caitlin, advocated leaving the age at six: "She is fine being young I think that her age shows her importance."

Students' prior experience had greater bearing on how they valued the various conversation topics that the virtual learning companions presented to them. Students with more prior experience seemed only to value the narrative topics about the station's backstory. Zac, for instance, "liked the one about Delphina, and stories about the station" but thought that hearing about "writing programs was boring." Jay added that he, too, liked "the stories because she says what she does with Murdock, and if you get to a point where she could reveal the secret, she could be useful." In general, these experienced students likely did not need the potential benefits that a virtual learning companion can provide, so they expressed greater

criticism than their less experienced peers. Emily, the experienced female student, wrote that "I liked her telling the story and her sister but the rest did not really do anything for [me]." Thinking critically on how to improve the learning experience of ENGAGE, Emily advised that the virtual learning companions should ask students for help with understanding the programming challenges, "because if your goal is to teach kids computer science and programming they should be able to explain it too and that would be a fun way to do it."

The students with less prior experience also liked the conversation topics about the station's backstory, but their appreciation of the virtual learning companions generally went beyond just the narrative intrigue. Hilary responded that, of all the topics, she thought that Adriana's conversation about "writing programs was the best since it was useful." Mary also listed "learning computer science" as one of her favorite conversation topics with the virtual learning companions. Jessica liked many of the conversation topics, including those that concerned the particular game challenges for which Adriana provided advice. In general, Mary and Jessica, the two Latina students mentioned above, both expressed especially positive feelings about the virtual learning companions. Mary remarked that "Adriana was a really nice experience for me. She reminds me of me because she's so childish, and I still am childish." Similarly, Jessica said, "I think her personality is kind of like a breath of fresh air after the challenges we went through. Yes I like her personality because it is endearing personally to me. Maybe Adriana could be a smidge more serious."

That last point of Jessica's—that she might prefer if Adriana was slightly less silly—echoed in many of the students' reflections. Therefore, in conjunction with increasing the virtual learning companions' ages following this prototyping study, I decided also to tweak their personalities. Additionally, since students responded positively to the multiple virtual learning companions of Episode 3, I knew that such an approach held merit. As the next section will explain, the inclusion of multiple virtual learning companions dovetails with the goal of supporting collaborative learning in ENGAGE.

7.2 Final Design and Implementation

The student feedback provided the necessary insight into how to design the virtual learning companions that will be studied in the proposed research for this dissertation. Following the prototyping study, I created four playable experiences with the two virtual learning companions, Adriana and Johnny. The overall purpose of Adriana and Johnny is to improve the collaborative learning experience of pairs of students as they play ENGAGE in the two-player mode. Indeed, unlike both the pilot study (described in Chapter 6) and the prototyping study, my data collection sees the virtual learning companions added to a typical ENGAGE classroom deployment: part of a nine-week oceanography elective with students playing in pairs during designated gameplay sessions. For this data collection, the pairs of students together choose how to respond to Adriana and Johnny, as the virtual learning companions tell stories about their experience collaborating together. Adriana is a more experienced programmer than Johnny, and they both describe the challenges and benefits that can come from such a dynamic. I hypothesized that the four interactions with the virtual learning companions, spread out over several weeks, would lead to more equitable collaborations between student pairs.

The four episodes all share a similar overarching design. For each conversation turn, the virtual learning companion utters a statement or question. The students can then choose one of three responses. In most cases, the options include one positive response (friendly and perhaps flattering to the virtual agent), one negative response (abusive to the virtual agent) and one neutral response. After the students click their chosen response, the virtual learning companion utters an appropriate reaction. The virtual learning companion then utters its next pre-scripted statement or question. As with the episodes from the prototyping study, the dialogue is non-branching but designed to offer the appearance of natural conversation. The only stylistic difference among the four episodes is that, while the first and fourth include both virtual learning companions, the second interaction only features Johnny and the third interaction only features Adriana. This allowed the second and third interactions to take up less time while still achieving all the desired goals.

7.2.1 Episode 1: Middle of Level 1

Students first encounter Adriana and Johnny on the second day of gameplay, in the same game location where students in the pilot study encountered Adriana. On the preceding gameplay day, the students play the tutorial part of the game up to the challenging Tri-Level Room (Figure 7-7). To navigate through this room, students must write a sequence of programs for moving platforms, and avoid falling in the water. As described in Chapter 6, this can be a particularly challenging room for students; the pilot study showed that a virtual learning companion could help with the heightened frustration that comes with the challenge.
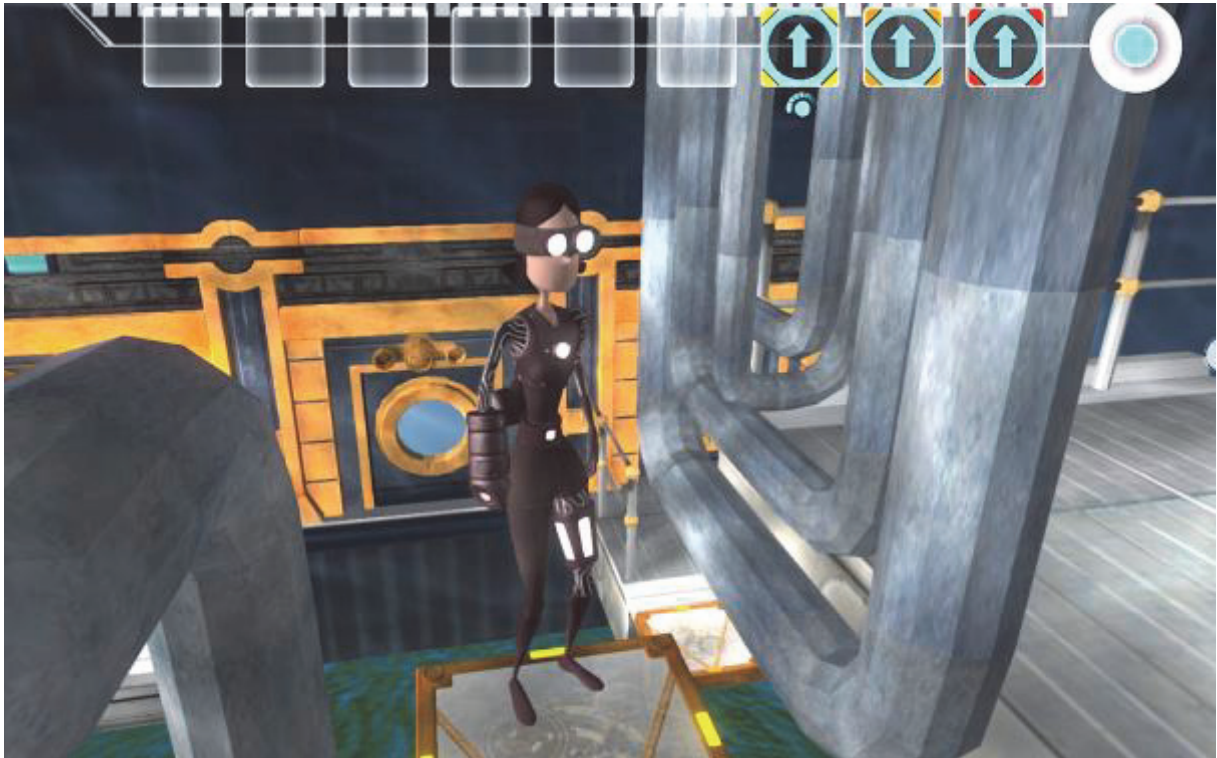
Figure 7-7. Screenshot of Tri-Level Room (Level 1)
In this room, the players must program a series of moving platforms to navigate their avatar around various obstacles, including pipes and water pits.

The second day of gameplay begins with the student pair having the conversation with Adriana and Johnny. The episode lasts approximately 8-10 minutes and then the pair spends the rest of the session (approximately 20-25 minutes) continuing Level 1 of ENGAGE. In the

conversation, the pair eventually experiences 9 conversation turns with each of Adriana and Johnny. They tell their two respective sides to the story about how they first pair-programmed together and empathize with the frustration that students might experience in the Tri-Level Room.

Figure 7-8. The virtual learning companions, Adriana (top) and Johnny (bottom)

7.2.2 Episode 2a: Early in Level 2

The student pairs experience their second interaction with the virtual learning companions approximately a week later. After completing Level 1, which students accomplish 1-2 days after that first interaction, they proceed to Level 2. Throughout ENGAGE's Level 2, students learn about binary numbers, and the first day introduces them to how sequences of bits can represent base-ten numbers. As with Level 1, the second day of Level 2 gameplay features increasingly complex challenges that can frustrate students, making it an appropriate location for inserting the affective scaffolding that the virtual learning companions can provide. Accordingly, on the second day, they start with the virtual learning companion interaction. In this second episode, Johnny talks to the students about the importance of being patient with your teammates, as well as the value of reading the programs thoroughly rather than doing a guess-and-check strategy. Johnny also provides some targeted advice for the upcoming rooms.
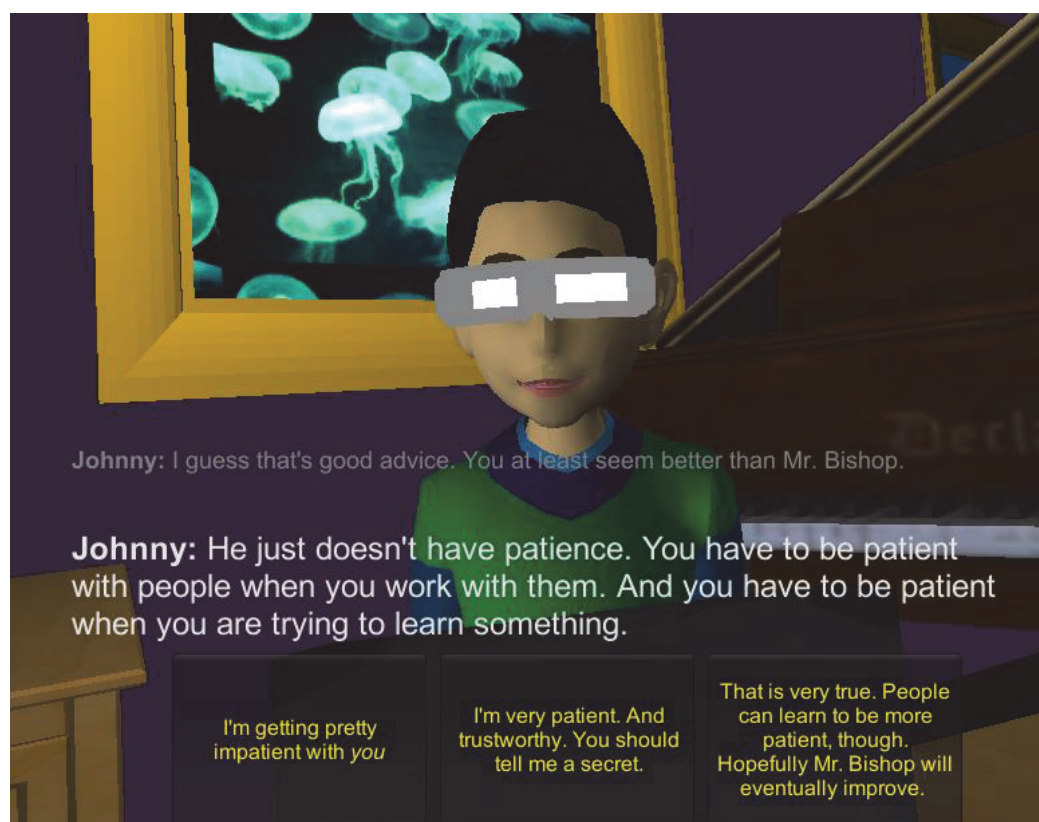
Figure 7-9. Screenshot of Johnny's conversation in Level 2

Immediately after their conversation with Johnny, the students begin gameplay in the Storage Room, where they learn about the computational concept, *variables*. The next room reinforces that concept. Previous observations of student gameplay have revealed that students often fail to learn the concept in this sequence of rooms, largely due to students employing the guess-and-check strategy rather than conscientiously reading the programs to understand them. I crafted Johnny's conversation specifically to dissuade students from using this harmful strategy.

Figure 7-10. Screenshot of Storage Room (Level 2)
This room contains 4 pressure pads distributed around the room. Students must navigate their avatar to step on the pressure pads to update the variable in the device's program.

7.2.3 Episode 2b: Late in Level 2

A couple of days later, the student pairs encounter the third episode (Episode 2b), this one with Adriana. In this conversation, Adriana reinforces the points that Johnny made (the importance of being patient with one's teammates and studying the programs rather than doing guess-and-

check). This episode with Adriana comes toward the end of Level 2; some students finish the level during the same session after conversing with Adriana, while other students may take an additional day to finish the level.
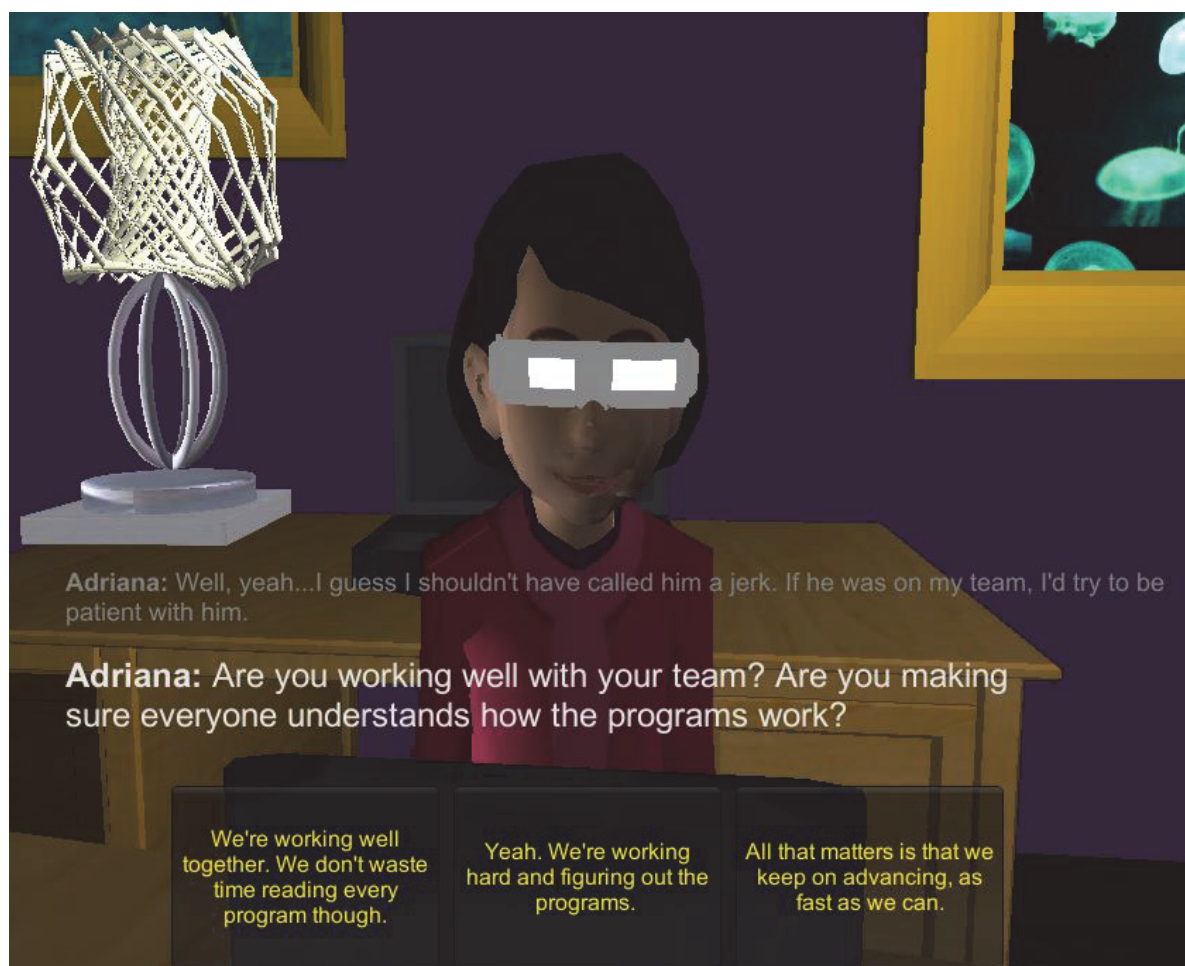
Figure 7-11. Screenshot of Adriana's conversation in Level 2

7.2.4 Episode 3: Middle of Level 3

The fourth and final episode with the virtual learning companions comes on the second day of playing ENGAGE's Level 3, in which students learn about "big" data. On the first day of this level, they write programs to visualize data and learn about metadata. The first day ends in a cliffhanger fashion with the students getting lured into a Trap Room. In order to escape this

Trap Room, students will need to write a bubble-sort program to reorder six large cans. Previous studies of ENGAGE revealed the Trap Room as the most time-consuming and potentially frustrating challenges in the entire game. Before attempting this challenge, the students begin the second day of Level 2 by having their conversation with Adriana and Johnny. The two virtual learning companions give specific advice on how to write the bubble-sort program. As with all their other conversations, Adriana and Johnny emphasize the importance of teamwork. After the conversation ends, students have approximately 25 minutes to play Level 3.
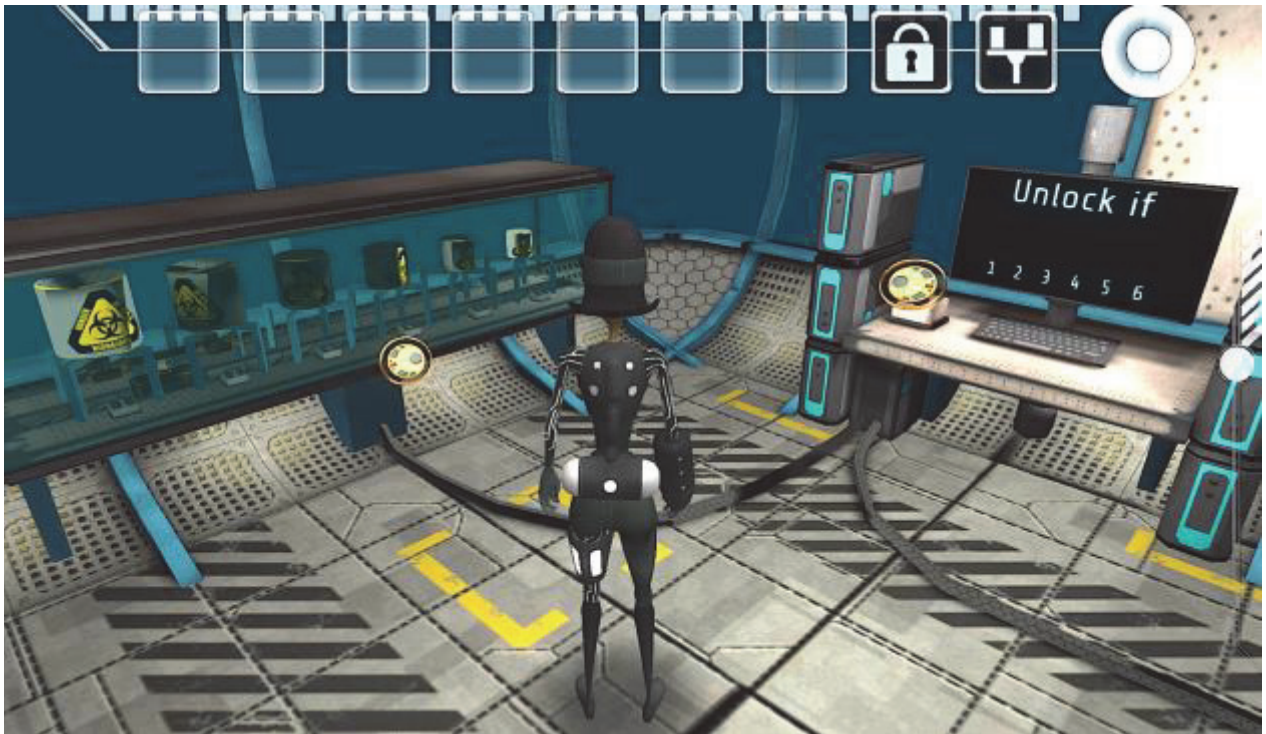
Figure 7-12. Screenshot of Trap Room (Level 3)
This room contains 6 cans that the students must reorder by writing a bubble-sort program in the block-based programming interface.

Chapter 8   STUDY DESIGN: VIRTUAL LEARNING COMPANIONS AND COLLABORATION

In Chapter 6, I detailed the promising results of a virtual learning companion pilot study; in Chapter 7, I explained how I iteratively developed a pair of virtual learning companions to support pairs of students collaborative playing ENGAGE. Building off that work, this chapter now describes the study design of the investigation into this dissertation's hypotheses, first listed in Chapter 1. In Section 8.2, this chapter contains details about the data collection of a two-condition experiment, including the study participants and types of data collected. The key type of data for addressing the study's hypotheses is audio of students' collaborative dialogue. Section 8.3 presents the dialogue act classification that will afford quantitative analyses of that dialogue. The quantitative analyses themselves will appear later in Chapter 9, with more in-depth, qualitative analyses following in Chapter 10.

8.1   Motivation

In general, this dissertation study's research goal is to explore opportunities for virtual learning companions to strengthen collaborative learning experiences. Specifically, this work focuses on *paired* collaboration. Among the numerous and less formal forms of collaboration (Kafai & Burke, 2013), this study is limited to collaborations in which two students are partners on a learning task. Secondly, this research specifically addresses the collaboration of *middle school students* in the United States. Thirdly, the collaborative learning for this study is situated within a *narrative-centered* learning environment. Accordingly, the virtual learning companions leverage storytelling to build rapport with the students and to contextualize the advice the learning companions supply. Finally, this research specifically concerns the domain of computer science education.

Due to the computer science subject matter, we can look to research on pair programming to understand the research problem. To wit, while research has established that pair programming affords certain benefits for students in general (Hanks et al., 2011), we know that troubling issues of equity can emerge among individual pairs (Lewis & Shah, 2015). At

times, one student in a pair might dominate the learning experience, while the other student takes on a passive role that can reduce the learning impact for that student. The proposed research addresses this issue of equity with a study of collaboration that will be conducted by analyzing data from a two-condition experiment in which some students used the baseline version of ENGAGE gameplay and others used an ENGAGE version that featured the addition of virtual learning companions. The learning companions were designed specifically to foster more equitable collaboration among student pairs.

The study examines the equity of student pairs through analyses of the student dialogue. The data includes an audio recording of the spoken utterances between each pair of students as the students played the game. This study relies on manually-transcribed transcripts of specific gameplay intervals, dispersed over multiple days for each student pair. Each student utterance in these transcripts was labeled according to a purpose-built dialogue act classification scheme. The labeled utterances then afforded an analysis of equity in each student partnership, based on frequencies of certain dialogue acts as well as sequences of dialogue acts.

8.2   Data Collection

8.2.1   Baseline Condition: ENGAGE with No Virtual Learning Companion

I collected the data that I used for this study during Spring 2016 and Fall 2016, as described in the preceding chapter. In late Spring 2016, twenty-four students at Carnage Middle School in Raleigh, North Carolina participated in the Oceanography class, in which pairs of students played the entire ENGAGE game over the course of several weeks. One student opted out of the Spring 2016 study, leaving eleven student pairs for analysis. The demographics for the 22 students in the analysis is as follows: 12 male and 10 female; 9 Indian, 4 Hispanic/Latino, 4 Caucasian, 3 Black/African-American, 1 Asian/Pacific Islander and 1 Multiracial. 12 students recalled having prior experiences with computer science and 10 students did not. Of the 11 pairs, 4 were male-female partnerships, 3 were female-female and 3 were male-male. Eight of the pairs included one student with prior computer science experience and one without, while

there were two pairs in which both the students reported prior computer science experience and one pair in which neither student reported prior computer science experience.

This dissertation will refer to these eleven pairs as part of the **Baseline Condition** because they played the original version of the ENGAGE game without virtual learning companions. During days that the classroom teacher designated as gameplay days, pairs of students played the game at their own pace. Because of the variance of paces between pairs, this could hypothetically result in faster pairs finishing many days before student pairs who progressed more slowly through the game. To mitigate this potentially undesirable issue, I designated several locations in the game beforehand as benchmarks at which students paused playing the game and only resumed gameplay on a later date when other students had also reached that benchmark. Figure 8-1 shows the program and an excerpt of collaborative dialogue from one of these benchmark locations: the Storage Room (which was described in the preceding chapter: Section 7.2.2).
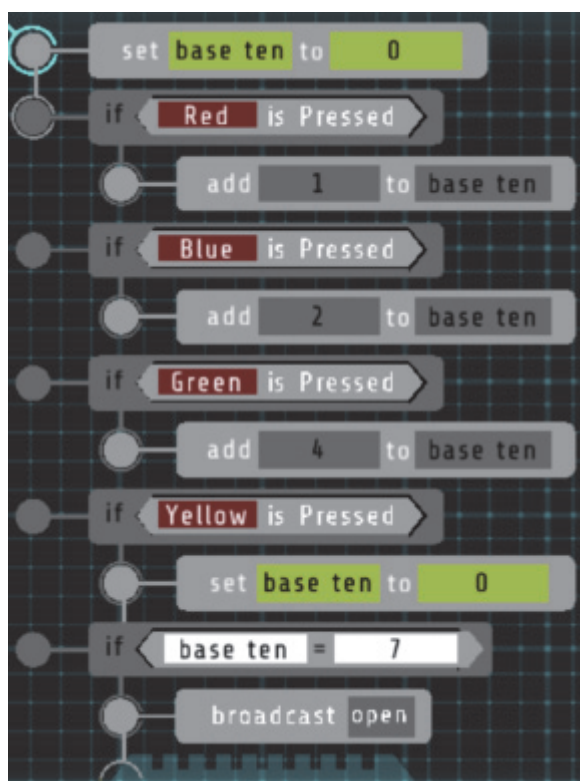
Speaker	Utterance
Dora	What number does the door want?
Dora	The door wants the open code.
Dora	How do you get open for the pressure pad?
Cecilia	Where's the pressure pad?
Cecilia	Oh there's so many pressure pads.
Dora	There's so many pressure pads
Cecilia	Wait, click that.
Dora	Don't ruin it, sorry.
Dora	Okay, if you see open…If base ten equals 7. So we have to make 7.
Dora	So we just press this, this, and that. Right?
Cecilia	Yeah, you just click all of them. It just goes…
Dora	Green, blue, red.
Cecilia	Green, blue red.

Figure 8-1. Example of a program and students' corresponding collaborative dialogue

Video and audio recordings were collected for each of the eleven student pairs in the Baseline condition. Each student pair played the game on a shared laptop. An audio recorder was placed on the laptop while students played the game, and this setup successfully recorded transcription-quality audio (*i.e.*, with the ability to hear both students' voices while minimizing the problem of background noise from other student pairs). For each student pair, a digital video recorder on a tripod captured the students playing the game. Because of the setup of the classroom, the tripods were placed in the aisles, thus recording students at a 45-degree angle as opposed to head-on (See Figure 8-2 for a still image that demonstrates the quality of the video). I used the video to identify speakers in the audio and to get clarification on how the two students were sharing the gameplay controls at a given time (e.g., if one student took over the keyboard to take care of the character controls at a specific time). In addition to the audio and video data, we also collected all data that is collected as part of ENGAGE's default logging. This includes game-trace data and survey data, as described in Chapter 3. During the last week of the course, after all students had completed the game, I conducted focus groups, each with six students. Video and audio were recorded for each of these focus groups.

Figure 8-2. Example of a still image from the video collected for the study

8.2.2 Companion Condition: ENGAGE with Virtual Learning Companions

In early Fall 2016, twenty-four students who had not previously played ENGAGE participated in the Oceanography class at Carnage Middle School in Raleigh, North Carolina. No student opted out of the study but one of the students was absent for an extended time, so there are eleven student pairs for analysis. The demographics for the 22 students in the analysis is as follows: 13 male and 9 female; 6 Indian, 4 Asian/Pacific Islander, 4 Caucasian, 3 Black/African-American, 2 Hispanic/Latino, 2 Native American and 1 Multiracial. 10 students recalled having prior experiences with computer science and 12 students did not. Of the 11 pairs, 1 was a male-female partnership, 4 were female-female and 6 were male-male. 3 of the pairs included one student with prior computer science experience and one without, while there were 3 pairs in which both the students had prior computer science experience and 4 pairs in which neither student had prior computer science experience.

This dissertation will refer to these eleven pairs as part of the **Companion Condition** because they played the ENGAGE game with the additional support of virtual learning companions. Students played the game in a similar manner—in pairs and stopping at the same benchmarks—as the students in the Baseline condition, with one difference. At four benchmarks in the gameplay (detailed in the previous chapter in Section 7.2), students conducted a conversation with a pair of virtual learning companions.[14] Since each of the conversations with the learning companions occurred immediately following a benchmark, students experienced these interactions at the beginning of a day's gameplay. The interactions were designed to last approximately 5-7 minutes, after which the students would return to the normal ENGAGE gameplay for the rest of the class session. Typically this allowed 25-30 minutes of gameplay following an interaction with the virtual learning companions.

---

[14] These conversation interactions were not integrated into the ENGAGE game interface. Students had to open a separate window to interact with the virtual learning companions, and were informed that this was part of the ongoing process of improving the game (in both conditions, the classroom teacher introduced the game as a work-in-progress).

As with the Baseline condition, video and audio recordings were collected for each of the eleven student pairs in the Companion Condition, as well as all the game-trace and survey data. I again conducted focus groups and collected video and audio for those. For the Companion Condition, there are two additional sources of data. First is the additional trace-data from the interactions that students had with the virtual learning companions. The second additional source of data comes in the form of post-surveys that contained questions about students' experience with the virtual learning companions. All the sources of data can be seen below in Table 8-1.

Table 8-1. Sources of Data for this Study

Prior to Starting Game	During Gameplay	After Finishing Game
Both Conditions		
- Knowledge Pre-Test - CS Attitudes Pre-Survey - Cognition - Spatial Reasoning - Prior Experience Survey - General Self-Efficacy	- Game-trace - Daily Engagement - Knowledge (Mid-Game) - Audio - Video	- Knowledge Post-Test - CS Attitudes Post-Survey - Engagement - Partner Ratings - Focus Group Feedback
	Additional Data in Companion Condition	
	Interaction-trace (with Virtual L.C.s)	L.C. Post-Survey

8.3   Dialogue Act Classification

This study primarily uses the audio data to address the research questions. A professional transcription service produced transcripts of the dialogue from each student pair, and then two independent taggers applied dialogue act tags to these transcriptions. I originally developed a proposed dialogue act classification by consulting ones that have been used to investigate equitable collaboration in computer science education (Lewis & Shah, 2015) as well as established learning sciences research (Correnti et al., 2015; Shaenfield, 2010). I also consulted dialogue act classification schemes that have informed a rich body of related work within the context of intelligent tutoring systems (Forbes-Riley, Litman, Huettner, & Ward, 2005; Stolcke

et al., 2000); I note, however, that the annotation scheme I developed for this study does not seek to analyze dialogue at as fine a level of granularity as some ITS research (Vail & Boyer, 2014). Finally, my familiarity with the dialogue of students playing the game (from classroom observations) also guided the development. To establish the stability of the classification scheme, I then followed best practices for verifying inter-annotator reliability (Cohen, 1988; Landis & Koch, 1977).[15] The refined dialogue act classification for this study contains six categories of dialogue act tags (some of which are families of more specific tags). The six categories consist of (a) Suggestions, (b) Questions, (c) Yields, (d) Emotives, (e) Other and (f) Abuse. The first five categories are mutually exclusive; the sixth (Abuse) was applied as a secondary tag to investigate an additional research question that emerged during the study.

8.3.1 Tag Descriptions

This section will provide details on each of the specific tags, including notes on how the tagging scheme evolved from conception to final form. In some cases, several individual tags in my original dialogue act classification scheme ended up collapsing into an overarching tag. In other cases, some tags were dropped due to reliability concerns (i.e., independent annotators had too low a level of agreement). The final tag listed below, Abuse, did not appear in the original scheme; as noted above I added it due to an emerging research question. Following these descriptions of the individual tags, the reader can find an overview of all the tags in Table 8-2, which also includes the frequency that each type of dialogue act appeared in the corpus.

Suggestions. The Suggestion tag applies to dialogue acts in which a student puts forth a suggestion on what the pair should do. In the initial dialogue act classification that I proposed before testing for inter-rater reliability, I formulated Suggestions as a family of dialogue act tags that included two more specific tags: *Assert* and *Guess*. The Assert tag referred to a dialogue move in which the student expresses confidence in the suggestion; for example, "I think we should use a Repeat block here". The Assert tag would also apply to confident utterances that go beyond mere suggestion, to the point of self-narration; for example, "I am

---

[15] Section 8.3.2 provides more details on this process.

going to use a Repeat block here". The Guess tag, on the other hand, would apply to dialogue moves in which the student expresses uncertainty in the suggestion; for example, "Maybe we should try going over there". However, the difference between Assert and Guess proved ambiguous enough that two independent taggers failed to agree on many of the Suggestion utterances. This family of tags thus collapsed into a single tag (Suggestion) for the dialogue act classification that I used for the study. As a final note on this category, another sub-category here could be "commands", which prior research on pair programming has considered significant for the purpose of analyzing equitable collaboration (Lewis & Shah, 2015). I excluded that tag from the proposed dialogue act classification due to my observation that few utterances in this corpus would qualify as a command (as defined by Lewis & Shah). For further explanation on this point, I will discuss in later chapters how the collaboration in this game-based learning experience differs from collaboration in pair programming.

Questions. The second dialogue act tag applies to all Questions. As with Suggestions, I initially conceived of a family of Question tags that included two distinct dialogue act tags: *Query* and *Rhetorical*. The *Query* tag would apply to questions to which the student is soliciting a response from the partner; for example, "How do you think we should solve this?". For question utterances tagged *Rhetorical*, the student would not be expecting any response from the partner; for example, "Will we ever get through this?". I anticipated that discerning between these two types of questions would be difficult for some student utterances, such as "What does this do?". In the end, the difficulty in differentiating the two Question types proved insurmountable for this study, so—like Suggestions—the Question family collapsed into a single tag.

Yields. The Yield tag applies to dialogue acts in which the student is responding to a partner's utterance without offering a novel suggestion. Yields can be difficult to tag in isolation; it was particularly important for the dialogue act taggers to consider the surrounding dialogue when applying a Yield tag to an utterance. In the proposed dialogue act classification (before refinement), Yield was part of a family of tags called Replies, which contained multiple individual dialogue act tags that included *Agree* and *Don't Know*. The Agree tag would refer to utterances in which the student accedes to the partner's suggestion; for example, if the

partner uttered the Assert, "Let's explore the room", the student's Agree might be "Ok". Don't Know utterances would be instances when the student says some variation of "I don't know". As I went through the transcripts with another independent tagger, we discovered few instances of unambiguous Don't Know utterances. Disagreement arose between the two taggers for many of the utterances that might have received this tag, such as "I guess. Don't ask me." So the Agree and Don't Know tags collapsed into the overarching Pass tag. A third tag that had been a member of the Reply family in the proposed dialogue act classification, *Answer*, proved difficult for taggers to differentiate from the Suggestion tag. For example, when one student asks the question, "And then, like, go forward?", and the other student replies by saying, "It was a go forward five times, I think", this second student's utterance could be considered both an answer and a suggestion. The Answer tag was therefore dropped for the refined dialogue act classification.

Emotive. The Emotive family encompasses all utterances that concern some affective response to the gameplay experience. Emotive utterances receive one of two more specified tags: *Emotive-Positive* and *Emotive-Negative*. Both Emotive tags cover a broad range of affective statements. Dialogue acts tagged as *Emotive-Positive* might include instances in which the student reacts happily to success in the game, such as "Yay! We did it!", as well as instances of the student providing encouragement, such as "Don't worry – we can do this". *Emotive-Negative* utterances include immediate reactions to specific game events, such as "Oh no! We fell!", as well as overarching negative feelings about the gameplay experience, such as "This is so frustrating".

Other. The fifth category of dialogue act tags, Other, contains dialogue acts that would not fit in any of the above categories. The Other family includes two specific tags, *Help* and *Off-task*, as well as a *Miscellaneous* tag to catch all remaining dialogue acts. The *Help* tag applies to utterances in which the student seeks external advice or assistance from somebody other than the student's partner (most frequently the teacher). Off-task utterances do not relate to the gameplay or learning experience: for example, if the student discusses weekend plans.

Abuse. The proposed dialogue act classification contained the above five categories, with the expectation of mutual exclusivity. During the study, I discovered the importance of

an additional tag, *Abuse*, that might be applied as a secondary tag. The Abuse tag applies to utterances in which one student is verbally abusing his or her partner (*e.g.,* "You're so stupid."). That we should need such a tag might, at first glance, seem shocking and problematic. Yet prior research on virtual learning companions has revealed that certain types of verbal abuse can potentially have a positive function with respect to rapport-building (Wang, Finkelstein, Ogan, Black, & Cassell, 2012). On the other hand, the empathetic researcher can easily imagine how verbal abuse might also have harmful effects. For this reason and others, I considered it imperative to track when students were making abusive statements directed at their partner.

8.3.2 A Note on Inter-Annotator Agreement

As noted above, I followed best practices for establishing the reliability of inter-annotator agreement (Cohen, 1988; Landis & Koch, 1977). After having first crafted the original, proposed dialogue act classification scheme, another researcher and I independently tagged a small subset (~8%) of the corpus. To quantify the agreement, I then calculated Cohen's Kappa, which can range from -1 (no agreement) to 1 (perfect agreement) (Cohen, 1960). The insufficient agreement during that testing round ($\kappa < .20$) prompted the refinement of the dialogue act classification scheme and further testing. I only proceeded to annotate the entire corpus once the calculated Kappa rose to a "substantial" level, which related work has established as above 0.60 (Landis & Koch, 1977). As I annotated the entire corpus, another researcher annotated a previously untagged subset of 21% (1467 utterances). The resulting agreement was $\kappa < .67$, indicating a sufficiently reliable dialogue act classification scheme.

The resulting dialogue act classification scheme afforded quantitative testing of the study's hypotheses, as detailed next in Chapter 9. I should note that recent work in the field of computer-supported collaborative work has warned that relying on Cohen's Kappa might result in Type I errors (Eagan et al., 2017), so I apply caution in proceeding to the quantitative analyses. As I discovered, however, caution may be superfluous for this particular work; the quantitative analyses largely failed to support the study's hypotheses. The next chapter will describe the results of those quantitative analyses for each of the hypotheses, and Chapter 10 will then provide further insight in the form of qualitative analyses.

Table 8-2. Dialogue Act Classification

Frequency	Tag	Description	Examples
.4105	Suggestion (SU)	Dialogue acts in which a student puts forth a suggestion on what the pair should do	*"Go over there."* *"Maybe we should try going over there."* *"I am going to use a Repeat block here."* *"I think we should use a Repeat block here."*
.1332	Question (QU)	Dialogue acts in which the student asks a question (might be rhetorical in nature)	*"What should we pick?"* *"What are you doing?"* *"Why is this so hard?"*
.0889	Yield (YD)	Dialogue acts in which the student is offering to "yield the floor" to their partner	*"Okay."* *"I don't know."* *"I guess so."*
.2029	Emotive		
	Positive (EP)	Dialogue acts that express a positive response to the gameplay experience	*"Yay! We did it!"* *"We can do this."* *"You're smart."*
	Negative (EN)	Dialogue acts that express a negative response to the gameplay experience	*"Oh no! We fell!"* *"This is so frustrating"* *"You're stupid."*
.1645	Other		
	Help (OH)	Dialogue acts spoken to somebody other than the student's partner	*"Excuse me, we need help."* *"Hey, how did you guys get past this?"*
	Off-task (OO)	Dialogue acts that do not relate to the learning experience	*"I have a soccer game tomorrow."* *"I need to do my homework."*
	Miscellaneous (OM)	Dialogue acts that would not fit in any of the above tags.	*"I wonder if it saved."* *"There's no platform."*
.0087	Abuse (AB)	Dialogue acts in which a student verbally abuses his/her partner	*"You're stupid."* *"Why are you so stupid?"*

Chapter 9   QUANTITATIVE ANALYSIS OF COLLABORATIVE DIALOGUE

The tagged transcripts discussed in the preceding chapter allow for a quantitative investigation of students' collaborative dialogue. For this study, analyzing the dialogue affords us the greatest opportunity to understand the nature of a student pair's collaboration. Each pair has a single game-trace, so that data does not allow us to differentiate the two individual students. As discussed in Chapter 8, the quality of the video data limits its use to supplemental analysis. This chapter will therefore analyze the coded transcripts with the goal of quantifying the collaborations, particularly to understand the impact of the virtual learning companions. As first detailed in Chapter 1, two primary research questions guided the design of the study:

- RQ1: To what extent does the addition of virtual learning companions affect collaborations between two middle school computer science learners, with respect to equitable contributions and as assessed through analyses of student dialogue?

- RQ2: To what extent are student characteristics of gender and prior computer science experience associated with the impact of the virtual learning companions?

In short, I am investigating the extent to which the students in the Companion Condition engaged in more equitable collaborations as compared to students in the Baseline condition, with careful consideration of student characteristics of gender and prior computer science experience. The term "equitable collaboration" may carry different connotations depending on one's perspective, so I have formalized my conception of the term in my hypotheses, guided by research in the learning sciences on how equitable collaboration relates to identity formation (Nasir & Hand, 2008; Shah et al., 2014). In doing so, I am adopting a situative perspective that "learning and a sense of identity are inseparable" and that students' learning depends on them having access to legitimate participation, which itself depends on the power dynamics within their learning environment (Lave & Wenger, 1991).

I hypothesized that various analyses of the tagged dialogue might afford quantitative measures of how equitably students collaborated. Section 9.1 details the results of analyses

that seek to quantify the impact of the virtual learning companions on equitable collaboration. Section 9.2 then looks at differences in the collaborative dialogue across gender, building on research into gender differences with virtual learning companions (Buffum, Boyer, et al., 2015). Finally, Section 9.3 examines the question of how students' prior computer science experiences might influence their collaborative dialogue. Throughout this chapter, I analyze the full data set to provide the reader broad insight into the trends that emerged. Chapter 10 will then analyze the data at a finer granularity, examining individual student pairs to gain a more nuanced understanding of the collaborative experiences of students.

9.1 Equitable Collaboration Across the Two Conditions

9.1.1 Occurrence of One Student Dominating the Collaboration

If one student is dominating the collaborative activity by making all the decisive moves, then the other student(s) in the collaboration might risk losing access to legitimate participation. With this in mind, I first analyzed the distribution of total utterances for each pair of students, calculating the proportion of the total utterances that came from the more "talkative" partner. I will refer to this as a student pair's *Total Distribution*. I then compared the two conditions of the study. For the Baseline Condition, the average proportion of total utterances made by the more "talkative" partner was 0.609. For the Companion Condition, the average proportion was a similar 0.6175. In other words, collaborations across both conditions saw, on average, a 60/40 split in the Total Distribution. This ratio was consistent across gender and prior computer science experience. Both conditions included one especially unbalanced collaboration, in which the more "talkative" partner made a little over 75% of the utterances. Both conditions also included a collaboration with almost even Total Distribution, in which the more "talkative" partner made no more than 53% of the utterances. In summary, there is no evidence to suggest that the virtual learning companions had any impact on the Total Distribution.

Prior research on equitable collaboration has established, however, that we must look beyond overall distributions and also examine the *content* of students' talk (Shah et al., 2014). In collaborative gameplay, one can imagine a scenario in which two students make an equivalent number of utterances, but while Student A makes many suggestions about how to

solve the given problem, Student B's utterances are largely limited to reactions to game events. In such a collaboration, Student B's more passive participation might not lead to as much learning compared to Student A. This consideration led to one of the study's hypotheses (first listed in Chapter 1):

> H1.1 The learning companions will significantly reduce the occurrence of one student dominating the collaboration. The ratios of suggestion utterances to emotive utterances will have greater convergence between student partners in the Companion condition than in the Baseline Condition.

To test this hypothesis, I first calculated the ratio of Suggestions to Emotives (Emotive-Positive + Emotive-Negative) for each individual student. I will refer to this ratio as the *Individual S:E Ratio*, and note that each student pair consequently had two Individual S:E Ratios. For each student pair, I then calculated the ratio of the larger of the two Individual S:E Ratios to the smaller one. I will refer to this ratio as the *Comparison S:E Ratio*. Intuitively, students with relatively larger Individual S:E Ratios were using most of their utterances to advance the problem solving task, while those with relatively small Individual S:E Ratios may have taken on more passive roles, merely reacting to the game events. A relatively large Comparison S:E Ratio, meanwhile, indicates an imbalanced collaboration, as in the scenario envisioned in the previous paragraph. On the other hand, a Comparison S:E Ratio of 1.00 would indicate a perfectly balanced collaboration by this metric.

Having calculated the Comparison S:E Ratio for each student pair, I investigated whether these ratios differed across the two conditions. Student pairs in the Baseline Condition had an average Comparison S:E ratio of 2.3043, while those in the Companion Condition had an average Comparison Ratio of 1.3515. Although this seems like a trend in support of the hypothesis, a Mann-Whitney U test revealed it to be statistically insignificant. The higher ratio for the Baseline Condition largely comes from a few collaborations with particularly high Comparison S:E Ratios. While none of the student pairs in the Companion Condition had particularly high Comparison S:E Ratios—the highest being 2.0297—this metric does not supply enough evidence to support the above hypotheses (H1.1).

As another analysis in the same vein, we might look at the distribution of Suggestions between the two students in a collaborative pair. Prior research on the collaborative dialogue of children pair programming with Scratch has demonstrated that the distribution of student "commands" can reveal the extent to which a collaboration is equitable (Shah et al., 2014). To be clear, I did not hypothesize that the results from this study would replicate those prior results; this dissertation study differs from that prior work both in the type of activity (game-based learning versus block-based programming) and in the style of collaboration (informal role-sharing versus a formal driver-navigator paradigm). Perhaps due to those key differences, students in this study rarely made utterances that could be tagged as a "command", which prior work has classified as "any statement that included a request to perform an action" (Lewis & Shah, 2015). The closest comparison is to the Suggestion tag, although I note that Lewis and Shah make clear that indirect requests (e.g., "We should…") were *not* tagged as "commands" in their work. In any case, I calculated for each pair the proportion of Suggestions (which I will refer to as the student pair's *Suggestion Distribution*) in the same way that I calculated the Total Distribution, described above. For the Baseline Condition, the average proportion of total utterances made by the more "suggesting" partner was 0.6667. For the Companion Condition, the average proportion was 0.6147. This is not a statistically significant difference.

From all the above analyses, I conclude that we cannot reject the null hypothesis that the virtual learning companions had no impact overall on the phenomenon of one student dominating the collaboration. Investigating this hypothesis (H1.1) has instead revealed the diversity of ways that individual collaborations can vary with respect to that phenomenon. Combining the above metrics, particularly the Suggestion Distribution and the S:E Ratio, can provide a fuller picture of the equitable collaboration within an individual student pair. The next chapter will take a closer look at specific student pairs and discuss how a virtual learning companion could support the students in having an equitable collaborative learning experience.

9.1.2 Extent to Which Students Seek Partner's Input

Another way to measure equitable collaboration is to look at the questions that students ask their partners. Every time a student seeks input on the problem-solving task from his or her partner, that partner is positioned as a legitimate source of expertise. We might expect an

equitable collaboration, in which both students perceive both themselves and their partners as legitimate sources of expertise, to include more questions than a less equitable collaboration. Furthermore, an equitable collaboration likely sees the two students asking each other an equivalent number of questions. Accordingly, one of this dissertation's hypotheses addressed the issue of questions:

> H1.2   The learning companions will increase the extent to which students seek their partner's input. There will be a greater number of questions in the Companion Condition than in the Baseline Condition, and the ratio of questions tagged *Query* to questions tagged *Rhetorical* will also be greater in the Companion Condition.

The first half of this hypothesis is straight-forward to examine. In the Baseline Condition, student pairs averaged 40.45 questions, compared to 52.11 questions in the Companion Condition. The large variance (with a range of 19 to 94 questions), however, makes it an insignificant difference. As with suggestions, we can also look at the distribution of questions within each student pair. The average Question was similar in both condition: 0.6183 in the Baseline Condition and 0.6278 in the Companion Condition. Indeed, I found no evidence that students' interactions with the virtual learning companions led them to utter more questions.

To get to the root of this issue, though, we should differentiate questions that students ask their partners from rhetorical questions (*i.e.*, questions that do not seek input from another person). Only the former style of question can position the partner as a source of expertise, so those questions are particularly important for the investigation of equitable collaboration. The second half of the hypothesis addresses this concern, but we unfortunately cannot fully test it. As mentioned in the description of the Question category in Section 8.3, differentiating between Query questions and Rhetorical questions proved too difficult for the independent taggers. Other related research has also encountered this challenge (Lewis & Shah, 2015), and it remains an open question of how to overcome it. The qualitative analyses in Chapter 10 will allow for a more nuanced look at the role of questions within equitable collaboration.

9.1.3 Responding to Partner's Uncertainty

A similar issue prevents any quantitative analysis of another of the hypotheses listed in Chapter 1:

H1.3 For instances in which one student issues an utterance tagged *Guess*, there will be a greater probability that the partner responds with an *Assert* in the Companion Condition than in the Baseline Condition.

As with the Question category, Section 8.3 describes how the refinement of the dialogue act classification led to the Suggestion category collapsing into a single, over-arching tag. The difference between a Guess and an Assert proved too ambiguous for many Suggestion utterances. Unlike with the matter of differentiating questions (discussed above in Section 9.1.2), even qualitative analysis proved difficult for this hypothesis. The motivation behind the hypothesis is that equitable collaborations might see both students' attempt to resolve each other's uncertainty. Chapter 10 will provide more insight into this hypothesis, although it remains the most dubious of all the dissertation's original hypotheses.

9.1.4 Responding to Partner's Negativity

A final hypothesized characteristic of positive collaboration concerns students' follow-up to negative utterances. If one student expresses frustration, for example, we might hope for the student's partner to respond with encouragement. The virtual learning companions explicitly encouraged students to support their partners through frustration, so I hypothesized in Chapter 1 that students in the Companion Condition would exhibit more of these positive behaviors:

H1.4 For instances in which one student issues an emotive utterance tagged *Emotive-Negative*, there will be a greater probability that the partner responds immediately with an emotive tagged *Emotive-Positive* in the Companion Condition than in the Baseline Condition.

Analysis of the coded transcripts revealed that this pattern was rare in both conditions, with some student pairs not producing a single instance of it. In some pairs, neither student made many Emotive-Negative utterances, so there were few opportunities for the pattern to emerge. Other pairs were prolific in their Emotive-Negative utterances but rarely engaged in the encouraging behavior envisioned by the above hypothesis. I calculated the conditional

probability for each group and then compared the average for both the Baseline and Companion conditions. The Baseline (0.0883), in fact, had a higher average than the Companion condition (0.0313), negating the hypothesis that the virtual learning companions would help produce more of these patterns among student pairs who interacted with them. The higher average for the Baseline Condition can be attributed to a single pair with a particularly high conditional probability. Chapter 10 will go into depth on that pair.

9.2   Gender and Equitable Collaboration

In addition to equitable collaboration in general, this dissertation takes particular interest in gender equity. As discussed in previous chapters, middle school computer science interventions must consider the underrepresentation of women in the field of computer science because students begin their career trajectory as early as middle school (Lent et al., 1994). Middle school computer science would therefore seem like fertile ground for the introduction of virtual learning companions, given the research showing their benefits for supporting female students (Arroyo et al., 2011). Accordingly, I included the following hypothesis relating to gender differences:

H2.1   The impact in the Companion Condition, as tested through hypotheses H1-H4, will be more pronounced with female students than male students.

As explained above, however, the preceding hypotheses did not bear significant results. Given the relatively small sample sizes of female-female pairs in the two conditions (3 such pairs in the Baseline Condition and 4 such pairs in the Companion condition), we should not expect to find statistically significant results on any of the metrics described in Section 9.1. I also did not find any evidence in the survey data that female students experienced any greater enjoyment or benefit from the virtual learning companions than male students did. This null finding contrasts to prior work on virtual learning companions, including my own research findings described in Chapter 6. I hypothesize now that the explanation lies in the fact that the virtual learning companions in this study support collaborative learning, whereas prior research has examined their impact with students working alone. I will go into more detail on this distinction as I analyze specific female pairs in Chapter 10.

While the virtual learning companions did not seem to have much impact among female students, an unexpected—and potentially alarming—finding emerged among male students in the Companion Condition. Following their interactions with the virtual learning companions, four out of the six male-male pairs exhibited verbally abusive behavior toward one another during their ensuing gameplay. None of the student pairs in the Baseline Condition exhibited such behavior, nor did any of the pairs with female students in the Companion Condition. Upon seeing this phenomenon in the transcripts (and adding the Abuse tag to be able to quantify its prevalence), I checked these student pairs' dialogue from the first day of Engage gameplay, before their first encounter with the virtual learning companions. I wondered if the Companion Condition happened to include male students who were already more verbally abusive to one another at the outset, compared to male students in the Baseline Condition. Yet I found no instances of verbal abuse in their prior day's gameplay (although the audio was missing for one pair).

I then looked at the interaction-trace of students' interaction with the virtual learning companion. For all the female-female pairs and the one female-male pair, there were no instances of the students selecting an abusive response to the virtual learning companions. On the other hand, five of the six male-male pairs chose at least two abusive statements to make toward the virtual learning companions. Prior research has found that male students often abuse virtual learning companions (Silvervarg et al., 2012), but this finding suggests that they might then transfer that verbally abusive onto their peers in collaborative learning experiences. Table 9.1 details the distribution of abuse among the six male-male groups. Chapter 10 will provide in-depth analyses of two of these groups, illustrating how sometimes the verbal abuse might be a form of innocuous rapport-building but other times it can cause real harm.

Table 9-1. Abusive Behavior by Male Students

Group #	Verbal Abuse: Student 1	Verbal Abuse: Student 2	Abuse Against Learning Companion
3	0	0	10
4	15	11	10
5	0	6	2
7	0	0	0
8	5	2	8
12	7	14	17

9.3 Prior Computer Science Experience and Equitable Collaboration

Finally, this dissertation takes keen interest in how middle school computer science interventions equitably support students with no prior computer science experience. In the interest of broadening participation in computing, we must design middle school interventions that appeal not only to students who have already shown interest in the field but also to those students who might not otherwise consider it for themselves. I designed the virtual learning companions with this grand challenge in mind, and hence included the following hypothesis:

> H2.2 The impact in the Companion Condition, as tested through hypotheses H1-H4, will be more pronounced among students without prior computer science experience than those with experience.

As with the preceding hypothesis on gender, this hypothesis proved difficult to analyze quantitatively given the null results for hypotheses H1-H4. Indeed, it seems likely that the virtual learning companions may not have had a pronounced impact on these students of interest. More work is needed on interventions for supporting these students in collaborative learning experiences, as certain metrics described above in Section 9.1 reveal that students with no prior computer science experience may have less access to legitimate participation when they are paired with a partner who does have some prior experience. Specifically, these pairs (*i.e.*, one student with prior computer science experience, one without) had an average Suggestion Distribution of 0.6800, while those pairs in which the two students had equal levels of prior computer science experience had a more balanced average Suggestion Distribution of 0.5975. A one-way ANOVA found this to be statistically significant ($F(1, 20) = 6.181$, $p <$

.05). Chapter 10 will look in-depth at how these unbalanced collaborations, in which the one student with prior computer science experience makes most of the suggestions, can lead to suboptimal learning experiences for the student with no prior computer science experience.

Chapter 10   QUALTITATIVE ANALYSIS OF COLLABORATIVE DIALOGUE

As Chapter 9 outlined, the quantitative measures of equitable collaboration revealed no differences between the two conditions. In other words, the preceding chapter provides evidence that the addition of the virtual learning companions did not, in fact, lead to more equitable collaboration between partner students playing Engage. The quantitative metrics in that chapter can, however, help us to understand better the nature of equitable collaboration and to identify instances in the student dialogue that may be detrimental to student learning. This chapter will examine dialogue excerpts of individual pairs, with the goal of uncovering patterns relevant to this dissertation's research questions (RQs). Section 10.1 looks at a series of issues relating to RQ1, which concerns equitable collaboration across the two conditions in general. Sections 10.2 and 10.3 then address RQ2 by examining differences based on gender and prior computer science experience, respectively. Finally, Section 10.4 discusses how analyzing equitable collaboration in game-based learning presents different challenges than doing so in pair programming.

10.1 Equitable Collaboration Across the Two Conditions

As mentioned above, the quantitative metrics revealed no significant differences in the equitable collaboration of student pairs in the Companion Condition versus those in the Baseline Condition. I believe that multiple considerations may factor into the explanation for the learning companions' apparent lack of success.

Firstly, I designed the learning companions to support student collaboration based on my understanding of what equitable collaboration looks like, yet I may have overestimated my prior ability to identify equitable collaboration from observations alone. Subsection 10.1.1 describes two pairs in the Baseline Condition. I designed the learning companion episodes with one of these pairs (Baseline Group 03) in mind, thinking that group to be an example of inequitable collaboration. Further analysis, however, has revealed that this group may have

exhibited more equitable collaboration than I thought, and that another group (Baseline Group 08) could have presented a more useful example of inequitable collaboration.

Secondly, I also may have underestimated the complexity of designing a learning companion interaction that supports a pair of students, as opposed to one that supports an individual student. Chapter 2 presented prior work that has established how virtual learning companions can benefit individual students, and I hypothesized that these benefits would translate to pairs of students collaborating in a game-based learning environment. That hypothesis might still hold promise, but it calls for further exploration of how the fundamental design of a learning companion interaction should differ from one designed for individual students. Subsection 10.1.2 illustrates a challenge that emerges when the two students in a pair have divergent responses to the virtual learning companions (e.g., one student is less receptive to them than the other student is). Subsection 10.1.3 then shows how, even when the two students do not have divergent responses, the nature of collaborative gameplay alters how they might receive the virtual learning companions.

10.1.1 The Challenge of Discerning Inequitable Collaboration

In preparation for my dissertation proposal, I originally transcribed and tagged a short segment of dialogue so as to provide an illustration of my proposed work. Table 10-1 below provides this one sample segment of dialogue, which comes from one of the student pairs in the Baseline Condition (Baseline Group 03). Based on classroom observations, I hypothesized that Baseline Group 03 represented an example of inequitable collaboration throughout the overall gameplay experience. We can start to see an example of inequity in this segment of dialogue that occurred on the second day of gameplay. Baseline Group 03 was a pair of girls, whom this document will reference as Ana and Benita.

Table 10-1. Sample excerpt of dialogue from Baseline Group 03
In which Ana makes minimal contributions to the problem-solving task

#	Speaker	Utterance	Tag
39	Benita	Ok, no we could just move this way and move this way and keep moving straight forever	SU
40	Ana	Yeah	YD
41	Benita	Let's do that. That might be easier.	SU
42	Benita	Uh…	EN
43	Ana	Uh, wrong	EN
44	Benita	Yeah. K.	YD
45	Ana	At least she enjoyed her walk in the water.	EP
46	Benita	K. Let's go. At least it saved all this.	EP
47	Ana	Whee. <inaudible>	EP
48	Benita	<inaudible>What are you doing?	QU
49	Benita	Oh my gosh.	EN
50	Ana	No! It's stuck	EN
51	Benita	It left without us.	EN
52	Ana	It's stuck	EN
53	Benita	Ok	OM
54	Ana	Ok so…	OM
55	Benita	Ok, let's go…	SU
56	Benita	This takes a while	EN
57	Ana	Mhm	YD
58	Benita	Just don't move and it will be ok….	SU
59	Benita	Ok...Ok, so let's redo this program. So we take this out.	SU
60	Ana	Mhm	YD
61	Benita	And what if we just keep going….	QU
62	Benita	Oops...Let's keep that too.	SU
63	Ana	Okay	YD
64	Benita	This better work	EN
65	Benita	Don't fall off! Ok, wow, that was close…	EP
66	Benita	Try and jump. Yeah…	SU
67	Ana	Noooo!	EN
68	Benita	Unh! Noooo!	EN
69	Benita	At least it works.	EP

At the most basic level—comparing raw counts of overall utterances—I noted that the student Benita dominated this excerpt of dialogue with 20 utterances compared to 11 from Ana. Looking more closely at the nature of the utterances, I observed that Benita's dialogue contributions included an approximately 1:1 ratio of suggestion utterances and emotive utterances (7 suggestions to 9 emotives), whereas Ana had no suggestions compared to 6 emotives. This segment of dialogue also contains only one question, and it is a rhetorical question asked by Benita (utterance number 48); I posited that a more equitable collaboration might see the two students actively soliciting more feedback from their partners. Finally, we also can see in this segment of dialogue that when one student makes an Emotive-Negative reaction, the other student responds with another Emotive-Negative reaction. I originally hypothesized that a more fruitful collaboration would feature more instances of students responding to their partners' Emotive-Negative reactions with encouragement (*i.e.*, an utterance tagged Emotive-Positive). I considered this excerpt of dialogue to be representative of the two girls' collaboration throughout their multi-day gameplay experience, as determined via my classroom observations of the pair.

Contrary to my expectations, however, the hypothesis testing process (see Chapter 9) revealed that Baseline Group 03 scored well on the two metrics that I consider most promising for defining equitable collaboration in the game-based learning experience. The Comparison S:E Ratio measures how similar the two students in a pair are in regard to their balance of making suggestions and emoting (a high Comparison S:E Ratio might indicate that one student's contributions are largely confined to emotive statements while the other student makes the majority of the crucial suggestions). The average Comparison S:E Ratio across both conditions was 1.853; Baseline Group 03 had one of the lowest such ratios at 1.1363. The Suggestion Distribution, meanwhile, represents how equally the two students in a pair shared the responsibility of making suggestions. Across both conditions, the average Suggestion Distribution was 0.6407 and Baseline Group 03 was just below this mean at 0.6321.

As it turned out, I seem to have misjudged the collaboration in Baseline Group 03. Indeed, although Ana, who is a year younger than Benita, often expressed less confidence in her abilities, she ended up making crucial contributions at times when the pair got stuck. Table

10-2 contains three such examples from the second benchmark day. In Excerpt 1, Ana temporarily takes over the controls from Benita after the two have been struggling for several minutes trying to get through a door. When Ana's idea works, Benita congratulates Ana and Ana provides an explanation for why it worked. Excerpt 2 again comes after a lengthy period of struggling to get past an obstacle, and Ana suggests that they ask the teacher for help. After Benita (who consistently expresses greater self-confidence) turns down that suggestion, Ana has a sudden "aha" moment, much to Benita's appreciation. Excerpt 3 again shows Ana feeling despondent at the beginning, but then making a realization that helps the pair solve the current challenge. By the end, she is laughing about how easy it is.

Unlike Baseline Group 03, another pair of girls—Baseline Group 08—actually did exhibit inequitable collaboration as measured by the metrics of Comparison S:E Ratio and Suggestion Distribution. As mentioned above, the average Comparison S:E Ratio across both conditions was 1.853; Baseline Group 08 had the second highest such ratio at 3:48 (a higher ratio indicates a less equitable collaboration). Meanwhile, the average Suggestion Distribution was 0.6407; Baseline Group 08 had the third most unbalanced of all groups at 0.7297. The reader might wonder why I did not think to base my study design off this group instead of Baseline Group 03, given this group's more inequitable collaboration. In truth, I erroneously thought—following my several weeks of classroom observations but before a comprehensive examination of the transcripts—that Baseline Group 08 was one of the groups that best exhibited *equitable collaboration*. It did not occur to me that I should maybe design the learning companion episodes more to support pairs like Baseline Group 08 than pairs like Baseline Group 03.[16]

---

[16] In hindsight, I recognize this as a failure of process. Had time allowed, it would have been more efficient to begin with the analysis of equitable collaboration on the Baseline groups, and then—following that analysis—to design the learning companion episodes to maximize impact. As is often the case, however, failure can be instructive.

Table 10-2. Additional excerpts of dialogue from Baseline Group 03
In which Ana makes significant contributions to the problem solving tasks

#	Speaker	Utterance	Tag
		Excerpt 1	
156	Ana	Let me see something. Okay go.	SU
157	Benita	Can we just walk through the door?	QU
158	Benita	Ah!	EP
159	Ana	I did it! I knew it!	EP
160	Benita	Yeah! Good job! Yes! Good job.	EP
161	Ana	Cuz it set binary zero, and it set binary zero again.	SU
162	Benita	Oh, so you just pressed it -	SU
163	Ana	Again	OM
164	Benita	Again	OM
		Excerpt 2	
216	Benita	I have no idea what they're talking about, though.	YD
217	Ana	Should we ask help from the teacher? That we don't understand this?	QU
218	Benita	I don't know, like…	YD
219	Ana	It doesn't hurt -	OM
220	Benita	I kinda want to figure this out by myself	SU
221	Ana	Oh look! Oh we can type something in here!	SU
222	Benita	No we can't - oh really!	EP
223	Ana	Yes!	EP
224	Benita	Really!? Great.	EP
		Excerpt 3	
267	Ana	Nothing happening - this is just great.	EN
268	Benita	Wait wait	OM
269	Ana	Wait, that's ten now.	SU
270	Benita	Yeah. We changed it.	SU
271	Ana	So it needs to be one	SU
272	Benita	Yeah	YD
273	Ana	So it's one, and it has to let us out…	SU
274	Ana	Oh my gosh!	EP
275	Benita	That was it, this whole time?	QU
276	Ana	That is…[Reading] "Don't worry, you can handle this"	OM
277	Benita	Wow, couldn't offer this encouragement earlier.	EN
278	Ana	[Laugh] This is so easy.	EP

I believe that the crucial difference between these two groups—the difference that led to my error in judgement—lies in the students' respective self-efficacy. On the general self-efficacy survey, both students in Baseline Group 08 scored 31 (which also happens to be the approximate average across all the middle school students who took this survey). In Baseline Group 03, Benita similarly scored 30, but Ana scored 24, the second lowest in their class. Ana's relatively low self-efficacy, combined with her younger age compared to Benita, may have led to their collaborative dialogue sounding more inequitable than it, in fact, was. In contrast, neither of the two girls in Baseline Group 08 had low self-efficacy and they both talked confidently and prolifically (while their Suggestion Distribution was highly unbalanced at 0.7297, their distribution of total utterances was 0.57, among the most balanced of all the groups). As I later discovered, however, one of these girls in Baseline Group 08, Cecilia, was both younger and had less prior computer science experience than her partner, Dora. While both girls in Baseline Group 08 were high-performing students in general (and seemed to know that fact about each other), Dora had a significant advantage in computer science expertise.

Thus, the general tone of Cecilia and Dora's dialogue—upbeat and positive on both sides—may have misled me to believe they were sharing a more equitable collaboration than they were.[17] In fact, Baseline Group 08 was the one group in either condition that exhibited frequent examples in which one student makes an Emotive-Negative utterance and the other student then follows up with an Emotive-Positive utterance (I mentioned this one exceptional group in the preceding chapter, Section 9.1.4). Table 10-3 contains an excerpt from their collaborative dialogue that exhibits some of this positive support. The reader should note that Cecilia and Dora appear to be directing some of their utterances at their avatar, and the two students are sharing the gameplay controls. These factors can make it difficult to fully understand the nature of their emotive utterances (although a review of the video of their gameplay brought clarification). On this point, Section 10.4 will address how the nature of the

---

[17] I also note that, as with Baseline Group 03, both students in Baseline Group 08 reported positive feelings about collaborating with their partners, both in their responses on the Daily Mini Engagement survey and in the end-of-term focus groups.

gameplay leads to collaborative dialogue that can be fundamentally different than the type of collaborative dialogue coming out of pair programming activities.

Table 10-3. Sample of dialogue from Baseline Group 08
With examples of encouragement

#	Speaker	Utterance	Tag
67	Cecilia	Oh my god.	EN
68	Dora	I laugh at you.	EN
69	Cecilia	I know. I can't even...I couldn't even jump last time.	EN
70	Dora	It's okay.	EP
71	Dora	Don't jump when you're really close, jump from far away. Jump forward.	SU
72	Cecilia	There you go.	EP
73	Dora	Now jump.	SU
74	Dora	Oh my god, all that...	EN
75	Dora	You can do it. There you go.	EP
76	Cecilia	There you go. Now I don't know how to get up there.	EP
77	Dora	I'm probably going to mess it up.	EN
78	Cecilia	There you go.	EP
79	Dora	There we go.	EP

We can thus see that Baseline Group 08 does exhibit some dialogue patterns that I had originally hypothesized would be indicative of equitable collaboration. Yet I maintain that this group's collaboration was, in fact, inequitable overall, based on the aforementioned Comparison S:E Ratio and Suggestion Distribution. A comprehensive review of their dialogue reveals that Cecilia, while chatty and confident, made comparatively few suggestions about how to solve the challenges in the game, deferring to the more experienced Dora. I argue that such a dynamic will limit the ability of a student in Cecilia's position to see herself as a full contributor in the computer science activity.

This realization, coupled with the fact that Baseline Group 08 was the only group that exhibited many instances of one student responding to a partner's negativity, leads me to question whether responding to a partner's negativity is, in fact, a characteristic of equitable collaboration. To be clear, I do not argue that such seemingly supportive utterances lead to negative outcomes; I only postulate that when students offer verbal support following a

partner's expression of frustration (or any other negative emotion), it might have little benefit in positioning the partner as a full contributor. If so, perhaps it is less crucial than I had thought that virtual learning companions encourage such behavior between paired students.

10.1.2 Pairs' Divergent Reception to the Learning Companions

We can now turn to examine the addition of the virtual learning companions, with respect to equitable collaboration between student partners. I had hypothesized that we might be able to bring the positive impacts of virtual learning companions to paired collaboration by simply focusing the learning companion episode on the issue of equitable collaboration. In other words, I hoped that creating an episode in which student pairs "talk" to a virtual learning companion about challenges in the game could have a similarly positive impact as I saw when I created an episode in which individual students "talk" to a virtual learning companion (see Chapter 6). Student pairs, however, invariably have a much more complex interaction with technology (e.g., the virtual learning companions) than do individual students. In the case of individuals, even those students who might not fully enjoy the presence of the virtual learning companions would experience a certain degree of immersion in the learning companion episode. For student pairs, on the other hand, each student is simultaneously navigating an interaction with his or her human partner in addition to the interaction with the virtual learning companion. A review of the dialogue from the Companion Condition leads me to believe that this multidimensionality can minimize the impact of the virtual learning companions.

Table 10-4 shows an excerpt from Companion Group 05, from when this pair was going through the first episode with the virtual learning companions. One of the boys, Lenny, repeatedly finds amusement in disparaging the virtual learning companion and making jokes. The other boy, Manny, who has ultimate control over the interaction by holding the mouse, seems to take the interaction with the virtual learning companions more seriously. In other words, the two boys diverge in their respective receptiveness to the virtual learning companions. As the interaction proceeds, Lenny's disparaging remarks seem to annoy Manny. He tells Lenny, in utterance #21, to "Shut up" and then, in #21, to "just read". At the end of this excerpt, after Lenny utters another abusive statement directed toward the virtual learning companion, Manny responds by uttering an abusive statement directed toward Lenny. One can

imagine that, were the two boys each experiencing the learning companion episode as individuals, they might have enjoyed more positive experiences. Manny could have immersed himself in the episode and reaped the benefits of the virtual learning companion, while Lenny could have abused the virtual learning companions as much as he liked (thereby at least gaining some sort of entertainment). As a pair, however, their interaction with the virtual learning companions ended up leading to friction in their interaction with each other.

Table 10-4. Sample of dialogue from Companion Group 05
With examples of verbal abuse

#	Speaker	Utterance	Tag
11	Lenny	Eww!	EN
12	Manny	Let's speak to Adriana. Okay.	SU
13	Manny	[READING OUT LOUD]	OM
14	Lenny	[crosstalk] for you?	QU
15	Manny	Nope.	SU
16	Lenny	[READING OUT LOUD]	OM
17	Manny	[READING OUT LOUD]	OM
18	Lenny	[READING TEXT OF WHAT HE WANTS TO CLICK]	SU
19	Manny	Yeah.	YD
20	Lenny	All right, we need to locate this bar.	SU
21	Manny	Shut up.	EN, AB
22	Manny	[READING OUT LOUD]	OM
23	Lenny	[READING TEXT OF WHAT HE WANTS TO CLICK] Do that one.	SU
24	Manny	[READING OUT LOUD]	OM
25	Manny	Let's go see that doesn't [inaudible]. Oh, I see.	OM
26	Lenny	Oh, no, press that one.	SU
27	Manny	I know what to do, [Lenny].	SU
28	Manny	Just read.	SU
29	Lenny	Look at her face, though. Ew.	EN
30	Manny	[Lenny], you're stupid.	EN, AB

I wondered if this divergent receptiveness might be especially present in mixed-gender pairs, given all the literature on gender differences with respect to virtual learning companions (see Section 2.1.3 in Chapter 2 for a full discussion on that literature). There was only one mixed-gender pair in the Companion Condition, and the transcripts for this pair do, in fact, exhibit a similar divergence. Companion Group 10 included a boy, Olly, who expressed a lot of distaste for the virtual learning companions, and a girl, Natasha, who made few utterances during the episode but later reported enjoying the virtual learning companions on a post-survey. Table 10-5 shows an excerpt of their dialogue. Olly controls the interaction with the virtual learning companions (holding the mouse), and narrates out-loud his thinking. Throughout the entire learning companion episode, and continuing after they have finished with the episode, he expresses his distaste for the experience. Natasha, meanwhile has little opportunity to immerse herself in the interaction with the virtual learning companions. She also shows little interest in interacting with Olly.

Natasha's reticence to interact with Olly would continue throughout the rest of the gameplay on this day, as she would only end up making 44 total utterances compared to Olly's 127 (many of his utterances throughout the gameplay seem to be in the form of self-narration, rather than directed toward Natasha). This imbalance in total utterances was among the most extreme of any pair from any day of gameplay. To be clear, there may not be enough evidence to conclude that the divergence in their receptiveness to the learning companions led to Natasha disengaging from her collaboration with Olly. Yet Natasha's independent feedback (on post-survey and in focus groups) do confirm that she saw potential value in the virtual learning companions, and one can imagine that she might have benefited more fully from the episode without Olly's negative commentary.

Table 10-5. Sample of dialogue from Companion Group 10
In which Olly criticizes the virtual agent and Natasha stays quiet

#	Speaker	Utterance	Tag
25	Olly	Oh, that's a nice option, but I don't want to hear more about the piano.	SU
26	Natasha	You don't have to read it. The last one's not nice.	SU
27	Olly	Oh, fine. I'm just going to talk about piano.	SU
28	Olly	[READING OUT LOUD]	
29	Natasha	Do the last one.	SU
30	Olly	Wait. Nope, nope, no, no more, no more. [inaudible]. It is.	SU
31	Natasha	[00:03:45] Why wouldn't it be spelled right?	QU
32	Olly	I don't know.	YD
33	Olly	[READING OUT LOUD]	
34	Olly	Don't want to hear anything. Don't need any extra dialogue.	EN
35	Olly	Where is [inaudible]?	QU
36	Olly	Must stop talking.	EN
37	Olly	[READING OUT LOUD]	
38	Olly	Oh, please don't say we have to talk to Adriana too. I think you have to talk to Johnny [inaudible] so far.	EN
39	Olly	Is that it?	QU
40	Natasha	I think so.	YD
41	Olly	Watch Adriana pops up right when I'm moving other page.	EN
42	Olly	Yeah, that's it. Yeah, that's it. Okay, got it. You did it. Enough.	EP
43	Olly	It's okay, it's okay. I can finally go to the full screen, I think. Continue game.	SU
44	Olly	Wow, that was awful.	EN

10.1.3 Pairs' Convergent Reception to the Learning Companions

In hindsight, we might have predicted that a divergence in the receptiveness to the virtual learning companions could dampen the impact for student pairs. Perhaps less predictable, however, was my discovery that pairs' collaboration may have diminished the impact of the virtual learning companions even when the two students shared a similar degree of receptiveness to the learning companions. While some of these groups did exhibit equitable collaboration in their dialogue, others did not. This variation mirrors that of the Baseline

Condition (as described above in Section 10.1.1), so there is little evidence to support the hypothesis that the virtual learning companions had any impact on the student pairs. As noted before, this contrasts to the demonstrated impact of virtual learning companions on individual students (see Chapter 6).

Companion Group 02 consisted of two girls who collaborated equitably. They did so despite one of them, Jessica, scoring low on general self-efficacy (21), while the other one, Kiara, had a higher self-efficacy score (31).[18] Table 10-6 illustrates how they both had a light-hearted reaction to the virtual learning companions. In utterance 16, Jessica disparages the virtual learning companion, and Kiara follows soon after by laughing about the scripted responses. After more laughing and disparagement by both girls, this excerpt ends with Jessica asking, "Why is this taking so long?" and Kiara responding with "I know, right?" That last utterance—"I know, right?"—was popular for both girls throughout the gameplay experience. Throughout their collaboration, whether they were interacting with the virtual learning companions or playing the baseline Engage gameplay, each girl consistently positioned her partner as a valued contributor. In many ways, Companion Group 02 stands as an exemplar for equitable collaboration, especially when we consider the differences in self-efficacy. Were we to see similar patterns in most pairs in the Companion Condition, it would be reasonable to conclude that the virtual learning companions had a wonderfully positive impact. That is not the case, however, and so I suspect that the positive behaviors that Jessica and Kiara exhibit are a testament to each of them as individuals and to their collective friendship, and not to their interaction with the virtual learning companions.

---

[18] In this regard, they are somewhat similar to Baseline Group 03, although their collaboration was more obviously equitable (perhaps because Jessica and Kiara are friends of the same age, whereas the girl with low self-efficacy in Baseline Group 03, Ana, was a year younger than her partner, Benita).

Table 10-6. Sample of dialogue from Companion Group 02
An equitable collaboration

#	Speaker	Utterance	Tag
11	Jessica	What will happen if they hated us? I wonder what would happen if we made them hate us?	QU
12	Kiara	[READING OUT LOUD]	
13	Kiara	What? That's not really a [inaudible].	EN
14	Kiara	All right - middle one, yeah.	SU
15	Kiara	Yay.	EP
16	Jessica	She's so weird, what the...	EN
17	Kiara	[READING OUT LOUD]	OM
18	Kiara	[Laughing] The rest of the answer choices are so sad.	EN
19	Jessica	I know, right?	YD
20	Kiara	[READING OUT LOUD]	OM
21	Jessica	[READING OUT LOUD]	OM
22	Jessica	[LAUGHING] Why are these other ones so sad?	QU
23	Kiara	[LAUGHING] What if you had picked the middle one?	QU
24	Jessica	[READING OUT LOUD]	OM
25	Kiara	Yay, we don't care.	EN
26	Jessica	[READING OUT LOUD]	OM
27	Kiara	I don't want to make one of them hate me.	SU
28	Jessica	[READING OUT LOUD]	OM
29	Jessica	[READING OUT LOUD]	OM
30	Jessica	Why is this so long?	QU
31	Kiara	I know, right?	YD

To illustrate that point, we can look at another pair of girls, Companion Group 11. As in Companion Group 02, one of the girls, Querita, scored much lower on general self-efficacy (27) than her partner, Paulina (35). As seen in Table 10-7, Paulina and Querita seem to share a business-like approach to the interaction with the virtual learning companions. Most of their utterances consist of strategizing about which scripted response will lead to the most positive reaction by the virtual learning companions. Paulina is at the controls, and Querita is advising which option to choose, mimicking the classic driver-navigator paradigm of pair programming. Just looking at their dialogue during the learning companion episodes could lead one to think that Companion Group 11 exhibited the behaviors of an equitable collaboration.

Table 10-7. Sample of dialogue from Companion Group 11
An inequitable collaboration

#	Speaker	Utterance	Tag
4	Paulina	I wouldn't say that one, because I don't feel like...it's just as dangerous.	SU
5	Querita	Don't do that one.	SU
6	Querita	Maybe that one?	QU
7	Paulina	Okay. Wow.	EN
8	Querita	Why does she have those glasses on?	QU
9	Querita	Oh, I know what they're talking about now. Before we got to the stairs.	SU
10	Querita	That one or that one.	SU
11	Paulina	I'm gonna do this one.	QU
12	Paulina	I wonder if he'll still want to chat with us.	SU
13	Paulina	This one?	QU
14	Querita	Yeah, I think so.	YD
15	Paulina	Wow he likes this a lot. Look at that. [inaudible].	EP
16	Paulina	He likes this more.	EP
17	Querita	The middle one.	SU
18	Paulina	Probably. That's what I was thinking too.	P
19	Querita	Like already at the top.	EP
20	Paulina	No the top is, like, right there.	SU
21	Querita	Why is it not going up?	QU
22	Paulina	Probably.	YD
23	Paulina	I just don't want him to [inaudible]. Like, I want him to line up with this one. He needs to talk faster.	EN
24	Querita	He's like dot, dot, dot, and then he went.	EN
25	Querita	Middle one.	SU
26	Querita	He didn't tell us anything.	EN

Yet a full review of their collaboration during the entire gameplay tells a different story. Paulina, the partner with higher general self-efficacy, made over two thirds of the pair's suggestions across the entire gameplay (their Suggestion Distribution of 0.6708 was among the highest of all groups, in contrast to Companion Group 02's Suggestion Distribution of 0.5489, which was among the lowest). Furthermore, the distribution was almost exactly opposite for questions, with Querita being the one to ask over two thirds of the pair's questions.

Prior research has described how such collaborations, in which one student utters significantly more suggestions (or commands) than their partner while their partner utters significantly more questions, are inherently inequitable (Lewis & Shah, 2015). The next section will continue the discussion on this pair's inequitable collaboration (as well as Companion Group 02's more equitable collaboration).

10.2 Gender and Equitable Collaboration

We now turn to this dissertation's second research question (RQ2), and specifically the question of gender differences. I originally hypothesized that the addition of the virtual learning companions would have an especially positive impact for girls. Meanwhile, implicit in that hypothesis was my belief that the virtual learning companions would have no worse than a neutral impact for boys. As described in the related work of Chapter 2 (specifically Section 2.1.3), middle school girls often benefit from virtual learning companions, whereas middle school boys show a tendency to abuse the virtual agents (although I have found little evidence suggesting that such behavior has any negative effect on the boys' learning). Chapter 9 has already described how the corpus of data fails to support the hypotheses connected to RQ2. Namely, the quantitative analyses showed no positive impact of the virtual learning companions for girls, while evidence exists that the virtual learning companions may, in fact, have negatively impacted the collaborative behaviors of boys. For a fuller understanding of the equitable collaboration by girls, Section 10.2.1 will continue the discussion of the groups described in the preceding section. Section 10.2.2 will then address the potentially negative issues found among many of the boys in the Companion Condition.

10.2.1 Girls

As described above in Section 10.1.3, even when two girls in a pair shared a mutual response to the virtual learning companions (i.e., there was no divergence in how receptive the two girls each were to the learning companion episodes), pairs varied in how equitably they collaborated during the rest of the gameplay. Whereas Companion Group 02 exhibited behaviors indicating equitable collaboration (a Suggestion Distribution of 0.5489 and a Question Distribution of 0.5849), the metrics indicate a less equitable collaboration for Companion Group 11 (a

Suggestion Distribution of 0.6709 and a Question Distribution of 0.6761). As mentioned above, I furthermore believe that the equitable collaboration seen in Companion Group 2 has little to do with the addition of the virtual learning companions; I suspect that these two girls might have collaborated in a similarly equitable manner without the virtual learning companions.

10.2.2 Boys

Chapter 9 first mentioned the negative outcomes among boys in the Companion Condition, and this chapter has already shown some evidence of it in Section 10.1.2, with the example of Companion Group 05 (Table 10-4). Namely, four of the six male-male pairs in the Companion Condition exhibited verbally abusive behavior toward each other. Section 10.1.2 provided evidence that the abuse may have stemmed from their abuse of the virtual learning companions, and Table 10-4 showed an example of how one student's abuse of the virtual learning companions led his partner to respond by verbally abusing him in the moment. Table 10-8 illustrates how the same pattern of behavior continued for this pair later in their gameplay. Now that they are past the episode with the virtual learning companions, Lenny has found a new virtual character to disparage: the players' avatar. In utterance 161, he jokingly notes that the avatar has no lips, and then tries to bring up the episode with the virtual learning companions. Manny shuts him down and continues with the current problem-solving task.

Unlike Lenny, Manny has prior computer science experiences and greater interest in computer science. He takes the gameplay seriously and rarely shows interest in off-task conversation. Lenny, however, has a greater proclivity for off-task conversation, perhaps because he does not share as much of an interest in the computational problem-solving of the game as Manny does. The episodes with the virtual learning companions—outside the baseline Engage game environment and involving no computational problem-solving—encourage his off-task tendencies. In this way, the learning companion episodes exacerbate existing tensions within the pair. To be clear, it is difficult to determine if Manny's abuse of Lenny stems directly from Lenny's abuse of the virtual characters, or if Manny would have verbally abused Lenny for any excessive off-task conversation, even if Manny's off-task conversation did not include any form of abuse. In the case of the former scenario, one might imagine that Lenny's abuse

of the virtual characters either inspires Manny to be abusive himself (i.e., he might not have even considered such behavior otherwise) or grants him some perceived permission to be abusive (i.e., since Lenny opened the door on abusive behavior, Manny feels permitted to walk through and be abusive himself). In any case, Manny's verbal abuse of Lenny was undeniably detrimental to Lenny's learning experience; midway through the term of instruction, he asked his teacher if he could switch partners.

Table 10-8. Sample of dialogue from Companion Group 05
In which Manny's verbal abuse follows Lenny's abuse of the virtual learning companion

#	Speaker	Utterance	Tag
160	Manny	Stop, don't move. Move forward a little bit.	SU
161	Lenny	Where's his lips? Oh, my goodness. Where's his lips?	EN
162	Manny	[Lenny].	EN
163	Lenny	How are we gonna talk to those kids if we don't have any lips?	QU
164	Manny	Shut up. You talk too much.	EN, AB
165	Manny	[inaudible] one. Okay, hold on.	SU
166	Lenny	You need to push the thing.	SU
167	Manny	So we gotta...move right. Move forward, move forward, move right, move forward. Yeah, let's try that.	SU
168	Manny	Okay, so...Move right, move for...move forward. Forward. Right. Move forward. Wait a minute.	SU
169	Manny	Okay, you ready, [Lenny]?	QU

For some other pairs of boys, their mutually abusive conversation may have been less detrimental, perhaps even acting as a form of rapport. Companion Group 12, for example, consisted of two boys—Ricky and Sonny—who both took great pleasure in abusing the virtual learning companions. Toward the beginning of the learning companion episode, Ricky issues the following directive: "Call him stupid and useless. He's just trying to make himself feel good about himself. So stupid!" The pair proceed to choose all the most abusive scripted responses, reveling in the virtual learning companion's sad reactions. At the end of the learning companion episode, Sonny laughs about how their behavior might hypothetically impact the narrative of the game: "I can only imagine. Johnny's being held hostage by like a group of people. Not by Mr. Mustache. He's like, nah, I don't want to talk to nobody. He...Mr. Mustache

was like, Mustache, gun to his head, and he's like no I'd rather die than talk to these jerks who hate everyone."

The two boys seem to take on the roles of "jerks", in an interesting form of "transformational play" that narrative-centered games can afford (Barab, Pettyjohn, Gresalfi, Volk, & Solomou, 2012). When they finish the learning companion episode and return to the main gameplay, Ricky and Sonny continue to play the same playfully antagonistic roles (especially Ricky). Table 10-9 depicts their conversation at this point, when they discover that they need to re-do a challenge from the day before. Ricky proposes that they fight over who controls the mouse, and then aggressively takes over the gameplay. Sonny concludes by saying he is going to tune out until Ricky gets them back to where they were before.

Table 10-9. Sample of dialogue from Companion Group 12
In which the verbal abuse follows immediately after the episode with the virtual learning companion

#	Speaker	Utterance	Tag
60	Ricky	Hey! You want to fight about that or something, because like you're not giving me the mouse.	EN, AB
61	Sonny	Huh! So mean. What?	EN
62	Ricky	Oh no, not this again!	EN
63	Sonny	At least you know what to do. Click through the wall, through this. I guess we didn't save hard enough.	EP
64	Ricky	Are you kidding me?	EN
65	Sonny	No, no.	EN
66	Sonny	Move forward twice. Rotate left four times...	SU
67	Ricky	Let's go in the wall again!	SU
68	Sonny	No, no, no.	EN
69	Sonny	[Inaudible] I'm just going to leave you to do it because I'm pretty sure you memorized it.	SU

Throughout their gameplay experience following the learning companion episode, both boys utter verbal abuse directed at their partner, although Ricky does so twice as often as Sonny. While it is difficult to definitively connect their verbal abuse to their interactions with the virtual learning companions, we can note that neither boy uttered any verbal abuse during the gameplay on the day before the episode with the virtual learning companions. Furthermore, over the course of the gameplay session that began with Ricky's opening directive about how

they should interact with the virtual learning companion ("Call him stupid and useless…"), Ricky ended up calling *Sonny* "stupid" six times.

As we can see in these two examples, the Abuse tag was applied to a variety of utterances, ranging from playful to potentially hurtful[19]. Prior research has demonstrated that playful face-threat between two students can be associated with positive learning outcomes, depending on the context of the students' existing relationship (Ogan, Finkelstein, Walker, Carlson, & Cassell, 2012). I did not have a measure for assessing the prior relationship of each student dyad in this study (i.e., the extent to which the two students were already friends), but my classroom observations and my conversations with the teacher lead me to believe that all the students who made abusive utterances did have existing friendships with their partners. In this respect, one might view the instances of verbal abuse with somewhat less alarm.

That point aside, I believe now that designers of virtual learning companions should guard against the potential encouragement of abuse. Future versions of Adriana and Johnny should not include pre-scripted menu responses that disparage the virtual learning companions. For systems that allow users to enter their own responses (e.g., by typing or speaking), the designers of the dialogue system must think critically about how the system should respond to abusive inputs. As virtual agents become more ubiquitous, they will inevitably have an increased role in training children how to treat their fellow humans. Technology designers thus have a responsibility to consider the potential social effects, regardless of the virtual agents' intended roles and primary purposes.

---

[19] Note that all tags were applied with consideration of the conversational context. Thus, the application of the Abuse tag did not extend to include utterly benign utterances that only seem like insults when taken out of context.

10.3 Prior Computer Science Experience and Equitable Collaboration

In addition to gender, Research Question 2 sought to explore differences in prior computer science experiences, with respect to students' equitable collaboration and receptiveness to the virtual learning companions. We might hypothesize that unbalanced groups (i.e., those in which one partner has prior computer science experience and the other does not) would be less likely to feature equitable collaboration. Indeed, the quantitative analysis in Section 9.3 provide support for that hypothesis. This section will bring further light by recalling a few of the collaborations that this chapter has already examined.

Unbalanced groups (in terms of prior computer science experience) exhibited inequitable collaboration regardless of the gender of the students. Companion Group 05 (Manny and Lenny), described above in Section 10.2.2, was one of two unbalanced groups in the Companion Condition. The learning companion episodes clearly failed to promote equitable collaboration in that group. Indeed, we can see in that example how the mismatch in prior computer science experience (and interest) may have led to the pattern of abuse; Manny, with prior interest in computer science, wanted to take the gameplay seriously, and he lashed out at Lenny whenever the latter made off-task comments. Earlier, in Section 10.1.1, I described the inequitable collaboration of another unbalanced group, Baseline Group 08. The two girls in that group, Cecilia and Dora, demonstrated much more supportive dialogue than Manny and Lenny, to such an extent that I originally mistook it for equitable collaboration. Yet Cecila, the student with no prior computer science experience, consistently deferred to the more experienced Dora. These two groups, so different in many ways, had almost identical Suggestion Distributions (among the most unbalanced groups); in both cases, the student with prior computer science made over 70 percent of the total suggestions. Collaborative dialogue might vary in many ways between boys and girls, but certain patterns of inequitable collaboration can emerge among any groups in which one student has more prior experience than the other.

Of course, inequitable collaboration can also emerge in balanced groups, such as those in which both partners have prior computer science experiences and interests. Section 10.1.2

introduced Companion Group 10, the sole mixed-gender pair in the Companion Condition. Both students in this group, Natasha and Olly, had significant prior computer science experiences. Yet Olly dominates the dialogue throughout the learning companion episodes (as seen earlier, in Table 10-5) and the rest of the gameplay. Even though Natasha sits next to him as a potentially valuable resource, Olly often misses out on her expertise as he instead self-narrates his own thinking. Table 10-10 shows an example of the teacher coming over to offer support to this group, asking questions to lead the two students through understanding the given challenge. Natasha jumps ahead to the ultimate solution while Olly is still processing the teacher's leading questions, so the teacher then tells Natasha to explain to Olly[20].

As with the two other groups above (Companion Group 05 and Baseline Group 08), this group had one of the five most unbalanced Suggestion Distributions, with Olly making over 70 percent of all the suggestions. Unlike those other two groups, however, the partner in this group making relatively few suggestions, Natasha, did not seem to be at risk of feeling less capable than her partner. Indeed, it is important to note that, in addition to the gender difference, Natasha is two years older than Olly. The relative infrequency of her suggestions—and utterances in general—might (at least in part) be due to her annoyance at being partnered with a younger boy, who himself shows great confidence in his ability and opinions. The two students exhibit minimal rapport throughout the gameplay, with Olly talking a lot but rarely expecting any reply from Natasha, and Natasha making minimal effort to contribute to the conversation. The teacher later explained to me that she intervened not because she noticed this group was stuck, but because she observed the unfortunate dynamic between Natasha and Olly. The intervention from the teacher seems to have helped, even if only temporarily, but it is difficult to conceive how an intervention involving virtual learning companions might have a positive impact on the collaboration of such a pair of students.

---

[20] The program that they are examining here is the one depicted in Figure 8-1.

Table 10-10. Sample of dialogue from Companion Group 10
An inequitable collaboration between two students with prior computer science experience

#	Speaker	Utterance	Tag
100	Teacher	Okay, if you press the red pressure pad, what is it going to turn to?	
101	Olly	It's going to go to one. If you press the blue, it goes to two.	SU
102	Teacher	Which would be what?	
103	Olly	Four.	SU
104	Teacher	One plus two?	
105	Olly	It's a…	OM
106	Natasha	All of them except yellow.	SU
107	Teacher	There you go.	
108	Olly	So, it's red, blue and green. So, there's four...	SU
109	Teacher	You got it. [Natasha]'s already got it. She already knows.	
110	Olly	Which one's which?	QU
111	Teacher	[Natasha], explain to him why you're not doing yellow.	
112	Natasha	Yellow sets him back to zero, but it's seven, so you have to go...	SU
113	Olly	Wait, what? Why would yellow send him back to zero? Does it say zero?	QU
114	Natasha	Yeah	YD
115	Teacher	So, if that's zero, so, [Olly], if you went on all the other pressure pads except for yellow, they add up to seven, which is what you're trying to get to. But if you clicked on the yellow, then it just throws it back to zero, which is not going to open unless it's based on seven.	
116	Olly	Oh.	YD
117	Teacher	She read it more quicker than, you know...she already clicked out, but she got it.	

10.4 Collaborative Dialogue in Game-based Learning Versus Pair Programming

I based this dissertation study's hypotheses on prior research on equitable collaboration (Lewis & Shah, 2015; Shah et al., 2014), yet that research was conducted in the context of a different type of learning environment. Although also analyzing middle school students learning computer science, that prior research involved pair programming with Scratch and Logo. Analyzing equitable collaboration within game-based learning presents additional challenges.

In this section, I note four issues that have affected the analysis for this study of equitable collaboration within the Engage game-based learning environment.

10.4.1 Rhetorical Questions and Hedged Suggestions

Chapter 8 described how the dialogue act classification was modified because the two independent taggers had difficulty coming to agreement on certain utterances. Specifically, it was difficult to differentiate rhetorical questions from non-rhetorical questions, and it was difficult differentiate assertive suggestions from guess suggestions. While the dialogue of two students pair programming might include rhetorical questions (*e.g.*, "What would happen if we tried X?") and hedged suggestions (*e.g.*, "Maybe we should try X."), I argue that the nature of gameplay leads to a richer variety of these types of soft utterances.

In a game-based learning environment like Engage, students often confront unfamiliar scenarios that require them first to comprehend the given challenge before they can begin solving it. For large portions of the gameplay, therefore, students are operating out of uncertainty but also with an element of playfulness. In these situations, they know that their partners cannot have definite answers (since the partners are also seeing the given challenge for the first time), so it might seem pointless to ask direct questions; instead they ask seemingly rhetorical questions, perhaps hoping that their partner will have some answer to propose. Likewise, they do not have a firm enough understanding of the task to utter outright commands to their partner, yet the nature of play makes it unnecessary to hedge their suggestions as mere guesses; their suggestions often come out somewhere in the gray area between the two extremes.

An excerpt from Baseline Group 12 (Table 10-11) exemplifies this phenomenon. At the beginning of the session (before this excerpt), Henry informs Georgia that he is quickly going to finish his homework for another class, so Georgia essentially plays solo for almost 5 minutes. Then Henry tunes back in with a suggestion in utterance 36. That first utterance of his may seem assertive, but Henry knows that Georgia has already been exploring the challenge for several minutes and thus has a greater understanding of what their task is (even if she does not yet know how to solve the task). Utterance 36, therefore, might be closer to a guess, with the intent of gaining clarification from Georgia; it is difficult to say for sure. Georgia later asks

a question in utterance 39. At this point, she knows that Henry has not been paying attention for the past several minutes and thus probably does not have an answer, so it seems to be a rhetorical question. Yet surely she hopes that her partner might have at least a guess in answer to her question. In short, the nature of the collaborative dialogue in this corpus makes it difficult to reliably differentiate between certain types of utterances.

Table 10-11. Sample of dialogue from Baseline Group 12
With rhetorical questions and hedged suggestions

#	Speaker	Utterance	Tag
36	Henry	We need to program them.	SU
37	Georgia	It *is* programmed.	SU
38	Georgia	[Reading Program] "Receive open from pressure pad".	SU
39	Georgia	*What* pressure pad?	QU
40	Georgia	Oh, my God.	EN
41	Georgia	Okay, maybe if we turn the yellow off it will work.	SU
42	Georgia	Okay, I hate this game officially.	EN
43	Henry	Are they all turned?	QU

10.4.2 Partners Valuing Different Aspects of the Game-based Learning Experience

Section 10.2.2 described how the two boys in Companion Group 05 (Lenny and Manny) diverged in their approach to the gameplay. Lenny made a lot of jokes, seemingly embracing the collaborative activity as a playful experience; Manny took the in-game tasks more seriously, seemingly considering the collaborative activity to be a learning experience. Part of the beauty of game-based learning is that it can incorporate both play and learning. Game-based learning can thus appeal to many learners, who might embrace different aspects of the rich interaction that it can afford. As evidenced by Lenny and Manny, however, when two partners diverge in their approaches to a game-based learning experience, conflict can arise. While conflicts can arise in other collaborative learning activities, such as pair programming, I argue that game-based learning presents a unique case in this regard.

10.4.3 Various Forms of Sharing the Game Controls

Another key way that the game-based learning in this study differed from pair programming was in how two students in a pair shared the game controls. Oftentimes in pair programming activities such as in the aforementioned prior research (Shah et al., 2014), the two students take turns being the driver—in full possession of the controls—and the navigator. If the navigator grabs the controls at any point before a scheduled switch, it is considered an exception and generally a negative behavior. In theory, the students playing Engage follow the driver-navigator paradigm, but neither the teacher nor our research team enforces it. Consequently, each student pair ends up negotiating how the collaboration will work. One popular strategy is for one of the students to control the keyboard (and thus the character controls) while the other student controls the mouse (with which a player can use the block-based programming interface). Companion Group 05 is one group employing this strategy. While Lenny controls their avatar with the keyboard, Manny writes the program in the programming interface (Table 10-12). Since they both have some form of game controls, they both issue commands to their partner (e.g., Lenny in utterance 148, Manny in utterance 154). When students share the controls in this manner, it is impossible to replicate some of the analyses into equitable collaboration that other researchers have performed on dialogue from students pair programming (Lewis & Shah, 2015).

In general, it might seem advantageous to allow both students to have some form of game control, as it better ensures that both students will be full contributors at all times. Yet if the students do not switch control of the programming interface (i.e., if one student consistently operates the mouse while the other student operates the keyboard), the student at the keyboard might miss out on being a full contributor *to the computational thinking element of the game*. The nature of the gameplay leads to multiple dimensions of equitable collaboration to consider. While any form of contribution might have some benefit to students, we must monitor students' collaborative gameplay to ensure that its equitability extends to the true learning objectives.

Table 10-12. Sample of dialogue from Companion Group 05

In which Manny controls the mouse (programming) while Lenny controls the keyboard (player movement)

#	Speaker	Utterance	Tag
134	Lenny	[Manny], you made the dude [inaudible].	EN
135	Manny	[Lenny], be quiet, please.	SU
136	Manny	You're so dumb.	EN, AB
137	Manny	Okay, you ready?	QU
138	Manny	Wait, we should do...move right.	SU
139	Lenny	No.	SU
140	Manny	No, yeah. Move forward. Move forward. Move forward. Move forward and to the right again. And then, wait one minute.	SU
141	Manny	All right, [Lenny], you ready?	QU
142	Manny	If you fall off...	EN
143	Lenny	All right, I got the thing. I...	EP
144	Manny	[Lenny]!	EN
145	Lenny	I didn't do it.	EN
146	Manny	No, we have to do it again.	EN
147	Manny	Okay, so maybe we need one more move forward right there and [inaudible] right there.	SU
148	Lenny	Move right again. Move right again.	SU
149	Manny	No.	SU
150	Lenny	Right here.	SU
151	Manny	No.	SU
152	Manny	Okay, you ready?	QU
153	Manny	Get up there. Yes! Go, go, go.	EP
154	Manny	All right, up the steps. Up the steps.	SU
155	Lenny	Can't we sprint?	QU
156	Manny	No, don't sprint.	SU

10.4.4 Directing Utterances toward Characters in the Game

Finally, Section 10.1.1 introduced Baseline Group 08 (Cecilia and Dora) with an excerpt of transcript (Table 10-3) that is, on first glance, difficult to understand. Once one realizes that several of Cecilia's utterances are directed not to Dora but rather to her avatar, the dialogue becomes clearer. Analyses of pair programming rarely need to consider the possibility that the speaker is directing a given utterance at a non-human, but many such instances might arise in the collaborative dialogue of students playing a game like Engage with virtual characters and avatars.

Chapter 11 CONCLUSION

In this dissertation study, I have aimed to further the research on how virtual learning companions can support the collaborative learning of middle school students. In considering this overarching goal, we should note that supporting collaborative learning holds the potential to produce both immediate benefits and long-term benefits for students. For one, the support immediately aids the students on the given learning task. In addition to this immediate scaffolding, the support can also potentially develop collaboration skills that will aid them in future learning and non-learning activities. Improving our understanding of how to support collaborative learning therefore carries great merit.

11.1 Summary

The study specifically investigated the potential impact of *virtual learning companions*, given the context of a *narrative-centered learning environment* and the subject domain of *computer science education*. Chapter 2 provided a discussion on the relevant literature in those three areas of research. To varying extents, this study can potentially make contributions to each of those research areas, with a focus on supporting collaboration.

Crucially, I conducted this study using data collected from in-school classrooms, rather than from a lab or out-of-school activity. One motivation for doing so concerns the student population. I wanted to study how virtual learning companions can impact all middle school students learning computer science. Were I to use data collected from an out-of-school activity, I would be analyzing pairs of students who do not necessarily represent the larger population. The analyses and discussion in Chapter 4 illustrate how in-school initiatives hold significant benefits over out-of-school ones—particularly so for middle school computer science education.

While the first research question for this study addressed the impact of the virtual learning companions on students' collaboration in general, the second research question examined how the impacts might vary based on student characteristics, such as gender. As described in Chapter 5, female students might particularly benefit from collaboration when

using ENGAGE. Chapter 6, meanwhile, provided early evidence to back the hypothesis that virtual learning companions can have an especially positive impact on female students using ENGAGE (in a pilot study in which students did not collaborate). I therefore hypothesized that combining the benefits of collaboration and virtual learning companions would have a particularly positive impact on female students. The impact on male students, of course, also warranted attention.

In Chapter 7, I described the process of iteratively designing the virtual learning companion episodes, modifying the design based on feedback from students. In the end, I developed four episodes, each approximately 5 minutes long and each designed for students to encounter when they reach a specific location in the gameplay. The episodes with the two virtual learning companions, Adriana and Johnny, specifically reinforce good collaboration practices. The companions tell stories about their successes and struggles collaborating, encouraging the students to persevere and support their partners.

The data for my proposed study comes from students who played ENGAGE at Carnage Middle School (in Raleigh, NC) in 2016. As detailed in Section 8.1, pairs of students who participated in Spring 2016 played only the ENGAGE game, without any support from the virtual learning companions; pairs of students who participated in Fall 2016 played ENGAGE and also experienced the four episodes with the virtual learning companions. I collected audio and video data from all sessions in both these conditions. To address my research questions, I analyzed the audio data by annotating the transcripts with a custom dialogue act classification scheme, detailed in Section 8.2.

Chapter 9 presented the quantitative analyses from my investigation into the research questions. In short, I found no evidence of the virtual learning companions having any sort of consistent impact on the equitable collaboration of student pairs. Chapter 10 then provided qualitative analyses that begin to explain that lack of impact. If there is one overarching conclusion from it all, it might be this: collaboration in game-based learning is a complex phenomenon, even without trying to understand how we might leverage virtual learning companions to help make that collaboration more equitable.

11.2 Contributions

This dissertation has produced greater insight into how to bring greater equity to collaborative learning experiences. As enumerated earlier in Chapter 1, it makes the following novel contributions:

C1. *Dialogue corpus of collaborative learning within a game-based learning environment.* In conducting this research, I created a dialogue corpus that consists of transcripts from 43 sessions of collaborative gameplay. Each transcript is annotated with dialogue act labels. Not including the transcribed dialogue, I collected over 150 hours of additional audio and video ripe for future analysis.

C2. *Dialogue act classification scheme.* I developed a custom dialogue act classification scheme for analyzing equitable collaboration within game-based learning. To develop the scheme, I referenced related work on student dialogue within pair programming (Shah et al., 2014) as well as long-standing work in computational linguistics (Stolcke et al., 2000).

C3. *Dialogue analysis of equity in pair programming, within the context of collaborative game-based learning.* Emerging lines of research are analyzing the dialogue of children collaboratively learning computer science, with a focus on equitable collaboration (Lewis & Shah, 2015). This dissertation adds to that research with the unique perspective of examining children collaboratively learning computer science within a game-based learning environment.

C4. *Empirical analysis of virtual learning companions' impact on collaborative learning.* Emerging lines of research are also starting to investigate how intelligent tutors can support collaborative learning (Kumar & Rosé, 2011). This dissertation makes a novel contribution by examining virtual learning companions supporting the collaborative learning of middle school students.

C5. *Design recommendations for virtual learning companions that support collaborative learning.* The work produces insights into how designers of intelligent learning environments can incorporate virtual learning companions into a system to support collaborative learning activities. It clarifies the challenges that such activities present,

in contrast to learning activities in which individual students interact with the computer-based learning environment.

C6. *Replication of analyses relating to gender differences on the impact of virtual learning companions, with novel application to collaborative learning.* Existing research suggests that virtual learning companions might particularly benefit female students (Arroyo et al., 2011). This dissertation study investigates how that hypothesis holds for virtual learning companions who support pairs of students in a collaborative learning activity. The results suggest that the particularly positive benefits for girls might not, in fact, extend to collaborative learning experiences.

C7. *Empirical analyses relating to prior experience differences on the impact of virtual learning companions.* Existing research has indicated that students' prior computer science experience may impact how well they collaborate (Katira et al., 2004). This dissertation adds to that research, and examines differences in the impact of the virtual learning companions based on students' prior computing experiences.

11.3 Future Work

Given its exploratory nature, I hoped for this dissertation to produce a launching pad for future inquiry, rather than merely a list of established contributions. In fact, the research reported here might broaden the possibilities for future research well beyond what I originally envisioned. As with any research that produces null results for many of its hypotheses, the findings from this dissertation could recommend a wide variety of future research directions. To bring clarity on which directions might hold greatest promise, one might ask oneself a series of branching questions. Firstly, is there any merit to pursuing further research on virtual learning companions that support collaborative learning? If the answer is 'yes', one might then ask if there is merit to continuing such research in the specific application of *game-based* collaborative learning. Alternatively, if the answer to the first question is 'no', one might then ask if game-based collaborative learning is significantly different than other forms of collaborative learning—in other words, is the collaboration that happens within game-based learning a unique form of collaboration? Figure 11-1 visualizes this branching set of questions.
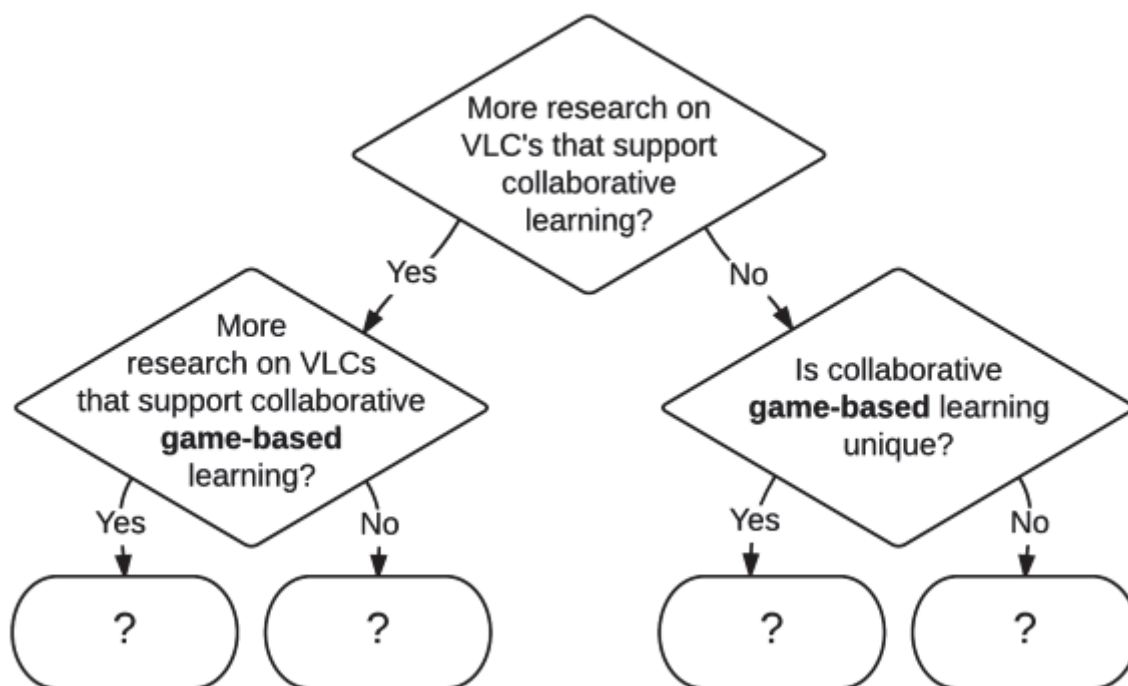
Figure 11-1. Many possible directions for future research

In truth, I find it surprisingly difficult to answer that first, fundamental question. As described in Chapter 10, paired collaboration brings significant complexity, with the potential for inequitable dynamics emerging in a diversity of forms. From my classroom observations, I found that the teacher's ability to correct inequitable collaboration far outstripped that of the virtual learning companions, as seen in the example of Natasha and Olly (Figure 10-10). Might we therefore better serve students in this regard by investing in teacher coaching and the development of curricular strategies, rather than putting time and resources into technological additions that have little value added? If so, any number of questions arise as to what those analog interventions might entail, and one could start investigating that line of research with a sense of urgency.

Regarding this dissertation's focus, one would soon want to clarify whether the paired collaboration that occurs in a game-based learning environment like Engage is distinct from that which occurs in other forms of paired collaboration, notably pair programming. I believe

this to be the case, but the question merits comparison studies of the two types of paired collaboration. If game-based collaboration is, in fact, a unique phenomenon, much more research is needed for understanding how inequitable patterns emerge between students playing a game, how we can design game-based learning environments that mitigate such patterns, and how we can coach classroom teachers to address inequitable collaboration with the greatest effectiveness. If, on the other hand, future research establishes a tight correlation between the collaborative dialogue of students in a game-based learning environment and the collaborative dialogue of students in pair programming activities, one can envision a wealth of synergistic future research. Researchers have only recently started the inquiry into examining equitable collaboration in pair programming; research on equitable collaboration in games like Engage could dovetail with that ongoing research while leveraging the unique advantages of game-based learning.[21] Indeed, despite this dissertation's focus on virtual learning companions, it could motivate substantial future research unrelated to virtual agents.

The above argument—that we might put aside research into how virtual learning companions can support collaborative learning—takes a short-term view, rooted in an urgency to address today's educational inequities. Indeed, the particularly stark inequities present in computer science education demand that designers of computer science learning initiatives think critically about the children who are currently in school classrooms, developing the skills and attitudes that will drive their future career and educational choices. I consider that concern to be of prime importance. Yet the counter argument—that we should continue research into how virtual learning companions can support collaborative learning—is oriented to the long-term view, and forward thinking can have great merit as well. To wit, we know that technology improves rapidly, and we are also seeing technology becoming increasingly dominant in classrooms, so one could argue that we must now be investigating the potential of technological innovations that emerging classroom technology will afford in the not-so-distant future.

---

[21] See Section 2.2 in the Related Work chapter for a review of the unique benefits that game-based learning environments such as Engage can afford.

This dissertation's findings lead me to believe that, for virtual learning companions to truly help address inequitable collaboration (beyond what a classroom teacher might provide), they will require technology that is not currently present in most classrooms. As seen among the twenty-two pairs included in this study, patterns of inequitable collaboration emerge in wide variety, and cannot be predicted merely by demographic information that one might obtain on a pre-survey. Future research could continue with analyses of collaborative dialogue, with the end-goal of a system being able to recognize a specific form of inequitable collaboration within a pair of students, via multi-modal inputs, and then have the virtual learning companion adaptively intervene. I designed this study's intervention optimistic for the potential of a one-size-fits-all approach, which has worked previously in studies in which individuals interact with virtual learning companions. While I did not theorize that one-size-fits-all might ultimately match a tailored, adaptive intervention in a hypothetical comparison study, I thought that it might provide some overall benefit at low cost. The results, however, suggest that a tailored, adaptive intervention may be the only viable solution to this line of research.

If one does seek to explore the potential of virtual learning companions to support collaborative learning, one might want to ask that secondary question: should one pursue that line of research in the context of game-based learning? I can imagine that one might answer 'no' to this question, arguing that other collaborative learning activities (*e.g.*, pair programming) might produce student dialogue in which inequitable collaboration is more easily detected. The existing and ongoing work on analyzing equitable collaboration in pair programming could inform the models that a virtual learning companion uses to intervene effectually. On the other hand, game-based learning might still present unique promise for this line of research on virtual learning companions, despite the potentially greater challenge of detecting equitable collaboration. At many times in this document, including in Section 2.1.1 of the Related Work chapter and in Chapter 7, I have mentioned the goal of designing virtual learning companions that can support students over long-term interventions, and the potential of narrative-centered learning environments to achieve that goal. In the end, this study has not clarified how virtual learning companions can support students over several sessions (as

supporting students over a single session remains a challenge). Yet if one retains hope for virtual learning companions to support collaborative learning, I argue for pursuing research into how they can support collaborative learning over multiple sessions, and for leveraging the power of narrative within game-based environments to accomplish that long-term goal.

REFERENCES

Adams, J. C. (2010). Scratching middle schoolers' creative itch. In *Proceedings of the 41st ACM Technical Symposium on Computer Science Education (SIGCSE '10)* (pp. 356–360). ACM. Retrieved from http://portal.acm.org/citation.cfm?id=1734385

Aimeur, E., & Frasson, C. (1996). Analyzing a new learning strategy according to different knowledge levels. *Computers and Education*, *27*(2), 115–127. http://doi.org/10.1016/0360-1315(96)00018-8

Anderson, N., & Gegg-Harrison, T. (2013). Learning Computer Science in the "Comfort Zone of Proximal Development." In *Proceeding of the 44th ACM Technical Symposium on Computer Science Education - SIGCSE '13* (pp. 495–500). New York, New York, USA: ACM Press. http://doi.org/10.1145/2445196.2445344

Aritajati, C., Rosson, M. B., Pena, J., Cinque, D., & Segura, A. (2015). A Socio-Cognitive Analysis of Summer Camp Outcomes and Experiences. In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education - SIGCSE '15* (pp. 581–586). New York, New York, USA: ACM Press. http://doi.org/10.1145/2676723.2677292

Arpaci-Dusseau, A., Astrachan, O., Barnett, D., Bauer, M., Carrell, M., Dovi, R., … Osborne, R. B. (2013). Computer Science Principles : Analysis of a Proposed Advanced Placement Course. In *Proceedings of the 44th ACM Technical Symposium on Computer Science Education - SIGCSE '13* (pp. 251–256).

Arroyo, I., Burleson, W., Tai, M., Muldner, K., & Woolf, B. P. (2013). Gender Differences in the Use and Benefit of Advanced Learning Technologies for Mathematics. *Journal of*

Educational Psychology, *105*(4), 957–969.

Arroyo, I., Woolf, B. P., Cooper, D. G., Burleson, W., & Muldner, K. (2011). The Impact of Animated Pedagogical Agents on Girls' and Boys' Emotions, Attitudes, Behaviors and Learning. In *2011 IEEE 11th International Conference on Advanced Learning Technologies* (pp. 506–510). IEEE. http://doi.org/10.1109/ICALT.2011.157

Arroyo, I., Woolf, B. P., Royer, J. M., & Tai, M. (2009). Affective Gendered Learning Companions. In *Proceedings of the 14th International Conference on Artificial Intelligence in Education - AIED '09* (pp. 41–48). http://doi.org/10.3233/978-1-60750-028-5-41

Astrachan, O., Barnes, T., Garcia, D. D., Paul, J., Simon, B., & Snyder, L. (2011). CS Principles: Piloting a New Course at National Scale. In *Special Interest Group on Computer Science Education (SIGCSE) Conference* (pp. 397–398).

Axelsson, A., Anderberg, E., & Haake, M. (2013). Can Preschoolers Profit from a Teachable Agent Based Play-and-Learn Game in Mathematics? In *Proceedings of the 16th International conference on Artificial Intelligence in Education - AIED '13* (pp. 289–298).

Baker, R. S. J. D., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be Frustrated than Bored: The Incidence, Persistence, and Impact of Learners' Cognitive–Affective States During Interactions with Three Different Computer-based Learning Environments. *International Journal of Human-Computer Studies*, *68*(4), 223–241. http://doi.org/10.1016/j.ijhcs.2009.12.003

Barab, S. A., Gresalfi, M., & Ingram-Goble, A. (2010). Transformational Play: Using Games

to Position Person, Content, and Context. *Educational Researcher*, *39*(7), 525–536. http://doi.org/10.3102/0013189X10386593

Barab, S. A., Pettyjohn, P., Gresalfi, M., Volk, C., & Solomou, M. (2012). Game-based curriculum and transformational play: Designing to meaningfully positioning person, content, and context. *Computers & Education*, *58*(1), 518–533. http://doi.org/10.1016/j.compedu.2011.08.001

Barker, L. J., O'Neill, M., & Kazim, N. (2014). Framing Classroom Climate for Student Learning and Retention in Computer Science. In *Proceedings of the 45th ACM Technical Symposium on Computer Science Education (SIGCSE '14)* (pp. 319–324). New York, New York, USA: ACM Press. http://doi.org/10.1145/2538862.2538959

Basawapatna, A. R., Koh, K. H., Repenning, A., Webb, D. C., & Marshall, K. S. (2011). Recognizing Computational Thinking Patterns. In *Proceedings of the 42nd ACM Technical Symposium on Computer Science Education - SIGCSE '11* (p. 245). New York, New York, USA: ACM Press. http://doi.org/10.1145/1953163.1953241

Basu, S., Dukeman, A., Kinnebrew, J. S., Biswas, G., & Sengupta, P. (2014). Investigating Student Generated Computational Models of Science. In *International Conference of the Learning Sciences* (pp. 1097–1101).

Baylor, A. L. (2011). The Design of Motivational Agents and Avatars. *Educational Technology Research and Development*, *59*(2), 291–300. http://doi.org/10.1007/s11423-011-9196-3

Baylor, A. L., & Ebbers, S. J. (2003a). Evidence that multiple agents facilitate greater learning. *Artificial Intelligence in Education: Shaping the Future of Learning Through*

Intelligent Technologies, 377–379.

Baylor, A. L., & Ebbers, S. J. (2003b). The pedagogical agent split-persona effect: When two agents are better than one. In *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications* (pp. 459–462).

Baylor, A. L., & Kim, Y. (2003). Validating Pedagogical Agent Roles: Expert , Motivator , and Mentor. In *World Conference on Educational Multimedia, Hypermedia and Telecommunications* (pp. 463–466).

Bell, S., Frey, T., & Vasserman, E. (2014). Spreading the Word: Introducing Pre-Service Teachers to Programming. In *Proceedings of the 45th ACM Technical Symposium on Computer Science Education - SIGCSE '14* (pp. 187–192). New York, New York, USA: ACM Press. http://doi.org/10.1145/2538862.2538963

Bian, Y., Yang, C., Guan, D., Xiao, S., Gao, F., Shen, C., & Meng, X. (2016). Effects of Pedagogical Agent's Personality and Emotional Feedback Strategy on Chinese Students' Learning Experiences and Performance. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16* (pp. 433–444). New York, New York, USA: ACM Press. http://doi.org/10.1145/2858036.2858351

Bickmore, T. W., & Cassell, J. (2001). Relational agents: A Model and Implementation of Building User Trust. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '01* (pp. 396–403). New York, New York, USA: ACM Press. http://doi.org/10.1145/365024.365304

Bickmore, T. W., Pfeifer, L. M., & Jack, B. W. (2009). Taking the Time to Care: Empowering Low Health Literacy Hospital Patients with Virtual Nurse Agents. In

Proceedings of the 27th International Conference on Human Factors in Computing Systems - CHI '09 (pp. 1265–1274). New York, New York, USA: ACM Press. http://doi.org/10.1145/1518701.1518891

Bickmore, T. W., & Picard, R. W. (2005). Establishing and Maintaining Long-Term Human-Computer Relationships. *ACM Transactions on Computer-Human Interaction*, *12*(2), 293–327. http://doi.org/10.1145/1067860.1067867

Bickmore, T. W., Schulman, D., & Yin, L. (2009). Engagement vs Deceit: Virtual Humans with Human Autobiographies. In *Proceedings of the 9th Annual International Conference on Intelligent Virtual Agents - IVA 2009* (pp. 6–19).

Biswas, G., Jeong, H., Kinnebrew, J. S., Sulcer, B., & Roscoe, R. (2010). Measuring Self-Regulated Learning Skills Through Social Interactions in a Teachable Agent Environment. *Research and Practice in Technology Enhanced Learning*, *5*(2), 123–152. http://doi.org/10.1142/S1793206810000839

Biswas, G., Leelawong, K., Schwartz, D. L., & Vye, N. (2005). Learning by Teaching: A New Agent Paradigm for Educational Software. *Applied Artificial Intelligence*, *19*(3–4), 363–392. http://doi.org/10.1080/08839510590910200

Brophy, S., Biswas, G., Katzlberger, T., Bransford, J., & Schwartz, D. L. (1999). Teachable agents: Combining insights from learning theory and computer science. *Learning Technology*, *50*, 21–28. http://doi.org/10.1.1.20.1877

Buffum, P. S., Boyer, K. E., Wiebe, E. N., Mott, B. W., & Lester, J. C. (2015). Mind the Gap: Improving Gender Equity in Game-based Learning Environments with Learning Companions. In *Proceedings of the 17th International conference on Artificial*

Intelligence in Education - AIED '15 (pp. 64–73).

Buffum, P. S., Frankosky, M. H., Boyer, K. E., Wiebe, E. N., Mott, B. W., & Lester, J. C. (2015). Leveraging Collaboration to Improve Gender Equity in a Game-based Learning Environment for Middle School Computer Science. In *RESPECT - IEEE Special Technical Community on Broadening Participation* (pp. 1–8).

Buffum, P. S., Frankosky, M. H., Boyer, K. E., Wiebe, E. N., Mott, B. W., & Lester, J. C. (2016a). Collaboration and Gender Equity in Game-Based Learning for Middle School Computer Science. *Computing in Science & Engineering*, *18*(2), 18–28. http://doi.org/10.1109/MCSE.2016.37

Buffum, P. S., Frankosky, M. H., Boyer, K. E., Wiebe, E. N., Mott, B. W., & Lester, J. C. (2016b). Empowering All Students: Closing the CS Confidence Gap with an In-School Initiative for Middle School Students. In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education - SIGCSE '16* (pp. 382–387). New York, New York, USA: ACM Press. http://doi.org/10.1145/2839509.2844595

Buffum, P. S., Lobene, E. V., Frankosky, M. H., Boyer, K. E., Wiebe, E. N., & Lester, J. C. (2015). A Practical Guide to Developing and Validating Computer Science Knowledge Assessments with Application to Middle School. In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education (SIGCSE '15)* (pp. 622–627). New York, New York, USA: ACM Press. http://doi.org/10.1145/2676723.2677295

Buffum, P. S., Martínez-Arocho, A. G., Frankosky, M. H., Rodríguez, F. J., Wiebe, E. N., & Boyer, K. E. (2014). CS Principles Goes to Middle School: Learning How to Teach "Big Data." In *Proceedings of the 45th ACM Technical Symposium on Computer*

Science Education - SIGCSE '14 (pp. 151–156). New York, New York, USA: ACM Press. http://doi.org/10.1145/2538862.2538949

Burke, Q., & Kafai, Y. B. (2012). The Writers ' Workshop for Youth Programmers Digital Storytelling with Scratch in Middle School Classrooms. In *Proceedings of the 43rd ACM Technical Symposium on Computer Science Education - SIGCSE '12* (pp. 433–438).

Burleson, W., & Picard, R. W. (2007). Evidence for Gender Specific Approaches to the Development of Emotionally Intelligent Learning Companions. *IEEE Intelligent Systems Journal*, *22*(4), 62–69.

Burns, R., Pollock, L., & Harvey, T. (2012). Integrating Hard and Soft Skills: Software Engineers Serving Middle School Teachers. In *Proceedings of the 43rd ACM Technical Symposium on Computer Science Education - SIGCSE '12* (pp. 209–2014). New York, New York, USA: ACM Press. http://doi.org/10.1145/2157136.2157199

Carlson, R., Keiser, V., Matsuda, N., Koedinger, K. R., & Rosé, C. P. (2012). Building a conversational simstudent. In S. A. Cerri, W. J. Clancey, G. Papadourakis, & K. Panourgia (Eds.), *Proceedings of the 11th international conference on Intelligent Tutoring Systems - ITS '12* (Vol. 7315, pp. 563–569). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-30950-2

Carter, E. J., Blank, G., & Walz, J. (2012). Bringing the Breadth of Computer Science to Middle Schools. In *Proceedings of the 43rd ACM Technical Symposium on Computer Science Education - SIGCSE '12* (pp. 203–208). New York, New York, USA: ACM Press. http://doi.org/10.1145/2157136.2157198

Cassell, J., & Jenkins, H. (1998). *From Barbie to Mortal Kombat: Gender and Computer Games*. Cambridge MA: MIT Press.

Castellano, G., Leite, I., Pereira, A., Martinho, C., Paiva, A., & McOwan, P. W. (2009). It's all in the game: Towards an affect sensitive and context aware game companion. In *2009 Humaine Association Conference on Affective Computing and Intelligent Interaction*. Ieee. http://doi.org/10.1109/ACII.2009.5349558

Castro, B., Diaz, T., Gee, M., Justice, R., Kwan, D., Seshadri, P., & Dodds, Z. (2016). MyCS at 5: Assessing a Middle-years CS Curriculum. In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education - SIGCSE '16* (pp. 558–563). New York, New York, USA: ACM Press. http://doi.org/10.1145/2839509.2844643

Chase, C. C., Chin, D. B., Oppezzo, M. A., & Schwartz, D. L. (2009). Teachable Agents and the Protégé Effect: Increasing the Effort Towards Learning. *Journal of Science Education and Technology*, *18*(4), 334–352. http://doi.org/10.1007/s10956-009-9180-4

Chou, C.-Y., Chan, T.-W., & Lin, C.-J. (2003). Redefining the Learning Companion: the Past, Present, and Future of Educational Agents. *Computers & Education*, *40*(3), 255–269. http://doi.org/10.1016/S0360-1315(02)00130-6

Chowanda, A., Flintham, M., Blanchfield, P., & Valstar, M. (2016). Playing with Social and Emotional Game Companions (pp. 85–95). Springer International Publishing. http://doi.org/10.1007/978-3-319-47665-0_8

Chris, B., Rebecca, L., & Lennart, N. (2011). Player Typology in Theory and Practice. In *Proceedings of DiGRA 2011 Conference: Think Design Play* (Vol. 6).

Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. (2013). *Digital Games for Learning:*

A Systematic Review and Meta-Analysis (Vol. May 2013).

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, *20*(1), 37–46. http://doi.org/10.1177/001316446002000104

Cohen, J. (1988). *Statisical Power Analysis for the Behavioral Sciences*. (L. Erlbaum, Ed.) (2nd Editio).

Coleman, B., & Lang, M. (2012). Collaboration Across the Curriculum: A Disciplined Approach to Developing Team Skills. In *Proceedings of the 43rd ACM Technical Symposium on Computer Science Education (SIGCSE '12)* (pp. 277–282).

Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A Systematic Literature Review of Empirical Evidence on Computer Games and Serious Games. *Computers & Education*, *59*(2), 661–686. http://doi.org/10.1016/j.compedu.2012.03.004

Cooper, S., Rodger, S. H., Schep, M., Stalvey, R. H., & Dann, W. (2015). Growing a K-12 Community of Practice. In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education - SIGCSE '15* (pp. 290–295). New York, New York, USA: ACM Press. http://doi.org/10.1145/2676723.2677255

Cordar, A., Robb, A., Wendling, A., Lampotang, S., White, C., & Lok, B. (2015). Virtual Role-Models: Using Virtual Humans to Train Best Communication Practices for Healthcare Teams (pp. 229–238). Springer International Publishing. http://doi.org/10.1007/978-3-319-21996-7_23

Correnti, R., Stein, M. K., Smith, M. S., Scherrer, J., McKeown, M. G., Greeno, J. G., &

Ashley, K. (2015). Improving Teaching at Scale: Design for the Scientific Measurement and Learning of Discourse Practice. In *Socializing Intelligence Through Academic Talk and Dialogue* (pp. 315–332). American Educational Research Association. http://doi.org/10.3102/978-0-935302-43-1_25

D'Mello, S. K., Lehman, B., Sullins, J., Daigle, R., Combs, R., Vogt, K., … Graesser, A. C. (2010). A Time for Emoting: When Affect-Sensitivity Is and Isn't Effective at Promoting Deep Learning. In *Proceedings of the 10th International Conference on Intelligent Tutoring Systems - ITS '10* (pp. 245–254). Berlin, Heidelberg. http://doi.org/10.1007/978-3-642-13388-6

De Jean, J., Upitis, R., Koch, C., & Young, J. (1999). The Story of Phoenix Quest: How Girls Respond to a Prototype Language and Mathematics Computer Game. *Gender and Education*, *11*(2), 207–223. http://doi.org/10.1080/09540259920708

Dede, C. (2009). Immersive interfaces for engagement and learning. *Science (New York, N.Y.)*, *323*(5910), 66–9. http://doi.org/10.1126/science.1167311

Dede, C. (2010). Comparing frameworks for 21st century skills. *21st Century Skills*, 51–76.

Denner, J., Bean, S., & Werner, L. (2005). Girls Creating Games: Challenging Existing Assumptions About Game Content. In *Proceedings of DiGRA 2005 Conference: Changing Views - Worlds in Play*.

DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A., … Morency, L. (2014). SimSensei Kiosk : A Virtual Human Interviewer for Healthcare Decision Support. *International Conference on Autonomous Agents and Multi-Agent Systems*, (1), 1061–1068.

Dewey, J. (1910). *How We Think*. Boston: D.C. Heath & Company. Retrieved from

http://catalogue.nla.gov.au/Record/2311170

Doering, A., Veletsianos, G., & Yerasimou, T. (2008). Conversational Agents and Their

Longitudinal Affordances on Communication and Interaction - ProQuest. *Journal of*

Interactive Learning Research, *19*(2), 251–270. Retrieved from

http://search.proquest.com.proxy-

library.ashford.edu/docview/211265423/fulltextPDF/14F73508A55349D1PQ/4?account

id=32521

Dong, W., Ehrlich, K., Macy, M. M., & Muller, M. (2016). Embracing Cultural Diversity:

Online Social Ties in Distributed Workgroups. In *Proceedings of the 2016 conference*

on Computer Supported Cooperative Work CSCW '16 (pp. 274–287).

http://doi.org/10.1145/2818048.2835198

Dweck, C. (2006). *Mindset: The New Psychology of Success*. Random House LLC.

Eagan, B. R., Rogers, B., Serlin, R., Ruis, A. R., Irgens, G. A., & Shaffer, D. W. (2017). Can

We Rely on IRR? Testing the Assumptions of Inter-Rater Reliability. In *CSCL 2017*

Proceedings (pp. 529–532).

Esper, S., Wood, S. R., Foster, S. R., Lerner, S., & Griswold, W. G. (2014). Codespells: how

to design quests to teach java concepts. *Journal of Computing Sciences in Colleges*,

29(4), 114–122. Retrieved from http://dl.acm.org/citation.cfm?id=2591468.2591490

Falkner, K., Falkner, N. J. G., & Vivian, R. (2013). Collaborative Learning and Anxiety: A

phenomenographic study of collaborative learning activities. In *Proceedings of the 44th*

ACM Technical Symposium on Computer Science Education (SIGCSE '13) (pp. 227–

232).

Finkelstein, S., Yarzebinski, E., Vaughn, C., & Ogan, A. (2013). The Effects of Culturally
Congruent Educational Technologies on Student Achievement. *Proceedings of the 16th
International Conference on Artificial Intelligence in Education - AIED '13*, 493–502.

Forbes-Riley, K., Litman, D., Huettner, A., & Ward, A. (2005). Dialogue-Learning
Correlations in Spoken Dialogue Tutoring. *Artificial Intelligence in Education -
Supporting Learning through Intelligent and Socially Informed Technology*, *125*, 225–
232.

Foster, S. R., Esper, S., & Griswold, W. G. (2013). From competition to metacognition:
designing diverse, sustainable educational games. In *Proceedings of the SIGCHI
Conference on Human Factors in Computing Systems - CHI '13* (pp. 99–108). New
York, New York, USA: ACM Press. http://doi.org/10.1145/2470654.2470669

Franklin, D., Conrad, P., Aldana, G., & Hough, S. (2011). Animal Tlatoque: Attracting
Middle School Students to Computing Through Culturally-Relevant Themes. In
*Proceedings of the 42nd ACM Technical Symposium on Computer Science Education -
SIGCSE '11* (pp. 453–458). ACM.

Franklin, D., Hill, C., Dwyer, H. A., Hansen, A. K., Iveland, A., & Harlow, D. B. (2016).
Initialization in Scratch: Seeking Knowledge Transfer. In *Proceedings of the 47th ACM
Technical Symposium on Computing Science Education - SIGCSE '16* (pp. 217–222).
New York, New York, USA: ACM Press. http://doi.org/10.1145/2839509.2844569

Franklin, D., Kiefer, B., Laird, C., Lopez, F., Pham, C., Suarez, J., … Almeida-Tanaka, P.
(2013). Assessment of computer science learning in a scratch-based outreach program.

In *Proceeding of the 44th ACM Technical Symposium on Computer Science Education - SIGCSE '13* (pp. 371–376). New York, New York, USA: ACM Press. http://doi.org/10.1145/2445196.2445304

Frankosky, M. H., Wiebe, E. N., Buffum, P. S., & Boyer, K. E. (2015). Spatial Ability and Other Predictors of Gameplay Time: Understanding Barriers to Learning in Game-based Virtual Environments. In *AERA Applied Research in Immersive Environments for Learning SIG*. Chicago, IL.

Gee, J. P. (2003). What Video Games Have to Teach Us About Learning and Literacy. *Computers in Entertainment*. http://doi.org/10.1145/950566.950595

Girard, S., Chavez-Echeagaray, M. E., Gonzalez-Sanchez, J., Hidalgo-Pontet, Y., Zhang, L., Burleson, W., & Vanlehn, K. (2013). Defining the Behavior of an Affective Learning Companion in the Affective Meta-tutor Project Design of the Affective Intervention. In *Proceedings of the 16th International conference on Artificial Intelligence in Education - AIED '13* (pp. 21–30).

Girard, S., & Johnson, H. (2010). What do children favor as embodied pedagogical agents? In V. Aleven, J. Kay, & J. Mostow (Eds.), *Proceedings of the 10th international conference on Intelligent Tutoring Systems - ITS '10* (Vol. 6094, pp. 307–316). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-13388-6

Goode, J., & Margolis, J. (2011). Exploring Computer Science. *ACM Transactions on Computing Education*, *11*(2), 1–16. http://doi.org/10.1145/1993069.1993076

Goodman, B., Soller, A., Linton, F., & Gaimari, R. (1998). Encouraging Student Reflection and Articulation using a Learning Companion. *International Journal of Artificial*

Intelligence in Education, *9*, 237–255.

Gorriz, C. M., & Medina, C. (2000). Engaging Girls with Computers through Software Games. *Communications of the ACM*, *43*(1), 42–49. http://doi.org/10.1145/323830.323843

Grawemeyer, B., Johnson, H., Brosnan, M., Ashwin, E., & Benton, L. (2012). Developing an Embodied Pedagogical Agent with and for Young People with Autism Spectrum Disorder. In *Intelligent Tutoring Systems* (Vol. 7315, pp. 262–267). http://doi.org/10.1007/978-3-642-30950-2_33

Gray, J., Haynie, K., Packman, S., Boehm, M., Crawford, C., & Muralidhar, D. (2015). A Mid-Project Report on a Statewide Professional Development Model for CS Principles. In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education - SIGCSE '15* (pp. 380–385). New York, New York, USA: ACM Press. http://doi.org/10.1145/2676723.2677306

Grover, S., Pea, R., & Cooper, S. (2014). Expansive Framing and Preparation for Future Learning in Middle-School Computer Science. In *International Conference of the Learning Sciences (ICLS) Conference* (pp. 992–996).

Grover, S., Pea, R. D., & Cooper, S. (2016). Factors Influencing Computer Science Learning in Middle School. In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education - SIGCSE '16* (pp. 552–557). New York, New York, USA: ACM Press. http://doi.org/10.1145/2839509.2844564

Gulz, A., Haake, M., & Silvervarg, A. (2011). Extending a Teachable Agent with a Social Conversation Module: Effects on Student Experiences and Learning. In *Proceedings of*

the 15th International Conference on Artificial Intelligence in Education - AIED '11 (pp. 106–114). Springer-Verlag. Retrieved from http://dl.acm.org/citation.cfm?id=2026506.2026524

Guzdial, M., Ericson, B., Mcklin, T., & Engelman, S. (2012). A Statewide Survey on Computing Education Pathways and Influences : Factors in Broadening Participation in Computing. *Proceedings of the 9th ACM Conference on International Computing Education Research - ICER '12*, 143–150.

Hall, J. V., Wyeth, P. A., & Johnson, D. (2014). Instructional Objectives to Core-Gameplay. In *Proceedings of the First ACM SIGCHI Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '14* (pp. 121–130). New York, New York, USA: ACM Press. http://doi.org/10.1145/2658537.2658696

Hanks, B., Fitzgerald, S., McCauley, R., Murphy, L., & Zander, C. (2011). Pair programming in education: a literature review. *Computer Science Education*, *21*(2), 135–173. http://doi.org/10.1080/08993408.2011.579808

Hart, J., Sutcliffe, A. G., & De Angeli, A. (2013). Love it or Hate it! Interactivity and User Types. In *Proceedings of the 31st Annual ACM Conference of Human Factors in Computing Systems - CHI '13* (p. 2059). New York, New York, USA: ACM Press. http://doi.org/10.1145/2470654.2481282

Harteveld, C., & Sutherland, S. C. (2015). The Goal of Scoring: Exploring the Role of Game Performance in Educational Games. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15* (pp. 2235–2244). New York, New York, USA: ACM Press. http://doi.org/10.1145/2702123.2702606

Hickey, D. T., Ingram-Goble, A. A., & Jameson, E. M. (2009). Designing Assessments and Assessing Designs in Virtual Educational Environments. *Journal of Science Education and Technology*, *18*(2), 187–208. http://doi.org/10.1007/s10956-008-9143-1

Hietala, P., & Niemirepo, T. (1998). The Competence of Learning Companion Agents. *International Journal of Artificial Intelligence in Education*, *9*(2), 178–192.

Hill, C., Dwyer, H. A., Martinez, T., Harlow, D., & Franklin, D. (2015). Floors and Flexibility: Designing a Programming Environment for 4th-6th Grade Classrooms. In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education - SIGCSE '15* (pp. 546–551). New York, New York, USA: ACM Press. http://doi.org/10.1145/2676723.2677275

Hyde, J., Kiesler, S., Hodgins, J. K., & Carter, E. J. (2014). Conversing with Children: Cartoon and Video People Elicit Similar Conversational Behaviors. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems - CHI '14* (pp. 1787–1796). New York, New York, USA: ACM Press. http://doi.org/10.1145/2556288.2557280

Iacobelli, F., & Cassell, J. (2007). Ethnic Identity and Engagement in Embodied Conversational Agents. In *Proceedings of the Seventh Annual International Conference on Intelligent Virtual Agents - IVA 2007* (pp. 57–63).

Isbister, K., Flanagan, M., & Hash, C. (2010). Designing Games for Learning: Insights from Conversations with Designers. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems - CHI '10* (pp. 2041–2044). New York, New York, USA: ACM Press. http://doi.org/10.1145/1753326.1753637

ISTE. (2016). *ISTE Standards: Students*. *International Society for Technology in Education*. Retrieved from http://www.iste.org/standards/iste-standards/standards-for-students

Jackson, G. T., & McNamara, D. S. (2013). Motivation and performance in a game-based intelligent tutoring system. *Journal of Educational Psychology*, *105*(4), 1036–1049.

Jenson, J., de Castell, S., & Fisher, S. (2007). Girls Playing Games: Rethinking stereotypes. In *Proceedings of the 2007 Conference on Future Play - Future Play '07* (p. 9). New York, New York, USA: ACM Press. http://doi.org/10.1145/1328202.1328205

Johnson, W. L. (2010). Serious use of a serious game for language learning. *International Journal of Artificial Intelligence in Education*, *20*, 175–195. http://doi.org/10.3233/JAI-2010-0006

Kafai, Y. B. (1995). Minds in Play: Computer Game Design as a Context for Children's Learning. Retrieved from http://dl.acm.org/citation.cfm?id=527173

Kafai, Y. B., & Burke, Q. (2013). The Social Turn in K-12 Programming : Moving from Computational Thinking to Computational Participation. In *Proceedings of the 44th ACM Technical Symposium on Computer Science Education - SIGCSE '13* (pp. 603–608).

Karoly, L. A., & Panis, C. W. A. (2004). *The 21st Century at Work: Forces Shaping the Future Workforce and Workplace in the United States. Summary*. *RAND Corporation*. Retrieved from http://www.rand.org/pubs/monographs/MG164/

Katira, N., Williams, L., Wiebe, E. N., Miller, C., Balik, S., & Gehringer, E. (2004). On Understanding Compatibility of Student Pair Programmers. *ACM SIGCSE Bulletin*. http://doi.org/10.1145/1028174.971307

Kay, J. S., Moss, J. G., Engelman, S., & McKlin, T. (2014). Sneaking in Through the Back Door: Introducing K-12 Teachers to Robot Programming. In *Proceedings of the 45th ACM Technical Symposium on Computer Science Education - SIGCSE '14* (pp. 499–504). New York, New York, USA: ACM Press. http://doi.org/10.1145/2538862.2538972

Ketelhut, D. J., Nelson, B. C., Clarke, J., & Dede, C. (2010). A multi-user virtual environment for building and assessing higher order inquiry skills in science. *British Journal of Educational Technology*, *41*(1), 56–68. http://doi.org/10.1111/j.1467-8535.2009.01036.x

Kim, J. M., Hill, R. W., Technologies, C., Lane, H. C., Forbell, E., Core, M., … Hart, J. (2009). BiLAT: A game-based environment for practicing negotiation in a cultural context. *International Journal of Artificial Intelligence in Education*, *19*, 289–308. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-78649543291&partnerID=40&md5=5e5c25a6939c7e398134844d858f19f4

Kim, J. M., Hill Jr., R. W., Durlach, P. J., Lane, C., Forbell, E., Core, M., … Hart, J. (2009). BiLAT: A game-based environment for practicing negotiation in a cultural context. *International Journal of Artificial Intelligence in Education*, *19*, 289–308. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-78649543291&partnerID=40&md5=5e5c25a6939c7e398134844d858f19f4

Kim, Y., & Baylor, A. L. (2006a). A Social-Cognitive Framework for Pedagogical Agents as Learning Companions. *Educational Technology Research and Development*, *54*(6), 569–596. http://doi.org/10.1007/s11423-006-0637-3

Kim, Y., & Baylor, A. L. (2006b). Pedagogical Agents as Learning Companions: The Role of Agent Competency and Type of Interaction. *Educational Technology Research and Development*, *54*(3), 223–243.

Kim, Y., Wei, Q., Xu, B., Ko, Y., & Ilieva, V. (2007). MathGirls: Toward Developing Girls' Positive Attitude and Self-Efficacy through Pedagogical Agents. In *Proceedings of the Thirteenth International Conference on AI in Education - AIED '07* (pp. 119–126). Retrieved from http://portal.acm.org/citation.cfm?id=1563601.1563625&coll=GUIDE&dl=GUIDE&CFID=58070637&CFTOKEN=64604507

Koh, K. H., Repenning, A., & Motter, P. (2013). Will it Stick? Exploring the Sustainability of Computational Thinking Education Through Game Design. In *Proceedings of the 44th ACM Technical Symposium on Computer Science Education - SIGCSE '13* (pp. 597–602).

Kumar, R., Ai, H., Beuth, J. L., & Rosé, C. P. (2010). Socially Capable Conversational Tutors Can be Effective in Collaborative Learning Situations. In V. Aleven, J. Kay, & J. Mostow (Eds.), *Proceedings of the 10th International Conference on Intelligent Tutoring Systems - ITS '10* (Vol. 6094, pp. 156–164). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-13388-6

Kumar, R., & Rosé, C. P. (2011). Architecture for Building Conversational Agents that Support Collaborative Learning. *IEEE Transactions on Learning Technologies*, *4*(1), 21–34. http://doi.org/10.1109/TLT.2010.41

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for

Categorical Data. *Biometrics*, *33*(1), 159–174.

Larkins, D. B., Moore, J. C., Covington, L. R., & Rubbo, L. J. (2013). Application of the
Cognitive Apprenticeship Framework to a Middle School Robotics Camp. In
*Proceedings of the 44th ACM Technical Symposium on Computer Science Education -
SIGCSE '13* (pp. 89–94).

Lasserre, P. (2009). Adaptation of team-based learning on a first term programming class. In
ITiCSE'09 (Vol. 41, pp. 186–190). http://doi.org/10.1145/1595496.1562937

Lave, J., & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*.
Learning in doing (Vol. 95). http://doi.org/10.2307/2804509

Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a Unifying Social Cognitive
Theory of Career and Academic Interest, Choice, and Performance. *Journal of
Vocational Behavior*, *45*, 79–122. http://doi.org/10.1145/1340961.1340974

Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., & Bhogal, R. S.
(1997). The Persona Effect: Affective Impact of Animated Pedagogical Agents. In
*Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI
'97* (pp. 359–366). New York, New York, USA: ACM Press.
http://doi.org/10.1145/258549.258797

Lester, J. C., Spires, H. A., Nietfeld, J. L., Minogue, J., Mott, B. W., & Lobene, E. V. (2014).
Designing Game-Based Learning Environments for Elementary Science Education: A
Narrative-Centered Learning Perspective. *Information Sciences*, *264*, 4–18.
http://doi.org/10.1016/j.ins.2013.09.005

Levy, R. B.-B., & Ben-Ari, M. (2007). We Work So Hard and They Don't Use It:

Acceptance of Software Tools by Teachers. In *Proceedings of the 12th Conference on Innovation and Technology in Computer Science Education - ITiCSE '07* (Vol. 39, pp. 246–250). New York, New York, USA: ACM Press. http://doi.org/10.1145/1268784.1268856

Lewis, C. M. (2010). How Programming Environment Shapes Perception, Learning and Goals: Logo vs. Scratch. In *Proceedings of the 41st ACM Technical Symposium on Computer Science Education - SIGCSE '10* (pp. 346–350). Retrieved from http://portal.acm.org/citation.cfm?id=1734383

Lewis, C. M. (2011). Is pair programming more effective than other forms of collaboration for young students? *Computer Science Education*, *21*(2), 105–134. http://doi.org/10.1080/08993408.2011.579805

Lewis, C. M., & Shah, N. (2015). How Equity and Inequity Can Emerge in Pair Programming. In *Proceedings of the eleventh annual International Conference on International Computing Education Research - ICER '15* (pp. 41–50). New York, New York, USA: ACM Press. http://doi.org/10.1145/2787622.2787716

Lewis, C. M., Titterton, N., & Clancy, M. (2012). Using collaboration to overcome disparities in Java experience. In *Proceedings of the 9th ACM Conference on International Computing Education Research - ICER '12* (p. 79). New York, New York, USA: ACM Press. http://doi.org/10.1145/2361276.2361292

Liebenberg, J., Mentz, E., & Breed, B. (2012). Pair programming and secondary school girls' enjoyment of programming and the subject Information Technology (IT). *Computer Science Education*, *22*(3), 219–236. http://doi.org/10.1080/08993408.2012.713180

Lim, M. Y., Aylett, R., Ho, W. C., & Enz, S. (2009). A Socially-Aware Memory for Companion Agents. In *Proceedings of the 9th Annual International Conference on Intelligent Virtual Agents - IVA 2009* (pp. 20–26).

Liu, J., Lin, C.-H., Hasson, E. P., & Barnett, Z. D. (2011). Introducing Computer Science to K-12 Through a Summer Computing Workshop for Teachers. In *Proceedings of the 42nd ACM Technical Symposium on Computer Science Education - SIGCSE '11* (p. 389). New York, New York, USA: ACM Press. http://doi.org/10.1145/1953163.1953277

Liu, J., Lin, C.-H., Potter, P., Hasson, E. P., Barnett, Z. D., & Singleton, M. (2013). Going Mobile with App Inventor for Android: A One-Week Computing Workshop for K-12 Teachers. In *Proceeding of the 44th ACM Technical Symposium on Computer Science Education - SIGCSE '13* (p. 433). New York, New York, USA: ACM Press. http://doi.org/10.1145/2445196.2445324

Liu, J., Lin, C.-H., Wilson, J., Hemmenway, D., Hasson, E. P., Barnett, Z. D., & Xu, Y. (2014). Making Games a "Snap" with Stencyl – A Summer Computing Workshop for K-12 Teachers. In *Proceedings of the 45th ACM Technical Symposium on Computer Science Education - SIGCSE '14* (pp. 169–174). New York, New York, USA: ACM Press. http://doi.org/10.1145/2538862.2538978

Lomas, D., Patel, K., Forlizzi, J. L., & Koedinger, K. R. (2013). Optimizing challenge in an educational game using large-scale design experiments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13* (pp. 89–98). New York, New York, USA: ACM Press. http://doi.org/10.1145/2470654.2470668

Maloney, J., Peppler, K., Kafai, Y. B., Resnick, M., & Rusk, N. (2008). Programming by
Choice : Urban Youth Learning Programming with Scratch. In *Proceedings of the 39th
ACM Technical Symposium on Computer Science Education (SIGCSE '08)* (pp. 367–
371).

Maloney, J., Resnick, M., Rusk, N., Silverman, B., & Eastmond, E. (2010). The Scratch
Programming Language and Environment. *ACM Transactions on Computing Education
(TOCE)*, *10*, 16:1–16:15. http://doi.org/10.1145/1868358.1868363

Matsuda, N., Yarzebinski, E., Keiser, V., Raizada, R., Stylianides, G. J., & Koedinger, K. R.
(2012). Motivational factors for learning by teaching: the effect of a competitive game
show in a virtual peer-learning environment. In S. A. Cerri, W. J. Clancey, G.
Papadourakis, & K. Panourgia (Eds.), *Proceedings of the 11th international conference
on Intelligent Tutoring Systems - ITS '12* (Vol. 7315, pp. 101–111). Berlin, Heidelberg:
Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-30950-2

McCoy, J., Treanor, M., Samuel, B., Reed, A. A., Mateas, M., & Wardrip-Fruin, N. (2014).
Social story worlds with Comme il Faut. *IEEE Transactions on Computational
Intelligence and AI in Games*, *6*, 97–112. http://doi.org/10.1109/TCIAIG.2014.2304692

McGill, M. M., Decker, A., & Settle, A. (2015). Does Outreach Impact Choices of Major for
Underrepresented Undergraduate Students? In *Proceedings of the 11th ACM
Conference on International Computing Education Research - ICER '15* (pp. 71–80).
New York, New York, USA: ACM Press. http://doi.org/10.1145/2787622.2787711

McQuiggan, S. W., Lee, S., & Lester, J. C. (2007). Early Prediction of Student Frustration. In
A. C. R. Paiva, R. Prada, & R. W. Picard (Eds.), *Proceedings of the 2nd International*

Conference on Affective Computing and Intelligent Interaction (ACII '07) (pp. 698–709). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-540-74889-2

McQuiggan, S. W., Rowe, J. P., & Lester, J. C. (2008). The Effects of Empathetic Virtual Characters on Presence in Narrative-Centered Learning Environments. In *Proceeding of the Twenty-Sixth Annual CHI Conference on Human Factors in Computing Systems - CHI '08* (p. 1511). New York, New York, USA: ACM Press. http://doi.org/10.1145/1357054.1357291

Miksatko, J., Kipp, K. H., & Kipp, M. (2010). The Persona Zero-Effect : Evaluating Virtual Character Benefits on a Learning Task with Repeated Interactions. In *Proceedings of the 10th Annual International Conference on Intelligent Virtual Agents - IVA 2010* (pp. 475–481).

Mondragon, A. L., Nkambou, R., & Poirier, P. (2016). Towards an Effective Affective Tutoring Agent in Specialized Education. In *Proceedings of the 13th International Conference on Intelligent Tutoring Systems - Volume 9684* (pp. 402–408). Springer-Verlag New York, Inc. http://doi.org/10.1007/978-3-319-39583-8_48

Mott, B. W., Callaway, C., Zettlemoyer, L., Lee, S. Y., & Lester, J. C. (1999). Towards Narrative-Centered Learning Environments. In *Proceedings of the 1999 Fall Symposium on Narrative Intelligence* (pp. 78–82). Retrieved from http://www.aaai.org/Papers/Symposia/Fall/1999/FS-99-01/FS99-01-013.pdf

Murphy-Hill, E., & Murphy, G. C. (2011). Peer interaction effectively, yet infrequently, enables programmers to discover new tools. In *Proceedings of the ACM 2011*

conference on Computer supported cooperative work - CSCW '11 (pp. 405–414). New

York, New York, USA: ACM Press. http://doi.org/10.1145/1958824.1958888

Nasir, N. S., & Hand, V. (2008). From the Court to the Classroom: Opportunities for

Engagement, Learning, and Identity in Basketball and Classroom Mathematics. *Journal

of the Learning Sciences*, *17*(2), 143–179. http://doi.org/10.1080/10508400801986108

Ni, L. (2009). What Makes CS Teachers Change?: Factors Influencing CS Teachers'

Adoption of Curriculum Innovations. *ACM SIGCSE Bulletin*, *41*(1), 544–549.

http://doi.org/10.1145/1539024.1509051

Nickel, A., & Barnes, T. (2010). Games for CS Education: Computer-Supported

Collaborative Learning and Multiplayer Games. In *Foundations of Digital Games

(FDG) Conference* (pp. 274–276).

O'Brien, H. L. O., & Toms, E. G. (2010). The Development and Evaluation of a Survey to

Measure User Engagement. *Journal of the American Society for Information Science*,

61, 50–69. http://doi.org/10.1002/asi

O'Rourke, E., Haimovitz, K., Ballweber, C., Dweck, C., & Popović, Z. (2014). Brain Points:

A Growth Mindset Incentive Structure Boosts Persistence in an Educational Game. In

*Proceedings of the 32nd annual ACM conference on Human factors in computing

systems - CHI '14* (pp. 3339–3348). New York, New York, USA: ACM Press.

http://doi.org/10.1145/2556288.2557157

Ogan, A., Aleven, V., Kim, J. M., & Jones, C. (2010). Intercultural negotiation with virtual

humans: the effect of social goals on gameplay and learning. In *Proceedings of the 10th

international conference on Intelligent Tutoring Systems - ITS '10* (pp. 174–183). Berlin

Heidelberg. http://doi.org/10.1007/978-3-642-13388-6

Ogan, A., Finkelstein, S., Mayfield, E., D'Adamo, C., Matsuda, N., & Cassell, J. (2012). "Oh dear stacy!": social interaction, elaboration, and learning with teachable agents. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12* (pp. 39–48). New York, New York, USA: ACM Press. http://doi.org/10.1145/2207676.2207684

Ogan, A., Finkelstein, S., Walker, E., Carlson, R., & Cassell, J. (2012). Rudeness and rapport: insults and learning gains in peer tutoring. In S. A. Cerri, W. J. Clancey, G. Papadourakis, & K. Panourgia (Eds.), *Proceedings of the 11th international conference on Intelligent Tutoring Systems - ITS '12* (Vol. 7315, pp. 11–21). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-30950-2

Ogan, A., Walker, E., Baker, R. S. J. D., Rebolledo Mendez, G., Jimenez Castro, M., Laurentino, T., & De Carvalho, A. M. J. B. (2012). Collaboration in Cognitive Tutor Use in Latin America. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12* (pp. 1381–1390). New York, New York, USA: ACM Press. http://doi.org/10.1145/2207676.2208597

Oliveira, O. L., Nicoletti, M. C., & del Val Cura, L. M. (2014). Quantitative Correlation Between Ability to Compute and Student Performance in a Primary School. In *Proceedings of the 45th ACM Technical Symposium on Computer Science Education - SIGCSE '14* (pp. 505–510). New York, New York, USA: ACM Press. http://doi.org/10.1145/2538862.2538890

Orji, R., Mandryk, R. L., Vassileva, J., & Gerling, K. M. (2013). Tailoring persuasive health

games to gamer type. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13* (pp. 2467–2476). New York, New York, USA: ACM Press. http://doi.org/10.1145/2470654.2481341

Paas, F. G. W. C., Renkl, A., & Sweller, J. (2003). Cognitive Load Theory and Instructional Design: Recent Developments. *Educational Psychologist*, *38*(1), 1–4. http://doi.org/10.1207/S15326985EP3801_1

Papastergiou, M. (2009). Digital Game-Based Learning in High School Computer Science Education: Impact on Educational Effectiveness and Student Motivation. *Computers & Education*, *52*(1), 1–12. http://doi.org/10.1016/j.compedu.2008.06.004

Papert, S. (1980). *Mindstorms*. *Current Directions in Psychological Science* (Vol. 15). http://doi.org/10.1111/j.1467-8721.2006.00431.x

Pareto, L., Haake, M., Lindström, P., Sjödén, B., & Gulz, A. (2012). A teachable-agent-based game affording collaboration and competition: evaluating math comprehension and motivation. *Educational Technology Research and Development*, *60*(5), 723–751. http://doi.org/10.1007/s11423-012-9246-5

Partnership for 21st Century Learning. (2015). P21 Partnership for 21st Century Learning. *Partnership for 21st Century Learning*, 9. Retrieved from http://www.p21.org/documents/P21_Framework_Definitions.pdf

Pasternak, A., & Vahrenhold, J. (2012). Design and Evaluation of a Braided Teaching Course in Sixth Grade Computer Science Education. In *Proceedings of the 43rd ACM Technical symposium on Computer Science Education - SIGCSE '12* (pp. 45–50). New York, New York, USA: ACM Press. http://doi.org/10.1145/2157136.2157154

Pecune, F., Cafaro, A., Ochs, M., & Pelachaud, C. (2016). Evaluating Social Attitudes of a Virtual Tutor (pp. 245–255). Springer, Cham. http://doi.org/10.1007/978-3-319-47665-0_22

PISA 2015 Draft Collaborative Problem Solving Framework. (2015).

Piwek, P. (2008). Presenting Arguments as Fictive Dialogue. In *Proceedings of 8th Workshop on Computational Models of Natural Argument (CMNA08; in conjunction with ECAI 2008)*. Retrieved from http://oro.open.ac.uk/12143/

Plant, E. A., Baylor, A. L., Doerr, C. E., & Rosenberg-Kima, R. B. (2009). Changing Middle-School Students' Attitudes and Performance Regarding Engineering With Computer-Based Social Models. *Computers & Education*, *53*(2), 209–215. http://doi.org/10.1016/j.compedu.2009.01.013

Rader, E., Echelbarger, M., & Cassell, J. (2011). Brick by Brick: Iterating Interventions to Bridge the Achievement Gap with Virtual Peers. In *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems - CHI '11* (pp. 2971–2974). New York, New York, USA: ACM Press. http://doi.org/10.1145/1978942.1979382

Radermacher, A. D., Walia, G. S., & Rummelt, R. (2012). Improving Student Learning Outcomes with Pair Programming. In *Proceedings of the 9th ACM Conference on International Computing Education Research - ICER '12* (pp. 87–92). New York, New York, USA: ACM Press. http://doi.org/10.1145/2361276.2361294

Ramirez Uresti, J. A., & Du Boulay, B. (2004). Expertise, Motivation and Teaching in Learning Companion Systems. *International Journal of Artificial Intelligence in Education*, *14*, 67–106. Retrieved from http://webdia.cem.itesm.mx/juresti/

Ring, L., Barry, B., Totzke, K., & Bickmore, T. W. (2013). Addressing Loneliness and Isolation in Older Adults: Proactive Affective Agents Provide Better Support. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction* (pp. 61–66). http://doi.org/10.1109/ACII.2013.17

Rodger, S. H., Dalis, M., Gadwal, C., Hayes, J., Li, P., Wolfe, F., … Liang, L. (2012). Integrating Computing into Middle School Disciplines through Projects. In *Proceedings of the 43rd ACM Technical Symposium on Computer Science Education - SIGCSE '12* (pp. 421–426). New York, New York, USA: ACM Press. http://doi.org/10.1145/2157136.2157262

Rodger, S. H., Qin, H., & Hayes, J. (2009). Engaging Middle School Teachers and Students with Alice in a Diverse Set of Subjects. In *Proceedings of the 40th ACM Technical Symposium on Computer Science Education (SIGCSE '09)* (pp. 271–275).

Rowe, J. P., Shores, L. R., Mott, B. W., & Lester, J. C. (2010). Individual Differences in Gameplay and Learning. In *Proceedings of the Fifth International Conference on the Foundations of Digital Games - FDG '10* (pp. 171–178). New York, New York, USA: ACM Press. http://doi.org/10.1145/1822348.1822371

Rowe, J. P., Shores, L. R., Mott, B. W., & Lester, J. C. (2011). Integrating Learning, Problem Solving, and Engagement in Narrative-Centered Learning Environments. *International Journal of Artificial Intelligence in Education*, *21*, 115–133.

Sabourin, J. L., & Lester, J. C. (2014). Affect and engagement in game-based learning environments. *IEEE Transactions on Affective Computing*, *5*, 45–56. http://doi.org/10.1109/T-AFFC.2013.27

Sauppé, A., Szafir, D., Huang, C.-M., & Mutlu, B. (2015). From 9 to 90: Engaging Learners of All Ages. In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education - SIGCSE '15* (pp. 575–580). New York, New York, USA: ACM Press. http://doi.org/10.1145/2676723.2677248

Schanzer, E., Fisler, K., Krishnamurthi, S., & Felleisen, M. (2015). Transferring Skills at Solving Word Problems from Computing to Algebra Through Bootstrap. In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education - SIGCSE '15* (pp. 616–621). New York, New York, USA: ACM Press. http://doi.org/10.1145/2676723.2677238

Schofield, E., Erlinger, M., & Dodds, Z. (2014). MyCS: CS for Middle-Years Students and their Teachers. In *Proceedings of the 45th ACM Technical Symposium on Computer Science Education - SIGCSE '14* (pp. 337–342). New York, New York, USA: ACM Press. http://doi.org/10.1145/2538862.2538901

Schroeder, N. L., Adesope, O. O., & Gilbert, R. B. (2013). How Effective are Pedagogical Agents for Learning? A Meta-Analytic Review. *Journal of Educational Computing Research*, *49*(1), 1–39. http://doi.org/10.2190/EC.49.1.a

Segedy, J. R., Kinnebrew, J. S., & Biswas, G. (2015). Coherence Over Time: Understanding Day-to-Day Changes in Students' Open-Ended Problem Solving Behaviors (pp. 449–458). Springer, Cham. http://doi.org/10.1007/978-3-319-19773-9_45

Seiter, L. (2015). Using SOLO to Classify the Programming Responses of Primary Grade Students. In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education - SIGCSE '15* (pp. 540–545). New York, New York, USA: ACM Press.

http://doi.org/10.1145/2676723.2677244

Shaenfield, D. (2010). Arguing with peers: examining two kinds of discourse and their cognitive benefits. In *International Conference of the Learning Sciences (ICLS) Conference* (pp. 254–260). International Society of the Learning Sciences. Retrieved from http://dl.acm.org/citation.cfm?id=1854360.1854393

Shah, N., Lewis, C. M., & Caires, R. (2014). Analyzing Equity in Collaborative Learning Situations: A Comparative Case Study in Elementary Computer Science. In *International Conference of the Learning Sciences (ICLS) Conference* (pp. 495–502).

Shashaani, L. (1994). Gender Differences in Computer Experiences and Its Influence on Computer Attitudes. *Journal of Educational Computing Research*, *11*, 347–367.

Shi, L., Bickmore, T. W., & Edwards, R. (2015). A Feminist Virtual Agent for Breastfeeding Promotion (pp. 461–470). Springer, Cham. http://doi.org/10.1007/978-3-319-21996-7_49

Silvervarg, A., Raukola, K., Haake, M., & Gulz, A. (2012). The Effect of Visual Gender on Abuse in Conversation with ECAs. In *Proceedings of the 12th Annual International Conference on Intelligent Virtual Agents - IVA 2012* (pp. 153–160).

Sjödén, B., & Gulz, A. (2015). From Learning Companions to Testing Companions (pp. 459–469). Springer, Cham. http://doi.org/10.1007/978-3-319-19773-9_46

Smith, N., Sutcliffe, C., & Sandvik, L. (2014). Code club: Bringing Programming to UK Primary Schools through Scratch. In *Proceedings of the 45th ACM Technical Symposium on Computer Science Education - SIGCSE '14* (pp. 517–522). New York, New York, USA: ACM Press. http://doi.org/10.1145/2538862.2538919

Snow, E. L., Jackson, G. T., Varner, L. K., & McNamara, D. S. (2013). Expectations of Technology : A Factor to Consider in Game-Based Learning Environments. In *Proceedings of the 16th International Conference on Artificial Intelligence in Education - AIED '13* (pp. 359–368).

Starrett, C., Doman, M., Garrison, C., & Sleigh, M. (2015). Computational Bead Design: A Pilot Summer Camp in Computer Aided Design and 3D Printing for Middle School Girls. In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education - SIGCSE '15* (pp. 587–590). New York, New York, USA: ACM Press. http://doi.org/10.1145/2676723.2677303

Stolcke, A., Ries, K., Coccaro, N., Shriberg, E., Bates, R., Jurafsky, D., … Meteer, M. (2000). Dialogue Act Modeling for Automatic Tagging and Recognition of Conversational Speech. *Computational Linguistics*, *26*(3), 339–373. http://doi.org/10.1162/089120100561737

Stolee, K. T., & Fristoe, T. (2011). Expressing Computer Science Concepts Through Kodu Game Lab. In *Proceedings of the 42nd ACM Technical Symposium on Computer Science Education - SIGCSE '11* (pp. 99–104).

Sullivan, A., Grow, A., Mateas, M., & Wardrip-Fruin, N. (2012). The design of Mismanor. In *Proceedings of the International Conference on the Foundations of Digital Games - FDG '12* (p. 180). New York, New York, USA: ACM Press. http://doi.org/10.1145/2282338.2282374

Swartout, W., Traum, D., Artstein, R., Noren, D., Debevec, P., Bronnenkant, K., … White, K. (2010). Ada and Grace : Toward Realistic and Engaging Virtual Museum Guides. In

Proceedings of the 10th Annual International Conference on Intelligent Virtual Agents - IVA 2010 (pp. 286–300).

Tashakkori, R. M., Parry, R. M., Benoit, A., Cooper, R. A., Jenkins, J. L., & Westveer, N. T. (2014). Research Experience for Teachers: Data Analysis & Mining, Visualization, and Image Processing. In *Proceedings of the 45th ACM Technical Symposium on Computer Science Education - SIGCSE '14* (pp. 193–198). New York, New York, USA: ACM Press. http://doi.org/10.1145/2538862.2538967

Tegos, S., Demetriadis, S., & Tsiatsos, T. (2016). An Investigation of Conversational Agent Interventions Supporting Historical Reasoning in Primary Education. In *Proceedings of the 13th International Conference on Intelligent Tutoring Systems - Volume 9684* (pp. 260–266). Springer-Verlag New York, Inc. http://doi.org/10.1007/978-3-319-39583-8_27

Tewari, A., & Canny, J. (2014). What did Spot Hide? A Question-Answering Game for Preschool Children. In *Proceedings of the 32nd annual ACM Conference on Human Factors in Computing Systems - CHI '14* (pp. 1807–1816). New York, New York, USA: ACM Press. http://doi.org/10.1145/2556288.2557205

Thiry, H., & Hug, S. (2014). "We Should All Help Each Other": Latina Undergraduates' Practices And Identities In The Figured World Of Computing. In *International Conference of the Learning Sciences* (pp. 214–221).

Vail, A. K., & Boyer, K. E. (2014). Identifying Effective Moves in Tutoring: On the Refinement of Dialogue Act Annotation Schemes. In *Proceedings of the 12th International Conference on Intelligent Tutoring Systems (ITS '14)* (pp. 199–209).

http://doi.org/10.1007/978-3-319-07221-0_24

VanLehn, K. (2006). The Behavior of Tutoring Systems. *International Journal of Artifical Intelligence in Education*, *16*, 227–265. Retrieved from http://iospress.metapress.com/index/al6r85mm7c6qf7dr.pdf

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems and other tutoring systems. *Educational Psychologist*, *46*(4), 197–221.

Vanlehn, K., Burleson, W., Girard, S., Chavez-Echeagaray, M. E., Gonzalez-Sanchez, J., Hidalgo-Pontet, Y., & Zhang, L. (2014). The Affective Meta-Tutoring Project: Lessons Learned. In S. Trausan-Matu, K. E. Boyer, M. Crosby, & K. Panourgia (Eds.), *Intelligent Tutoring Systems* (Vol. 8474, pp. 84–93). Cham: Springer International Publishing. http://doi.org/10.1007/978-3-319-07221-0

Vygotsky, L. (1978). Mind and society: The development of higher mental processes. Cambridge, MA: Harvard University Press.

Wang, W. Y., Finkelstein, S., Ogan, A., Black, A. W., & Cassell, J. (2012). "Love ya, jerkface": using Sparse Log-Linear Models to Build Positive (and Impolite) Relationships with Teens. In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue - Sigdial 2012* (pp. 20–29).

Warren, S. J., Dondlinger, M. J., & Barab, S. A. (2008). A MUVE Towards PBL Writing. *Journal of Research on Technology in Education*, *41*(1), 113–140. http://doi.org/10.1080/15391523.2008.10782525

Webb, D. C., Repenning, A., & Koh, K. H. (2012). Toward an Emergent Theory of Broadening Participation in Computer Science Education. In *Proceedings of the 43rd*

ACM Technical Symposium on Computer Science Education - SIGCSE '12 (pp. 173–178). New York, New York, USA: ACM Press. http://doi.org/10.1145/2157136.2157191

Webb, H. C. (2011). Injecting Computational Thinking into Career Explorations for Middle School Girls. In *2011 IEEE Symposium on Visual Languages and Human Centric Computing VLHCC* (pp. 237–238). http://doi.org/10.1109/VLHCC.2011.6070410

Webb, H. C., & Rosson, M. B. (2011). Exploring Careers While Learning Alice 3D: A Summer Camp for Middle School Girls. In *Proceedings of the 42nd ACM Technical Symposium on Computer Science Education (SIGCSE '11)* (pp. 377–382). ACM. Retrieved from http://portal.acm.org/citation.cfm?id=1953275

Werner, L., Denner, J., & Campe, S. (2012). The Fairy Performance Assessment : Measuring Computational Thinking in Middle School. In *Proceedings of the 43rd ACM Technical Symposium on Computer Science Education - SIGCSE '12* (pp. 7–12).

Wiebe, E. N., Williams, L., Yang, K., & Miller, C. (2003). *Computer Science Attitude Survey*. Raleigh, NC: Dept. of Computer Science, North Carolina State University Technical Report TR-2003-1.

Wing, J. M. (2006). Computational Thinking. *Communications of the ACM*, *49*(3), 33. http://doi.org/10.1145/1118178.1118215

Woolf, B. P., Arroyo, I., Muldner, K., Burleson, W., Cooper, D. G., Dolan, R. P., & Christopherson, R. M. (2010). The Effect of Motivational Learning Companions on Low Achieving Students and Students with Disabilities. In *Proceedings of the 10th International Conference on Intelligent Tutoring Systems - ITS '10* (pp. 327–337).

Berlin, Heidelberg. http://doi.org/10.1007/978-3-642-13388-6

Wouters, P., van Nimwegen, C., van Oostendorp, H., & van der Spek, E. D. (2013). A Meta-Analysis of the Cognitive and Motivational Effects of Serious Games. *Journal of Educational Psychology Psychology*, *105*(2), 249–265.

Yadav, A., Zhou, N., Mayfield, C., Hambrusch, S., & Korb, J. T. (2011). Introducing Computational Thinking in Education Courses. In *Proceedings of the 42nd ACM Technical Symposium on Computer Science Education - SIGCSE '11* (p. 465). New York, New York, USA: ACM Press. http://doi.org/10.1145/1953163.1953297

Yarzebinski, E., Ogan, A., Rodrigo, M. M. T., & Matsuda, N. (2015). Understanding Students' Use of Code-Switching in a Learning by Teaching Technology. In *Proceedings of the 17th International conference on Artificial Intelligence in Education - AIED '15* (pp. 504–513). http://doi.org/10.1007/978-3-319-19773-9

Yee, N., & Bailenson, J. (2007). The Proteus Effect: The Effect of Transformed Self-Representation on Behavior. *Human Communication Research*, *33*(3), 271–290. http://doi.org/10.1111/j.1468-2958.2007.00299.x

Yee, N., Ducheneaut, N., Shiao, H.-T., & Nelson, L. (2012). Through the Azerothian Looking Glass: Mapping In-Game Preferences to Real World Demographics. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12* (pp. 2811–2814). New York, New York, USA: ACM Press. http://doi.org/10.1145/2207676.2208683

Yoo, D., Schanzer, E., Krishnamurthi, S., & Fisler, K. (2011). WeScheme: The Browser Is Your Programming Environment Danny. In *Proceedings of the 16th Conference on*

Innovation and Technology in Computer Science Education - ITiCSE '11 (pp. 163–167).

New York, New York, USA: ACM Press. http://doi.org/10.1145/1999747.1999795