Storytelling with robots:
Effects of robot language level on children's language learning

by

Jacqueline Kory

B.S. Cognitive Science
Vassar College, 2011

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning
in partial fulfillment of the requirements for the degree of
Master of Science in Media Arts and Sciences
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2014

(c) 2014 Massachusetts Institute of Technology. All rights reserved.

Author ....…..... **Jacqueline Kory**
Program in Media Arts and Sciences,
School of Architecture and Planning
August 8, 2014

Certified by …... **Cynthia Breazeal**
Associate Professor of Media Arts and Sciences
Thesis Supervisor

Accepted by …... **Patricia Maes**
Professor of Media Arts and Sciences
Interim Academic Head, Program in Media Arts and Sciences

Storytelling with robots:
Effects of robot language level on children's language learning

by
Jacqueline Kory

B.S. Cognitive Science
Vassar College

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning
on August 8, 2014 in partial fulfillment of the
requirements for the degree of
Master of Science in Media Arts and Sciences

Abstract

Children's oral language skills in preschool can predict their academic success later in life. Increasing children's skills early on could improve their success in middle and high school. To this end, I examined the potential of a sociable robotic learning/teaching companion in supplementing children's early language education. The robot was designed as a social character, engaging children as a peer, not as a teacher, within a relational, dialogic context. The robot targeted the social, interactive nature of language learning through a storytelling game, mediated by a tablet, that the robot and child played together. During the game, the robot introduced new vocabulary words and modeled good story narration skills. In a microgenetic study, 17 children played the storytelling game with the robot eight times each over a two month period. With half the children, the robot adapted its level of language to the child's level – so that, as children improved their storytelling skills, so did the robot. The other half played with a robot that did not adapt. I evaluated whether this adaptation influenced (i) whether children learned new words from the robot, (ii) the complexity and style of stories children told, and (iii) the similarity of children's stories to the robot's stories. I expected that children would learn more from a robot that adapted, and that they would copy its stories and narration style more than they would with a robot that did not adapt. Children's language use was tracked across sessions. I found that children in the adaptive condition maintained or increased the amount and diversity of the language they used during interactions with the robot. While children in all conditions learned new vocabulary words, created new stories during the game, and enjoyed playing with the robot, children who played with the adaptive robot improved more than children who played with the non-adaptive robot. Understanding how the robot influences children's language, and how a robot could support language development will inform the design of future learning/teaching companions that engage children as peers in educational play.

Thesis Supervisor: Cynthia Breazeal
Title: Associate Professor of Media Arts and Sciences

Storytelling with robots:
Effects of robot language level on children's language learning

Jacqueline Kory

The following people served as readers for this thesis:

Thesis Advisor:

____

Dr. Cynthia Breazeal
Professor of Media Arts and Sciences
Personal Robots Group
MIT Media Lab

Thesis Reader:

____

Dr. M. Ehsan Hoque
Assistant Professor of Computer Science
University of Rochester

Thesis Reader:

____

Dr. Edith K. Ackermann
Professor of Developmental Psychology
University of Aix-Marseille 1, France
Visiting Scientist at the MIT Media Lab

"We are all in the gutter, but some of us are looking at the stars."
– Oscar Wilde

Acknowledgments

Many people contributed to the creation of this thesis. I'd like to briefly acknowledge the wisdom, help, and encouragement of the following people:

My readers, Cynthia Breazeal, for her ideas, support, and magnificent vision of a world filled with social robots; Ehsan Hoque, for his many thoughtful comments on my ideas and for helping me think about technology and adaptation; and Edith Ackermann, for her great insights into interaction and children's behavior, and for understanding the importance of context.

My lab, the Personal Robots Group: the most awesome and supportive lab group one will ever find. I specifically want to acknowledge my officemate Sooyeon, for our conversations over tea and her knowledge of our robots and codebase; and Siggi, for teaching me a lot about debugging the DragonBots.

My mentors, past and present, some of the best an aspiring scientist could hope to work with: Ken Livingston, for seeing my potential when I was a mere college freshman, and for offering continuing support and encouragement as I delved into cognitive science. John Long, for the same, and for sharing his wisdom on perception-action loops. Gwen Broude, for the many books, and for the classes where I realized the importance of context and interactivity. Jan Andrews, for the cog sci language classes where I realized how important language really is. Jenny Walters, for introducing me to computer science. Roz Picard, for her thoughtful commentary and knowledge of emotion, and for reminding me to speak up. Sidney D'Mello, for introducing me to affective computing, showing me how to run a good study, and for reminding me to pay attention to detail.

Finally, my family, my sisters, and my husband Randy, for always encouraging me to explore.

"The day I know everything, I might as well stop!"
– *Doctor Who*

Table of Contents

Abstract....2
Acknowledgments....5
1 Introduction....8
2 Learning Language with Robots....9
  2.1 Social robotic learning companions....9
  2.2 Related work: Robots in schools....10
  2.3 Areas for improvement....11
  2.4 Robotic language learning companions: Research questions....12
  2.5 A storytelling robot for children's language development....12
3 Storyspace: A simple storytelling game....12
  3.1 Choice of game....12
  3.2 Game design....13
  3.3 Pilot study....16
4 Microgenetic Study: Storybot....21
  4.1 Overview....21
  4.2 Methods....22
  4.3 Results....33
  4.4 Discussion....53
5 Conclusion....58
  5.1 Contributions....58
  5.2 Future work....59
6 References....62

1 Introduction

Research from the past two decades has revealed that children's early oral language knowledge and vocabulary skills are a primary predictor of later learning and academic success. For example, children whose preschool instructors regularly used unfamiliar words and engaged them in cognitively challenging tasks generally entered kindergarten with higher language abilities (Snow, Porche, Tabors, & Harris, 2007). Differences in kindergarteners' vocabulary skills predicted differences in reading ability in middle and high school.

A more surprising example comes from Hart and Risley's (1995) classic work. Hart and Risley (1995) examined the number of words addressed to children, taking this as a measure of children's cumulative language experience. They found that preschool-age children raised in families with lower socio-economic status (SES) had significantly smaller vocabularies than children whose parents had higher SES. These differences magnified over time. There was a 30 million word gap by age 3 between the estimated number of words that children from low SES families were exposed to versus children from high SES families. Furthermore, vocabulary use at age 3 was predictive of language skill at ages 9-10, and strongly associated with scores on later vocabulary and reading comprehension tests. Other researchers have found similar deficits in the language abilities of children raised in environments with an impoverished exposure to novel English words or rich vocabulary-building curricula (e.g., Fish & Pinkerman, 2003; Páez, Tabors, & López, 2007).

The studies mentioned above highlight the importance of early language exposure and oral language development, particularly during the preschool years, for children's later success. These results are generally interpreted to mean that we should expose children to more words – to as many words as possible! – since if children do not have sufficient exposure to a sufficient quantity of words, they are less likely to be academically successful. The emphasis becomes on teaching vocabulary.

However, two issues should be taken with this view. First, the importance of language and literacy is not just about academic success. As Neil Gaiman (2013) recently wrote, language is key to understanding other people. Through understanding language, using language with imagination, and being able to read stories, we build empathy. Psychologist Paul Bloom (2000, pp. 259) writes similarly: "language allows us to express our thoughts and understand those of others—to become full-fledged members of the human community." Language helps us understand our differences, our similarities, and how we can change through time. This understanding may be as important as, if not even *more* important than, academic achievement to the future of humanity.

Second, "teaching vocabulary" may not be a sufficient goal. Prior work suggests that children's language development is not solely a result of *exposure* to words. Children learn language through a *dialogic context* – through social interactions in which meaning is communicated, where words happen to be the means of communicating (Duranti & Goodwin, 1992; Vygotsky, 1978). This kind of context is interactive and social by definition. Children engage as active participants, both as speakers and listeners. And they engage for a reason. Something needs to be communicated; this is why the dialogic context arises in the first place. As such, if our goal is to encourage further language development, we should not simply expose children to more words. We should re-engage children in a dialogic context, and we should provide richer dialogic contexts in which children can learn to understand and communicate many more meanings. Indeed, if we look at the studies above again, they are all studying language in some kind of dialogic context: instructors speaking with children, words addressed to children, environments in which dialogues are occurring and available.

This thesis explores how technology can create and support dialogic contexts for preschool children's oral language development. While the focus is on oral language, this kind of technology may also support other precursors to literacy, such as joint pointing, the performative power of words, and pretense reading (Ackermann, 2002). A social robotic learning/teaching companion is introduced that plays a storytelling game with children. The robot is the interactive dialogue partner; the game becomes the reason for meaning to be shared. During the interaction, the robot can introduce new words and model more complex uses of language, but more importantly, the robot provides the social context for language use.

2 Learning Language with Robots

2.1 Social robotic learning companions

Acquiring language is an inherently social activity. Language, embedded in culture and context, is for communicating meaning with others. Accordingly, prior research suggests that social interaction matters a great deal to children's language development. Social cues that facilitate joint attention, rapport, and a sense of joint purpose – such as eye-gaze, motor, and affective mimicry and synchrony (Chartrand & van Baaren, 2009; Valdesolo & DeSteno, 2011; Wiltermuth & Heath, 2009) – are critical both for language learning (Bloom, 2000; Meltzoff, Kuhl, Movellan, & Sejnowski, 2009) and for children's readiness or willingness to engage with instructors (Corriveau, Harris, Meins et al., 2009; Harris, 2007). In addition, the lack of a social other or interaction partner seems to impair learning: Infants may readily learn to differentiate new phonemes presented by a live person, but may have difficulty picking up this information from a video of a person (Kuhl, 2007). Young children may learn new words from educational television, but might not learn grammar or complex sentence structures (Naigles & Mayeux, 2001).

The exact mechanisms by which children learn words are unknown, as Paul Bloom (2000) rightly notes. Nevertheless, available evidence points to a key role for social interaction, social presence, and social cues. Given this, any technology designed to help children learn and practice oral language should be social itself. To this end, I hypothesize that *robotic learning/teaching companions* that act as peers or tutors could be a highly beneficial technology to supplement children's early language education, for three key reasons. First, utilizing technology for language learning has several benefits: (a) accessibility—being able to deploy at-scale as technology becomes cheaper and more accessible, (b) ease-of-use—rapid customization and the addition of new content, and (c) versatility—it can be used alone, with peers, or with a caregiver. Second, sociable robots share physical spaces with humans and can leverage the ways people communicate with one another to create more intuitive interfaces for interaction. For example, these robots may use behaviors such as speech, movement, expressions of affect, and nonverbal behaviors such as mimicry, gaze following, and synchrony. These are all cues that humans easily interpret. As such, children willingly treat sociable robots as companions from whom they can learn (e.g., see Freed, 2012; Kahn et al., 2012; Movellan, Eckhardt, Virnes, & Rodriguez, 2009; Sage & Baldwin, 2010; Tanaka & Matsuzoe, 2012). Third, sociable robots could combine critical aspects of social interaction with student-paced educational software and individual attention.

It is important to emphasize here that educational technologies, such as robots, are not designed to be replacements for parents or teachers—quite the opposite. The goal is to *supplement* what caregivers are already doing and *scaffold* or *model* beneficial behaviors that caregivers may not know to use, or may not be able to use. A robot can be a peer, with whom an enjoyable activity can be shared, from whom things can be learned. For example, a robot could play an educational game with the child, during which it could introduce new words and information, model more advanced speech patterns, and ask questions that spark conversation. Freed (2012) developed a simple food-sharing game for learning vocabulary that parents and children played with a robot. Without prompting, parents aligned their guidance and reinforcement of children's behavior during play with language learning goals. The robot's presence encouraged communication and discussion between children and their caregivers. These technologies may be especially useful for parents who may not be proficient English speakers themselves, or who may not be able to spend a lot of time with their children; for teachers who want to extend their capabilities to work with individual or small groups of children in the classroom; or even for children who play on their own, perhaps as an alternative to less beneficial activities such as just watching TV. In general, the goal of introducing a social robot into a child's learning environment is to augment the existing relationship between children and their parents, siblings, and teachers.

2.2 Related work: Robots in schools

Some robotic learning companions for children have already been developed. They are taken to schools for an afternoon (e.g., Tanaka & Matsuzoe, 2012), or for a series of play sessions over several weeks (e.g., Chang, Lee, Chao, Wang, & Chen, 2010; Movellan et al., 2009; Movellan, Malmir, & Forester, 2014). Activities varied, though the focus is generally on teaching children new vocabulary. For example, the RUBI-4 and RUBI-5 robots from Javier Movellan's lab played simple vocabulary games with preschool children on the screen embedded in the robot's stomach (Movellan et al., 2009; Movellan et al., 2014). In one game, four objects were presented on the screen, and the robot asked children to touch one of the objects (a flashcard-type game). A similar game involved taking and giving back physical objects. Tanaka and Matsuzoe's (2012) robot played a verb-learning game, in which the experimenter asked either the preschool child or the robot to act out novel verbs. When either performed the desired action incorrectly, the experimenter would demonstrate the verb's action for the child or directly teach the robot the action. After a couple rounds, the experimenter asked the child to teach the robot the action. They found that teaching the robot helped children remember the verbs, as well as inspiring further teaching-verbs play.

Chang et al. (2010) wanted to support second-language learning practice, beyond just vocabulary. Their robot, developed as a “teaching assistant” for fifth-grade classrooms (in a Wizard-of-Oz scenario), performed five different language activities. It read aloud stories, led students in reciting vocabulary and sentences, performed physical actions when instructed by students, asked and answered simple questions, and acted as a “cheerleader” to encourage students to participate in other classroom games. Justine Cassell and colleagues developed a virtual agent that also supported broader language development through storytelling games (Ryokai, Vaucelle, & Cassell, 2003). The agent appeared on a screen as a virtual child, projected behind a toy castle. It listened to stories children told about a figurine in the toy castle and told stories of its own, modeling more advanced narrative language for the child. This was effective because the agent, although virtual, was still a socially-situated peer that could engage with children in a natural play environment. It gave both verbal and nonverbal feedback to encourage children to continue talking—smiles, nods, 'uh-huh' sounds, and prompting questions. It enabled children to produce and construct stories as well as consume them.

2.3 Areas for improvement

In light of this prior work, this thesis addresses several areas in which robotic learning companions have been under-explored and/or can be improved. First, these robots should more fully utilize the social, personal aspects that give them such potential. As was discussed earlier, social cues are crucial for engaging children in the dialogic context. Some researchers have made progress in this area. For example, Tanaka and Matsuzoe's (2012) Nao greeted preschool children by name and shook their hands in an attempt to build rapport. Movellan et al.'s (2009) RUBI-4 supplemented the screen-based activities with the robot's physical behavior. It expressed some emotions with its face and voice, used its arms for gesturing and exchanging objects with children, and talked about the activities on its screen. The virtual agents in Cassell's (2004) story-listening systems used many social cues, giving both verbal and nonverbal feedback. However, there is much untapped potential.

In a recent study, I investigated how a robot's social behavior changed children's perceptions of the robot as an informant from whom they can learn (Breazeal, Harris, DeSteno, Kory, Dickens, & Jeong, in review). Preschool children talked with two robots that provided information about unfamiliar animals. The robots differed in a subtle way: one robot attended to the child in a socially contingent fashion, signaled via head and gaze orientation (e.g., looking at the child when she or he was speaking), and the timing of backchanneling behaviors (such as 'uh-huh' sounds). The other robot was just as expressive, but its behavior was not contingent on the child's (e.g., looking away when the child was speaking). Children spent more time attending to the contingent robot. They preferred to seek and endorse information from the contingent robot. This suggests that the robot's social responsiveness may have a significant effect on learning outcomes. In this thesis, the robot uses these contingent social cues and is highly socially responsive.

A second area for improvement is in adaptation or personalization to individual children. Children learn at different paces. A learning companion may be more successful at helping children learn if tailored to the needs of individuals. Some interactive systems for older children or adults have incorporated adaptation to individuals, generally finding greater engagement and learning outcomes than the same systems without adaptation (e.g., Castellano, Leite, Pereira, Martinho, Paiva, & McOwan, 2013; D'Mello & Graesser, 2012; Kasap & Magnenat-Thalmann, 2012). Change in the robot's speech and behavior over time may be crucial for maintaining engagement over multiple encounters and in building a long-term relationship (Bickmore, Schulman, & Yin, 2010; Kidd & Breazeal, 2008; Lee, Forlizzi, Kiesler, Rybski, Antanitis, & Savetsila, 2012). Little work has been done in creating adaptive learning companions for preschool children. In this thesis, I examine how increasing the robot's language abilities over time affects children's learning.

Finally, we do not know what makes a robot a better learning companion. Several major themes emerge for further study. First, the robot's appearance. This includes questions such as how the robot should look and what it should sound like, as well as questions of how the robot should be presented to the child. It may be that framing the robot as a “social other” versus as a machine, versus free play with no framing at all will have a significant impact on how a robot appears to a child (Coeckelbergh, 2011a, 2011b; Kory & Kleinberger, 2014). Second, the robot's behavior. Which social or affective cues should the robot use? How will the robot express these? As noted earlier, social responsiveness impacts whether children will perceive a robot as someone from whom they can learn. Third, what content should the robot present? This includes questions about the curriculum the robot should follow, and what expertise the robot should have. Expertise is one factor children take into account when determining who to learn from (Harris, 2007). All these factors will impact how effective any given robot will be. This thesis contributes insights into how the robot's behavior and content interact, informing the design of future robotic learning/teaching companions.

2.4 Robotic language learning companions: Research questions

In this thesis, I asked to what extent a robot could facilitate preschool children’s long-term oral language development. Specifically, I investigated how the language used by the child’s play companion during a storytelling game influenced the language the child uses over time (such as the complexity of stories and similarity to the play companion’s language style), and whether the child learns new words. How does increasing the companion's ability over time affect the child's language and learning? Will *adapting* the companion’s ability to the child’s (“growing” with the child) be more beneficial than *not* adapting (falling behind the child, as the child grows)? As a secondary aim, I asked how children construed the robot during the long-term interaction, as well as what the kinds of relationships children developed with the robot over time.

2.5 A storytelling robot for children's language development

To address these questions, I created a storytelling game that was the basis of the interaction between the robot and child. I tested this game with children at the Boston Museum of Science, using the results to inform both further development of the game and the robot's behavior and stories while playing the game with children (Section 3). Next, I performed a microgenetic study in which children played the storytelling game with the robot eight times over two months. This study allowed me to observe changes in the children's language abilities over time in relation to the robot's abilities (Section 4).The design of the study is also described in Kory & Breazeal (2014). Conclusions and thoughts on future work follow (Section 5).

3 *Storyspace*: A simple storytelling game

3.1 Choice of game

In choosing a storytelling game as the context for studying whether a robot can aid in children's language development, I considered multiple factors. First, play and narrative are highly intertwined and highly important in children's learning and development (Golinkoff, Hirsh-Pasek, & Singer, 2006; Nicolopoulou, 1993). Young children enjoy playing with robots, as has been shown in many studies (e.g., Breazeal et al., in review; Freed, 2012; Movellan et al, 2009). Could children engage in symbolic or pretend play with a robot, creating narratives together, as children do when they play with each other? However, the space of the robot's speech and actions had to be limited due to technological constraints. As such, the robot would have to play a more structured game, rather than opening the field to the entire realm of pretend play. The goal of such a game would be to encourage the child to speak more, practice known words, learn new words, and build up more advanced language structure. Within the game context, I wanted to explore how children learn through doing, rather than from being explicitly taught.

Storytelling was an ideal choice for the game: a socially-situated activity in which children readily engage, rooted in free play, and allowing creative conversation as well as space for learning about language. As Justine Cassell (2004) describes, storytelling can support *emergent literacy*, which she defines as "the ability to make meaning for others across space and time." Stories can help children learn metalinguistic knowledge about language patterns, structure, and function; vocabulary; "decontextualized" language that can be understood outside its original context; as well as supporting cognitive, communicative, and linguistic development more broadly (Curenton, Craig, & Flanigan, 2008; Engel, 1995). Storytelling activities with peers function as a platform for learning, practicing, and constructing one's own language in a collaborative, social context. As described earlier, Justine Cassell (2004) and colleagues created several *story-listening systems* that encouraged children to tell stories with peers as a means for developing literacy, supporting children as producers as well as consumers of stories.

To these ends, I developed and tested *Storyspace*, a simple game that affords the imaginative creation of stories.

3.2 Game design

3.2.1 Design constraints and game play

Several design constraints were evident. First, the robot needed to be *able* to play the game with the child. To this end, the game was situated on a tablet computer embedded in a small play table at which the child and robot could both sit. Game events from the tablet could be streamed to the robot, and the robot could send commands to the game over the network. This allowed the robot to be a full participant in the game with control over the game's elements.

Second, the focus of the game had to be on telling stories, not on some other aspect of play. The game also had to be easy to learn and to play. As such, the game was kept very simple. The tablet displayed a limited set of characters and objects against a background image (the story setting – see next section). The characters and objects could be dragged across the screen, allowing the storyteller to walk them through a story – analogous to "virtual stick puppets." However, this draggable behavior was the *only* provided interaction on the screen. No other finger gestures (e.g., double taps, scrolling, rotating) were used. No animations were included. These choices were deliberate, because the focus was to be on the story, and on telling stories with another person, not on viewing animations on the screen. This was not a narrated movie. As will be discussed later (Section 3.3.2), when presented with animations, children often simply seek to replay these over and over, rather than engage in actual storytelling.

The game play progressed as follows: The robot would select a story scene and tell a story, moving the characters to act out the story. Then, guided by the robot, the child would be invited to tell a story about the same characters in the same scene (similar in structure to the *Sam the CastleMate* game described in Ryokai et al., 2003). During the child's turns, the robot would use backchanneling behaviors (e.g., smiles, “mmhm” sounds, gaze) to encourage children to continue telling their story. Then, the robot would take another turn, selecting a new story scene and telling a new story. The child would be invited to take another turn as well.

Finally, the game needed to engage children over multiple play sessions. The best way to combat boredom is through variability. Multiple story settings were created for the game (8 scenes total), with different character present in each. In addition, when playing the game with the robot, the robot would have more than story available for each story scene, so that if a scene was encountered twice, the story the robot told would be different each time.

While a tablet-based game does not contain the same embodied flavor as physically playing with a doll in a castle (as in *Sam the CastleMate* described in Ryokai et al., 2003), creating stories on a tablet may support a wider range of make-believe scenarios and could cater to children's diverse interests. Future improvements could be made: In the next version of the game, children could select among various characters or scenes in telling their story, tailoring the game to their own ideas. The game could include a connection to past children who played it, similar to the *StoryMat* (Cassell & Ryokai, 2001), by playing back audio from other children's stories or showing animations of what children had their characters do. Children could playback their own stories in order to revise or expand them.

This is different from prior work with storytelling agents foremost because the game will be played with an embodied robot, rather than a virtual peer. Past work has shown that people rate interactions with physically embodied agents more positively than with virtual agents, and will generally be more compliant toward trust-related tasks (Bainbridge, Hart, Kim, & Scassellati, 2011; Kidd & Breazeal, 2008). I expect that embodying the child's play companion will have similar benefits. The child may be more likely to adapt their language to match the robot's or may be more willing to play with the robot more. In addition, few researchers have incorporated robots into storytelling activities. Those that did used the robot as a *character* in the story, not as a *participant* in the activity (see Chen & Wang, 2011, for a survey of these). Here, the robot will be a peer, engaging in the activity as much as the child is. It is also different from the activities that other robots have played with children, because the game is rooted in free play, rather than a structured adult-led game or activity. Details of the robot-child interaction will follow in Section 4.2.

3.2.2 Game scenes

Eight scenes were created for the story game. Each scene included three characters and up to three objects against a background image. The scenarios and characters were chosen to be appealing to children. Not all the scenes were complete before the pilot study began, which allowed feedback from children about desired scenes or characters to be incorporated into new scenes. The scenes are listed in Table 1 and depicted in Figure 1. The game was built in Unity 4 / Mono.

Figure 1: Images of the eight scenes created for the storytelling game. A variety of scenarios were selection, with several different characters present in each scene.

Story	Characters	Objects
Castle with moat	Prince, Princess, Crocodile	None
Dinosaur pond	Stegosaurus, T-Rex, Brontosaurus	None
Living room in a house	Girl, Cat, Dog	Ball, Box, Box
Iceberg with snowman	Penguin, Penguin, Seal	Snowman's Hat
Mars-like planet	Green alien, Green alien, Blue alien	Space shuttle
Pine forest	Rabbit, Turtle, Squirrel	None
Playground	Girl, Girl, Boy	Ball, Ball
Meadow with trees	Dragon, Butterfly, Bird	None

Table 1: A list of the scenes created for the storytelling game, the characters present in each scene, and any draggable objects also present in the scene.

3.3 Pilot study

I tested the storytelling game with children to find out (1) whether the game was to easy to learn and to play, (2) whether children enjoyed the game, (3) whether children thought the game would be fun to play more than once, (4) whether play focused on storytelling and language use, not just on what was happening on the screen, (5) how children interacted with the game more generally, (6) children's gaze patterns during storytelling, and (7) what kinds of stories they told.

3.3.1 Methods

3.3.1.a Participants

Children ages 2-6 played the game at the Boston Museum of Science Discovery Center Living Laboratory. Some children played at the same time as a sibling, and some played with their parents or caregivers. A total of 30-40 children played the game. However, audio and video data was only recorded for 10 children because only some parents allowed recording.

3.3.1.b Interaction

Visitors to the Discovery Center were invited to participate in this pilot study. After consenting to participate (with or without data recording), children were invited to join the human experimenter at a short table in the corner of the Discovery Center, where they played the game. The experimenter introduced the story game in a way similar to how the robot would introduce it, as follows: “We're going to play a story game! We can make up stories about these animals. Here, I'll go first to show you how it works. Then it'll be your turn.” The experimenter then told a story, using her finger to move the characters around on the screen. An example story is as follows:

> *Once upon a time, there was a little penguin named George. He enjoyed sliding down icebergs. One day, he went to an iceberg with Gabby, the other penguin. They climbed up to the very top of the iceberg. George wanted to slide down, but Gabby was scared. "Why are you scared, Gabby? It's so much fun to slide!" "It's so high up!" "Okay, I'll slide with you!" So George and Gabby slid all the way down together. They went so fast! Gabby thought it was fun. "Let's do that again!" The end.*

Following this, the child was prompted to take a turn: “Now it's your turn to tell a story!” Further prompts (e.g., “And then what happens?”, “What does X do next?”) were used as needed to encourage storytelling.

Children were invited to tell stories about one or more scenes in the game, based on their interest. When they were done, the experimenter asked them several semi-structured interview questions regarding their opinion of the game, whether it was fun, what else they might want to tell stories about, and their age. Finally, all children were given a sticker as thanks for participating.

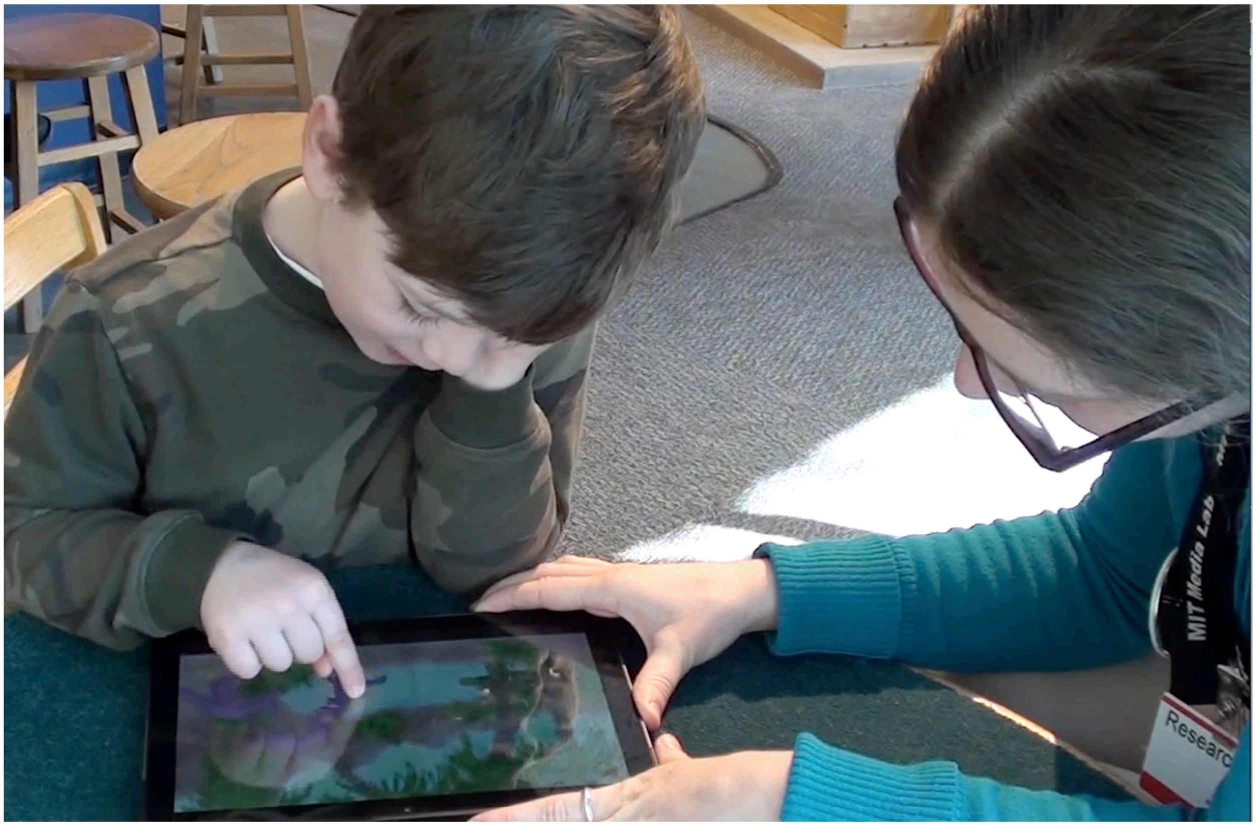

Figure 2: A child tells a story using the Storyspace app on the tablet with a researcher.

3.3.1.c Data analysis

Children's stories were transcribed and analyzed to determine (a) word count; (b) story complexity via qualitative analysis and Coh-Metrix, a text-analysis software (McNamara, Louwerse, Cai, & Graesser, 2013); and (c) general topics and themes incorporated in stories. Videos of children were analyzed to determine gaze patterns during storytelling. This information was used to inform the robot's language and behavior (see Section 4.2).

3.3.2 Results & Discussion

As noted above, the pilot study asked seven main questions regarding ease of game play, children's enjoyment of the game, potential to play more than once, the focus of game play, children's general interactions with the game, children's gaze patterns, and the stories children told. The results were informative on all counts.

3.3.2.a Ease of play

Children readily learned how to play the game. Most were familiar with tablets or iPads and had used such a device at least once before. Younger children (2-3 years) had more trouble dragging objects on the screen than older children (4-6 years), but with guidance, all children successfully moved the objects. When the experimenter modeled how to tell a story with the game, taking the first turn, children easily figured out how to tell a story themselves. Younger children told fewer stories, and shorter stories, than older children, likely because their expressive language abilities are lower and not as developed.

3.3.2.b Enjoyment

Nearly every child enjoyed playing the game, voluntarily staying to tell stories about multiple scenes. One notable child stayed for upwards of twenty minutes, only leaving because her caregiver insisted she give other children a turn. This child proceeded to slip back to the Living Laboratory corner several more times to play again, only to be carted off the minute her caregiver discovered where she had gone.

One parent noted that the game seemed to be most enjoyable when played with another person. By oneself, the game is boring: You drag objects, they move, that's all. With another person, it's suddenly exciting! You can take turns creating stories!

3.3.2.c Playing multiple times

When asked, children often said they would like to play the game again. Some children and caregivers asked if the game was available on the app store, indicating interest in playing the game on their own.

3.3.2.d Focus of game play

When the experimenter modeled how to tell a story with the game, children were more likely to tell stories themselves. If, however, a child was shown the game, but was not shown how to tell a story with it (i.e., move the characters while narrating), the child tended to poke the screen and drag objects, with no apparent goal or purpose.

This behavior likely results because most of the games children play on tablets or iPads function with a "poke and see what happens" game style. The games are not about creation outside of the screen; they are about touching to make things happen. Storyspace differs significantly from that – the screen is a setting, a platform; the primary game is not in the screen, but in the stories told aloud with words. As such, for the game to work, I found that storytelling must be modeled by the experimenter (or, in the future, by the robot), so that the child can learn that *this* game operates differently than other tablet or iPad games.

3.3.2.e General Interactions

Children were, for the most part, happy to take turns telling stories with the experimenter. However, some children tried to interrupt the experimenter's story or grab the tablet away. In addition, children frequently managed to accidentally exit the game through the tablet's menu bar located at the bottom of the screen. When the game is played with a robot instead of with an adult, these kinds of activities will need to be prevented, because the robot may not be able to restart the app or take the tablet back from the child. Two solutions arise. First, the tablet could be embedded in a small play table that hides the bottom menu bar, physically preventing children from reaching the “back” button that would exit the app. This also situates the tablet as a shared surface, rather than a device that can be grabbed and held by just one participant in the game. Second, during the robot's turns, all touch events on the tablet screen could be disabled, so that children literally cannot move the characters when the robot is telling a story, preventing touch-based interruptions. The robot should also have some dialogue available during its turns for dissuading the child from interrupting (e.g., “Hey, it's still my turn!”, “Please wait, I'm not done!”, or “Wait! You get the next turn!”).

Parents offered many suggestions regarding the game. One parent suggested the game include multiple scenes with the same characters, allowing those characters to go on adventures. Another parent gave the idea of including objects as well as characters, and increasing the detail of the images, to allow for more complex interactions. Another idea was to include a record-playback function, so that children could record, edit, and replay their stories (similar to the record/playback functions of StoryMat described by Cassell, 2004, which played back *other* children's stories as well). In general, parents liked the focus on exploration and interactivity in the service of telling stories.

As described earlier, no animations beyond simple dragging of objects were included in the story scenes. In order to verify that this was a valid design decision, one animation was added to the Iceberg scene. If the user tapped the sun in that scene, clouds would appear and snowflakes would fall. Once this snowing animation was discovered, children's attention was captured by starting and stopping the snowstorm, rather than on creating new stories. They might still tell a story, but a very simple one: “Then it snowed! Then it stopped snowing. Then it snowed!”. This is an argument *against* putting complex animations in the story – rather, it suggests that keeping the scenes as simple as possible while still affording interesting or detailed scenery and characters could lead to higher engagement and more interesting stories. Complex animations may be distractions. They lead children to search for hidden animation triggers in the scene, a treasure hunt of sorts that devolves into randomly poking objects in the scene, rather than focusing on making up a story of their own.

3.3.2.f Gaze

I performed an analysis of children's gaze patterns during storytelling and story-listening. Unfortunately, not many parents allowed video to be recorded, and not all the video had a good view of children's eyes. As such, these observations are qualitative in nature. In reviewing the videos, I found that children tended to look up at their partner during exciting moments in their stories, for emphasis, or possibly to check that their partner was paying attention. They generally spent the remainder of the time looking down at the tablet screen. When listening to their partner's stories, children tended to glance up at their partner approximately every 4-6 seconds. Their gaze stayed up for a brief time (1-3 seconds), and then they would look back down at the tablet screen. This behavior likely varies by child, and the limited number and quality of videos prevented a more definitive analysis from being performed. However, this was enough to inform the development of the robot's behavior (Section 4.2.3).

3.3.2.g Children's stories

Some children told no stories (often due to shyness), and some told several. The length of their stories varied. Of the stories that were recorded (8 children, 20 stories), the mean word count across all the stories was 101.8 (*SD*=54.9, *min*=46, *max*=230). Qualitatively, their stories were highly narrative. They tended to use simple sentence structures and easier words.

I also assessed children's stories using (1) Flesch-Kincaid Grade Level (FKGL), (2) Coh-

Metrix (McNamara et al., 2013) indices of difficulty (narrativity, syntactic simplicity, word concreteness, referential cohesion, deep cohesion), where higher values of the indices are generally easier to read and understand (they range from 0-100). The scores for narrativity (*M*=87.4, *SD*=18.0) and syntactic simplicity (*M*=89.5; *SD*=12.2) reflected the qualitative assessment. It should be noted, however, that Coh-Metrix may not give accurate or informative results when analyzing small samples of text (less than 200 words), so, since children's stories were generally short, these scores should be interpreted cautiously. The mean Flesch-Kincaid Grade level across stories was 1.8 (*SD*=1.5, *min*=0.1, *max*=3.7), which reflects the use of shorter sentences and simpler words with fewer syllables.

Children covered a range of themes in their stories. Some children incorporated elements of the experimenter's stories into their own – e.g., if the experimenter told a story about animals playing hide-and-seek together, the child might also tell a story about hide-and-seek games. Some examples follow, ordered by complexity and length:

> *(Iceberg scene)*
> *They go up and go down. Or they'll go over. And go down. And then he play hide and seek. Up and down down down down. Up on the snowman. And back down and wheeee. Then he hide. And he goes around and on. The end.*
>
> *(Meadow scene)*
> *Yeah, yeah. He goes hide. So he, the birdy will hide. The birdy is not. Now it's the butterfly's turn. Wheee. Now he's this he found her. It's this guy's turn. This guy can't find him. This guy couldn't find him. Now all of them are hiding. And now, this bird, this bird is counting to ten, until they can find him. He didn't find him!*
>
> *(Mars scene)*
> *One day there was an alien mom, an alien kid, and an alien father. They lived in a nice home. One day the alien father had to go on a long trip. He got in the rocket ship and flew away. Then the mom and the kid were so sad. He came back in the rocket ship. They had a great big party. The end.*
>
> *(Castle scene)*
> *Once there was a princess that lived in the castle up there. But she couldn't get down. But she could get down, actually, by the chimney. But there was a mean old alligator that wouldn't let the prince back in. So So the princess slid down and then she ran across and then ran back and he went up there, and then she went up there. The end.*
>
> *(Iceberg scene)*
> *One day, Daisy was building a snowman with her friend. They didn't have a carrot for their nose, so Daisy said, "I'm going to go find a carrot at the top of the mountain." So he went up to find one. His friend didn't want to be alone, so he followed, but they didn't. But Daisy didn't know his friend was following him. And when they got up there, and she slid back down with the carrot nose, they bonked into each other and they got a big surprise and then they put the carrot nose. The End.*

(Iceberg scene)
Once upon a time there was a little snowman. Then there was a little penguin. And he was the best. Then his mother came up and then this seal. And then the seal slid all the way down. And then the little baby did. And then the mom. And then they all and then the mom went up one iceberg and the little baby went up the other and the mom slid down because it was time to eat, and the baby didn't want to eat. He was not hungry. he stayed at the top and he slid all the way up to the highest one. Then he hid behind a rock. Then he jumped out and he slid down and he went up the other one and he went down. And then mom said, lunch is ready and he did not want to go and he just sat there. Then he said I'm hungry and he slid down the high one and he ate his lunch. The end.

(Castle scene)
Once upon a time, there was a prince that lived at the top of the castle. And a princess living at the top of the castle. Someone else living near the castle. A big dragon. And one day, he sneaked and he went up and scared the other people ROAR and he went down the path again. And then he went down the other path. And but one day they thought he was come back but he didn't. He stayed and hided at the lake. And the princess walked down the path and into the castle. And she said, it's safe to go down and play! And the prince said, okay! And he went down and walked and said I don't think it is, and they walked out and saw the dragon. And he ran into the castle! And he said to the princess, the dragon! and she said I'm not sure. Oh yes! And they went up and one day the dinosaur really got hidden. He hided just like that. And when they went down and tried to see the dragon, they couldn't. And they played happily together. And then they lived happily ever after in the castle together. The end!

The stories told by children and the information about the stories' complexity were used to develop the robot's stories (Section 4.2.7).

4 Microgenetic Study: Storybot

4.1 Overview

4.1.1 Research questions

Language development takes time. The goal of this study was to investigate whether playing the storytelling game with a robot could be beneficial for children's language development. The study was set up microgenetically (Siegler & Crowley, 1991); children played the storytelling game with the robot multiple times. This long-term interaction allowed me to ask how the language used by the robot in its stories influenced the children's language through time. How did the complexity of children's stories change through time? Did children learn new words from the robot, and did they use these words in their stories? The long-term interaction also allowed me to examine how the relationship between the child and the robot influenced how they played the game together. If children play with a new robot or game just once, are any results just due to the novelty? How effective is the robot or the game at engaging children more than once?

In addition, during the pilot study of the story game, I had found that some children did not want to tell stories due to shyness and the unfamiliarity of the experimenter. After multiple sessions with an initially unfamiliar robot, will shy children become more comfortable, and more willing to tell stories? The long-term interaction also allowed me to test whether *adapting* the robot's language ability to the child's – so that it "grows" with the child – influences children's language differently than a robot whose language ability does *not* adapt – and thus, "falls behind" as the child grows. Specifically, does playing with the *adaptive* companion lead to greater learning outcomes, as measured through analysis of the complexity children's stories, and vocabulary and language tests? Finally, what kind of relationships would children develop with the robot over time? How would children construe a robot that they encountered repeatedly?

4.1.2 Conditions

Two conditions were tested in the microgenetic study. The first condition was a control. The robot did *not adapt* to the child. It told simpler, easier stories to all children. This meant that as children learned and developed, the robot would appear to "fall behind," becoming a peer with a lesser language ability than the child. In the second condition, the robot's language ability adapted to the child's, so that the robot told simpler stories to children with lower language ability and more complex stories to children with higher language ability. The robot's language only differed during its stories in the storytelling game.

4.1.3 Hypotheses

I expected that children would feel most comfortable playing the storytelling game with a robot at their language level because it would act and speak most like themselves. A robot with too advanced an ability may be intimidating, while a robot with too low an ability may appear boring or stupid. However, I also expected that children would learn more from a robot with greater ability than themselves, and that they would copy its stories and narration style more than they would with a robot of lesser ability. Vygotsky's theory of the zone of proximal development suggests that children may learn more readily when slightly challenged [31]. A more advanced robot may present a slight challenge. As such, all the robot's stories were designed to be slightly more complex than the stories told by a child of comparable language ability to the robot (i.e., the robot's simpler stories are slightly more complex than stories told by a child of lower language ability and the robot's more complex stories are slightly more complex than stories told by a child of higher language ability – see Section 4.2.7), thus causing the robot to appear as a slightly older peer. That said, I also expected that playing with a robot of lesser ability could prompt teaching or mentoring behavior from children which could also be beneficial to language learning (Tanaka & Matsuzoe, 2012). This might occur when the non-adaptive robot continues to tell simpler stories to children of high language ability.

4.2 Methods

4.2.1 Procedure

Each child participated in 9 sessions, spread over 10 weeks. During the first session, children were given a language assessment, a subset of the Preschool Language Scale, 5th

Edition (PLS-5; Zimmerman, Steiner, & Pond, 2011), to assess aspects of their expressive and receptive language ability (Section 4.2.9). During the following eight sessions, children played the storytelling game with the robot for 10-15 minutes. Half the children (8 children) were randomly assigned to play with the *non-adaptive* robot; the other half (9 children) played with the *adaptive* robot, whose story level was adapted to theirs after the first 4 sessions. Each session has five phases: (i) vocabulary test, (ii) introductory chat with robot, (iii) storytelling game with robot, (iv) closing chat with robot, (v) vocabulary test. The set of target vocabulary words tested were those that the robot used during the session (Section 4.2.8).

4.2.2 Participants

Participants were 17 children ages 4-6 from two Boston-area preschools (9 from the first and 8 from the second). 10 were female; 7 were male. There were three 4-year-olds, thirteen 5-year-olds, and one 6-year-old. The 6-year-old girl did not complete the final session, and one 4-year-old girl completed only the first 4 sessions. Children in this age range were targeted because their expressive language abilities are developed enough to be able to tell stories. They are still in the process of developing their narrative abilities. Younger children, as was discovered during pilot testing of the storytelling game, may not tell stories at all and are less likely to understand and follow the rules of the game.

Thirteen children said they had played with a tablet or iPad before. Ten children said they had played with a robot before. This included experience with a different DragonBot during a previous study that had been conducted at one preschool, a remote-control motorcycle, and two different toy robots. All of the children except one said they liked stories, liked reading (usually with parents or at school; few could read on their own), and liked making up their own stories.

Figure 3: Two DragonBots, designed to look like fluffy baby dragons.

4.2.3 Robot

The robot companion used the existent DragonBot platform, a fluffy robot designed to be appealing to children (Freed, 2012; Kory, Jeong, & Breazeal, 2013; Setapen, 2012). This platform has several advantages. Most physical robots are difficult to build, expensive, often do not scale well as a deployable technology, and may not be very expressive or customizable. However, this robot was designed with field studies with children in mind. First, the robot's design is based on "squash and stretch" principles of animation, which creates more natural and organic motion and allows for a range of expressive body movements, while keeping the actuator count low. Second, this robot is based around an Android phone. The phone is the robot's computational power, and its screen displays the robot's face. Because the face is animated, it can be highly expressive and customizable. In addition, the phone's sensors are used to support remote presence interactions. This Android-driven platform affords rapid customization, accessibility, and scalability. The robot can easily be taken out of the lab for testing with children.

Figure 4: The DragonBot can be controlled with this interface, which runs on a tablet. Buttons on the left trigger expressive animations. The center picture shows a live camera feed of what the robot sees. It can be touched to direct the robot's gaze. The slider on the right controls facial expressions.

The robot can be controlled in Wizard-of-Oz fashion by a human. The robot's gaze, facial expressions, and actions can be controlled. The wizard can live-stream his or her voice to the robot, or play pre-recorded speech. These capabilities allow the robot to appear as a social agent while avoiding the difficulties of children's speech recognition and natural language processing.

In this thesis, the speech the robot used was scripted. A dialogue system presented the wizard with limited dialogue options. The wizard selected the most appropriate option for the robot to say. The wizard also triggered the robot's facial expressions and animated motion. However, the robot's gaze was automatically directed to look either up at the child or down at the storytelling game, based on data collected during the pilot study regarding when children would look up at their play partner or down at the game during play.

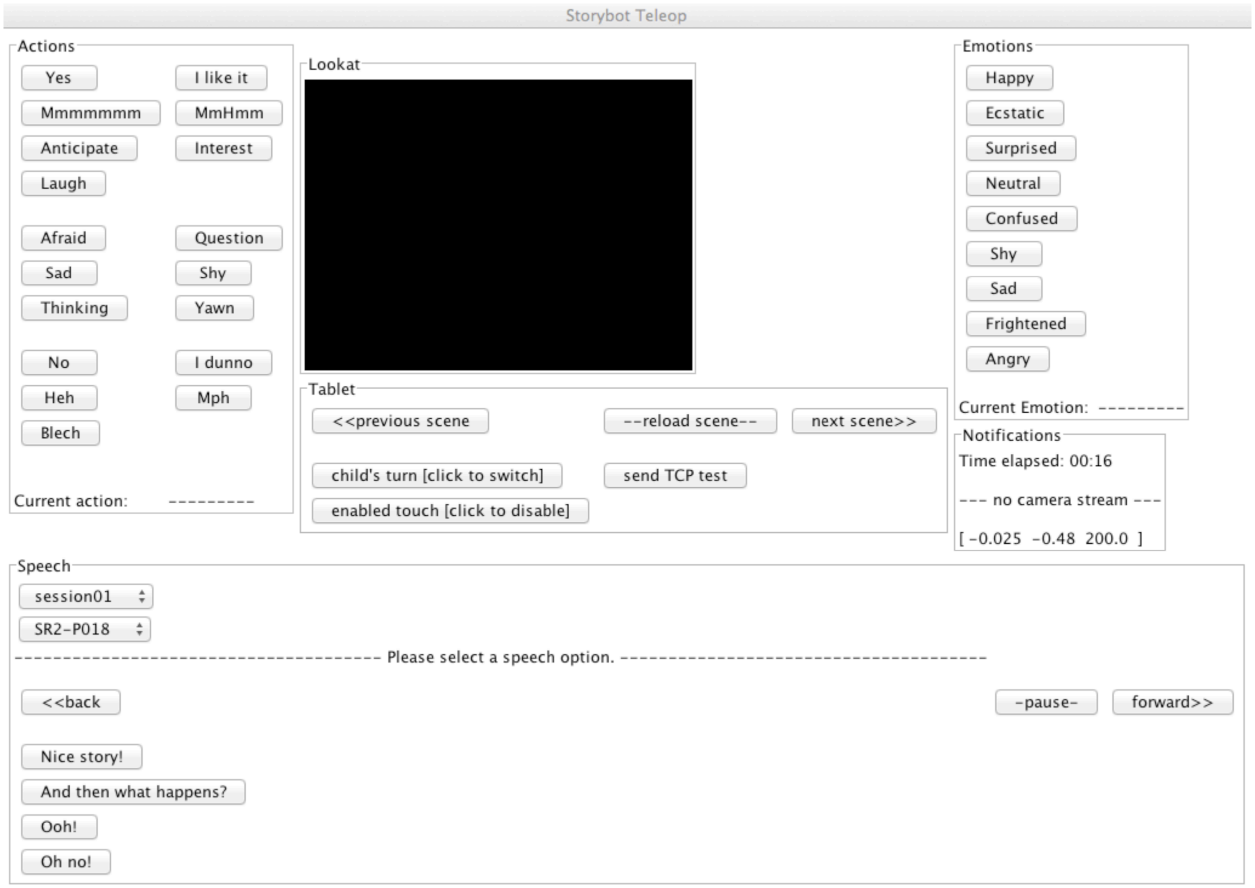

Figure 5: An interface for controlling the DragonBot via a computer, created for this work. It includes some of the same controls as the tablet version: Buttons on the left trigger expressive animations. The center box can show a live camera feed of what the robot sees. It can be clicked to direct the robot's gaze. Buttons on the right controls facial expressions. This interface also has some new additions: There are buttons in the center that send commands to the tablet running the story game, which can, for example, enable or disable touch events on the tablet, and advance to the next scene in the story game. Finally, the buttons at the bottom present the robot's dialogue options, including options to move forward or backward in the robot's dialogue script.

4.2.4 Robot role

Much of the existing work on robotic learning companions has put the robot in the role of a teacher or tutor, not as a peer. However, making the robot a peer allows the child to take on the role of "friend" or "teacher" rather than "student" (e.g., Tanaka & Matsuzo, 2012). This changes the dynamic of the relationship. Some children may feel more free to talk and play when interacting with a peer. As such, the robot in this interaction was designed to be a peer. It talked to the child as an equal, modeling speech and stories without explicitly teaching or leading.

4.2.5 Adaptation & Conditions

As was discussed earlier, children may learn best at different paces or through different methods. A learning/teaching companion that tailors its behavior, language, or an activity's content to an individual's needs through time may lead to greater learning outcomes. Many possible avenues of adaptation are available, including (a) adapting to the child's interests, (b) adapting to the child's particular learning style, (c) adapting social behaviors to the child's specific communication style, and (d) adapting to the child's language ability, to name just a few. Each means of adapting has it own difficulties and merits; it is outside the scope of this thesis to discuss them all here. In this thesis, I focused on a robot that adapted the “level” of the language it used, such that the robot spoke as a slightly older peer relative to the child. Ryokai et al. (2003) presented preliminary evidence that by taking turns telling stories with a slightly more advanced virtual peer, children's stories became more sophisticated. This follows from Vygotsky's (1978) “zone of proximal development,” which suggests that children will learn most readily when slightly challenged, and thus, that a more competent child can help a less competent child learn.

The robot adapted the language level used in the stories it told during the storytelling game, based on the child's current language ability. The child's language ability was measured via a language and vocabulary assessment and then classified as either “high ability” or “low ability.” The assessments are discussed in depth in Section 4.2.9. Nine children were classified as “high” and eight children as “low.” So the adaptive robot told easier, simpler stories to children of low ability, and harder, more complex stories to children of higher ability. The non-adaptive robot told easier, simpler stories to all children. Figure 21 shows a table indicating how many children were in each condition in relation to their language ability and the level of the robot's stories. For example, a more advanced story (vs. an easier story) used slightly more complex sentence structures, included more abstract words (vs. concrete words), was slightly longer (more total words), and had increased narrativity and cohesion. The intent was for the robot to model good storytelling and more advanced uses of language for the child. I expected that this might implicitly lead children to improve their own stories, through copying the robot and the increased exposure to the more advanced stories.

Number of children	Condition	Child language ability	Robot level
4	Adaptive	High	Hard
5	Adaptive	Low	Easy
5	Non-adaptive	High	Easy
3	Non-adaptive	Low	Easy

Table 2: This table indicates how many children were in each condition, when divided by their language ability. The table also shows the level of the robot's stories for each condition.

In addition, the robot's dialogue outside of the storytelling game varied each session. Adapting a robot's dialogue over time can help build rapport and increased the user's continued engagement (Kasap & Magnenat-Thalmann, 2012; Kidd & Breazeal, 2008; Lee et al., 2012). For example, Autom, a weight loss coach robot, varied its dialogue through a combination of interaction scripts and factors including the time of day, time since last interaction, and the estimated state of the relationship with the user (Kidd & Breazeal, 2008). The robot used non-present speech to talk about what it and the human had done yesterday, or might do tomorrow.

Inspired by these systems, the robot here greeted children differently each play session, asked different, relevant questions about the child's day, and made references to past play sessions during conversation. I should note, however, that these variations were scripted for the robot; they were not dynamically generated. Future work could include autonomous generation of varied dialogue.

4.2.6 Robot dialogue & teleoperator

The robot has a scripted set of dialogue options to lead the child through the interaction. The robot's stories were pre-scripted. The full text of the stories and interaction scripts are included in the supporting materials (available for download from http://jakory.com/storybot/). The robot's speech was triggered by a human operator – i.e., a human controlled the timing, so that the robot spoke at the right times and did not interrupt the child. Sometimes, what the robot said depended on what the child said. In these cases, the human operator selected among a couple dialogue options for the robot (the human was kept in the loop primarily because of the lack of good automatic speech recognition for children). However, there was a main storyline that the robot always returned to. In addition, as noted earlier, each introductory chat and each closing chat followed the same format with slight variations. Finally, I should note that the robot was named “Green” and referred to as such, in a distinctly non-gendered way, by the experimenter throughout the study.

With regards to the specific choices the teleoperator made when selecting robot speech and actions, several general rules were followed. First, the teleoperator made the robot's behavior as socially contingent as possible – reacting to the child as closely to as a human would in the same circumstance. When the child spoke, the robot would acknowledge through speech, verbal exclamations such as “Ooh!” and “Oh no!”, smiles, and short affirmative non-linguistic noises. These acknowledgements were primarily triggered during pauses in the child's speech. The same sounds or animations were not triggered twice in close succession, though the same sounds and animations were often used multiple times per session. Finally, the teleoperator made the robot's behavior as consistent as possible across participants, using the same set of sounds and animations with approximately the same frequency for all children.

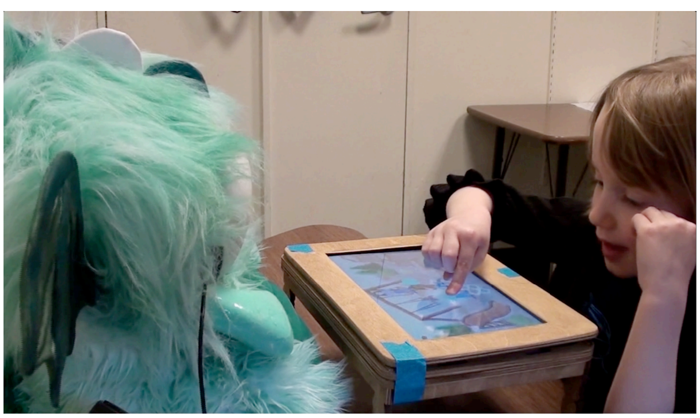

Figure 6: A child tells a story to the robot and moves the characters on the tablet screen while playing the storytelling game.

4.2.7 Stories

As mentioned earlier, the robot told stories at different levels of complexity and introduced target vocabulary words during its stories. The stories written for the robot were based on stories told by children during pilot testing of the storytelling game.

There are eight scenes in the storytelling game. Two stories were written for each scene, for a total of sixteen stories. Each of these stories was manipulated to create two versions of the story – one easier, simpler version (EASY), and one harder, more complex version (HARD). The dimensions manipulated were narrativity, sentence length, word frequency, syntactic simplicity, and referential cohesion. I assessed significant differences in story difficulty by comparing the EASY and HARD stories on two measures: (1) Flesch-Kincaid Grade Level (FKGL), (2) Coh-Metrix (a text-analysis software, McNamara et al., 2013) indices of difficulty (narrativity, syntactic simplicity, word concreteness, referential cohesion, deep cohesion – see Table 3), where higher values of the indices are generally easier to read and understand (they range from 0-100). In the future, I could compare subjective human ratings of the stories as well.

I ensured that the FKGL were at least one grade level different (mean 2.2 grade levels different). EASY stories were grade 2.4 or below (mean 1.8); HARD stories were grade 3.2 and up (mean 4.0). Note that FKGL was intended to be a measure of reading level, not oral language difficulty, so a child who cannot read at a FKGL of 2 may be perfectly able to understand spoken speech at that level. In addition, all EASY stories were shorter (mean word count 200) than the HARD stories (mean word count 228).

I ensured that the EASY and HARD texts differed significantly on all but one of the Coh-Metrix indices (average $p < 0.05$), with the EASY stories having higher scores than the HARD stories on all indices except narrativity (higher for HARD – more complex stories may include more narrative structure). I purposefully varied deep cohesion scores across the stories, though I expect to find that more complex stories have higher deep cohesion scores.

Dimension	Attributes
Narrativity	score is high when a text conveys a story with a sequence of actions and events; affiliated with word familiarity and prior knowledge
Syntactic Simplicity	score is high for sentences with fewer words and clauses, more common structure; fewer words before the main verb
Word Concreteness	score is high if more words are concrete versus abstract
Referential Cohesion	score is high when words and ideas are connected across sentences, such as words that appear in adjacent sentences
Deep Cohesion	score is high when a text contains more connective words that help clarify relationships between events, ideas, and information
FKGL	score is high when a text contains longer sentences and words with more syllables
Word count	score is high if text contains many words

Table 3: A summary of the text dimensions analyzed, listing what attributes of a story impact each dimension.

4.2.8 Vocabulary words

Twenty-four vocabulary words were selected from Andrew Biemiller's "Words Worth Teaching" lists (Biemiller, 2010). Specifically, words were chosen from the list of words the majority of children should know by the end of second grade that could also easily fit into the context of the story scenes. These included 10 nouns such as *structure, chunk,* and *moat;* 12 verbs such as *expect, reveal,* and *wonder*; and 2 adjectives such as *massive* and *ancient*. Table 4 shows the full word list. Each story has three key vocabulary words presented. However, the two stories about the same scene use two of the same vocabulary words as well as one word that is shared with a story from a different scene. For example, both of the robot's stories for the meadow scene that has a butterfly character have the key words *flutter* and *wonder*, but one story also has the key word *reveal* while the other has *plunge*. This way, some words (like *plunge*) are encountered in more than context (the meadow scene with the butterfly and a second scene).

Vocabulary word list

ancient (adj.)	flutter (v.)	slip (v.)
approach (v.)	massive (adj.)	sofa (n.)
board (v.)	moat (n.)	span (v.)
buttercup (n.)	palm (tree) (n.)	spruce (n.)
chunk (n.)	pass (v.)	structure (n.)
clump (n.)	plunge (v.)	whiff (n.)
discover (v.)	reveal (v.)	wish (v.)
expect (v.)	shuttle (n.)	wonder (v.)

Table 4: List of target vocabulary words. The words were selected from the "Words Worth Teaching" list (Biemiller, 2010) of words that children should know by the end of the second grade, that also fit into the context of the robot's stories.

4.2.9 Learning measures

Children's learning was measured in three primary ways. First, during the first session, children were given a standard language assessment. This assessment was a subset of the PLS-5 (Zimmerman et al., 2011) that measured aspects of expressive communication and auditory comprehension that were especially relevant to narration and storytelling. Table 5 shows a full list of the questions included from the Auditory Comprehension (AC) questions and the Expressive Communication (EC) questions. Scores on this assessment could range from 0 (no items correct) to 18 (all items correct).

Second, children were assessed on the target vocabulary words introduced during the robot's stories. Six words were introduced per session during the first four play sessions; during the latter four play sessions, all the vocabulary words had appeared in a story before and were heard again in new stories. Children were tested on each word four times:

- time 1: at the beginning of a session, before the word was introduced in a story

- time 2: at the end of a session, after the word was introduced in a story
- time 3: at the beginning of a later session, after a delay of approximately one week
- time 4: at the end of a session, after the next time the word was used in a story

For example, if the word *expect* was first used in a story in Session 2, then children would be tested on the word (i) at the start of Session 2, before playing with the robot; (ii) at the end of Session 2, after playing with the robot; (iii) at the start of Session 3, approximately one week later, before playing with the robot again; and (iv) at the end of the next session where the word appeared in a story. This assessment was a picture-based test based on the Peabody Picture Vocabulary Test (PPVT, Dunn & Dunn, 2007). For each target word, four images were shown to the child. The child was asked to point to image depicting the target word. This measure was used to determine whether children learned any target words while playing with the robot.

Finally, children's language use during the storytelling game was analyzed for content, structure, and complexity using Coh-Metrix and measures such as the mean length of utterances (MLU), word count, and the type-token ratio. These measures were used to determine whether children's stories grew longer or more complex after playing with the robot.

Children's scores from the PLS-5 subset and their scores for the time 1 vocabulary assessment (before each word was introduced) were used to classify children as “lower language ability” or “higher language ability.” Children were classified as “higher language ability” if they achieved two of the following three: PLS-5 AC score of 9 (all the AC items correct), PLS-5 EC score of 8 or 9 (nearly all the EC items correct), and/or a time 1 vocabulary score of 12 or higher (i.e., knew half or more of the target words). These classifications were used in determining the robot's language level for children in the *adaptive* condition.

	Auditory comprehension		**Expressive communication**
AC.46	Understands complex sentences	EC.49	Uses modifying noun phrases
AC.48	Understands modified nouns	EC.51	Repairs semantic absurdities
AC.52	Understands time/sequence concepts	EC.58	Retells story with introduction
AC.53	Recalls story details	EC.59	Retells a story with sequenced events
AC.54	Identifies a story sequence	EC.60	Retells a story with a logical conclusion
AC.55	Identifies the main idea	EC.61	Formulates sentences
AC.56	Makes an inference	EC.62	Uses synonyms
AC.57	Makes a prediction	EC.64	Uses past tense forms
AC.62	Makes grammatically correct judgments	EC.67	Uses time/sequence concepts

Table 5: Items selected from the PLS-5 that target aspects of auditory comprehension and expressive communication especially relevant to narration and storytelling.

4.2.10 Engagement and perception measures

To assess children's enjoyment of the play sessions, every session, the robot would ask the child whether he or she would come back and play again later. In addition, after children's fourth session with the robot (midway interview), and again after their last session (final interview) with the robot, they were asked questions regarding how much they liked playing the story game with the robot. They were also asked questions pertaining to their impression of the robot as a social and mental other; some of these were adapted from questionnaires used by Kahn et al. (2012). The questions included:

1. How much do you like playing the story game? A lot, a little bit, or not very much?
2. How much do you think Green likes playing the story game? A lot, a little bit, or not very much?
3. How smart is Green? Really smart, kind of smart, or not very smart?
4. Is Green smarter than you, or are you smarter than Green?
5. Can Green be happy?
6. Can Green be sad?
7. Does Green have feelings?
8. Do you like playing with Green, or are you a little bored by Green?
9. If you were lonely, do you think you might want to play with Green?
10. What would you do if Green was sad?
11. Do you think you could trust Green with a secret?
12. Is Green your friend? (if no) Can a robot be your friend?
13. How much is Green like one of your friends? A lot, a little bit, or not very much?

After the last session (during the final interview), children were asked additional questions regarding their perception of Green as an animate living being versus inanimate artifact, and their impressions of the stories the robot told during the game. The first eight questions below are taken from a study by Jipson and Gelman (2007) that investigated children's inferences about living and nonliving kinds, with two questions pertaining to each of aliveness, psychological abilities, perceptual abilities, and qualities of being an artifact.

1. Does Green eat?
2. Does Green grow?
3. Can Green think?
4. Can Green be happy?
5. Can Green see things?
6. If I tickled Green, would Green feel it?
7. Did a person make Green?
8. Can Green break?
9. Do you think you learned anything from Green?
10. Is Green a boy or a girl?
11. How much is Green like you? A lot, a little bit, or not very much?
12. Is Green alive? How alive is Green? A lot, a little bit, or not very much?
13. How much like a machine or a computer is Green? A lot, a little bit, or not very much?
14. How much did you like Green's stories? A lot, a little bit, or not very much?

15. Were they really interesting, kind of interesting, or kind of boring?
16. Were they easy to understand, kind of easy to understand, or hard to understand?
17. Would you want to play the story game with Green again? A lot, a little bit, or not very much?
18. Would you want to play the story game with your friends? A lot, a little bit, or not very much?

Finally, children were asked about the similarity of the robot to persons versus iPads/tablets, a series of questions that we have asked children before (unpublished):

1. When a robot remembers something, is it more like a person or more like a iPad?
2. When a robot tells you something you didn't know, is it more like a person or more like a iPad?
3. When a robot answers a question, is it more like a person or more like a iPad?
4. When a robot thinks about something, is it more like an iPad or more like a person?
5. When a robot teaches you something, is it more like an iPad or more like a person?
6. When a robot is interested in something, is it more like an iPad or more like a person?

4.2.11 Data analysis

Audio and video of children's interactions with the robot were recorded with a microphone situated near the robot, a camera behind the robot facing the child, and for some sessions, a lower-quality web camera behind the child facing the robot. All touch actions on the tablet screen in the story game were recorded: finger-down (finger touches screen), tap (finger taps screen), drag-begin (start of a dragging motion), drag-move (during a dragging motion), drag-end (end of a dragging motion). Children's responses to the PLS-5 subset, the vocabulary assessments, and interview questions were also recorded.

The recorded audio was used to transcribe and code children's speech. Coding conventions and code entry into the transcripts were consistent with the SALT software (Miller & Chapman, 2012). From the transcripts, the text of each of the children's stories was extracted. Stories of twenty words or longer were analyzed to get the five Coh-Metrix scores, Flesh-Kincaid Grade Level (FKGL), and word count. These scores for stories told by the same child in the same session were averaged to get a single score for each measure, for each child across each session.

In addition, the full transcripts were analyzed with the SALT software (Miller & Chapman, 2012) to obtain other language measures, such as the mean length of utterances, words per minute, and the type-token ratio. The logs of tablet touch actions were analyzed to determine the number of times each action was performed each session. During analysis, drag-end actions were determined to give the same information as drag-begin actions (being the start and end of the same drag action), so the results for the drag-end actions are not reported. Finally, videos of the children's sessions were viewed to gain a qualitative understanding of children's behavior across sessions. Future work includes more detailed coding of the videos.

4.3 Results

4.3.1 Engagement and perception measures

4.3.1.a Play again

Each session, Green asked children if they would come back and play again. Most children said they would. Some were shy and did not respond. One child consistently misinterpreted the question: thinking the robot was asking if she would come back and play again *that day* (rather than *later another day*), she always insisted that she could not, she had to go back to class. At the end of one session, one boy followed his enthusiastic "Of course!" by asking if Green would be coming to the preschool's picnic that evening.

4.3.1.b Midway interview

The midway interview about children's perceptions of the robot revealed that the majority of children believed that Green had mental states insofar as they said that Green was smart (88.2%), liked the story game (94.1%), could be happy (100%), could be sad (58.8%), and had feelings in general (88.2%). For example, one child said, "I know Green likes to play with me, so I know he's happy." Another said, "Yes, everybody has feelings, even robots do!" The difference in perceptions of the robot's ability to be happy or sad is likely due to the fact that the robot was portrayed as an optimistic, happy character. Sad animations were rarely used. Interestingly, despite so many children saying that Green was smart, 82.4% of children said that *they* were smarter than the robot! One child justified this judgment by saying that, "The wrong dinosaurs in the story ran fast and slow. Green didn't know that."

The majority of children also believed the robot was a social other insofar as they said that they liked playing with Green (88.2%), might want to play with Green if they were sad (82.4%), believed they could trust Green with a secret (70.6%), might comfort Green if Green was sad (88.2%), that Green was their friend (88.2%), and that Green was similar to their other friends (70.6%). Children who dissented usually said they were not sure, rather than saying "no" directly. With regards to sharing secrets with Green, two different children said they could not "because you can't tell secrets at school." Other justifications included: "No, because she's a dragon," and, "Yeah, she couldn't tell anyone else because she's a robot." When asked what they would do if Green was sad, most children said they would play with Green, though one child said he would "buy ice cream to make him happy, robot ice cream."

The final questions pertained to the storytelling game. The majority of children said they liked playing the game a lot (76.5%); only a few said they liked it a little (11.8%) or not very much (11.8%). In addition, children were very good at understanding the robot's expressed feelings about the game: 94.1% indicated that Green liked playing the story game very much.

4.3.1.c Final interview

The final interview session repeated the same questions as the midway interview, and added several more that further probed children's perceptions of Green. There were no significant changes in children's responses to the midway interview questions. One child did not complete the final interview, so the responses of only 16 children are considered here.

First, a minority of children believed the robot was alive, insofar as they said that Green eats (33.3%) and Green grows (40.0%). One child justified this judgment by saying, “yes, dragons grow!” When asked directly if Green was alive, responses were evenly divided: 26.7% said Green was very alive, 26.7% thought Green was kind of alive, and 40.0% thought Green was not alive. One child justified his response that Green was alive by saying, “He's a robot!”, while another said, “because she can talk.” Another said Green was *not* alive because he was “just a robot, not a real dragon.” One child said the question was too hard “because if people made him then he's not alive.”

A majority of children, however, believed the robot had psychological abilities, insofar as they said that Green could think (73.3%), could be happy (100%), and was like themselves (81.3%). They also believed the robot had perceptual abilities, insofar as they said that Green could see things (93.3%) and could feel tickles (73.3%). One child said, “he can see because he has eyes.” Finally, the majority *also* indicated that Green had some qualities of being an artifact, insofar as they said that a person made Green (80.0%), that Green could break (73.3%), and that Green was very much like a computer or machine (73.3%). For example, one child said, “Robots are computers; they have computers as their brain.”

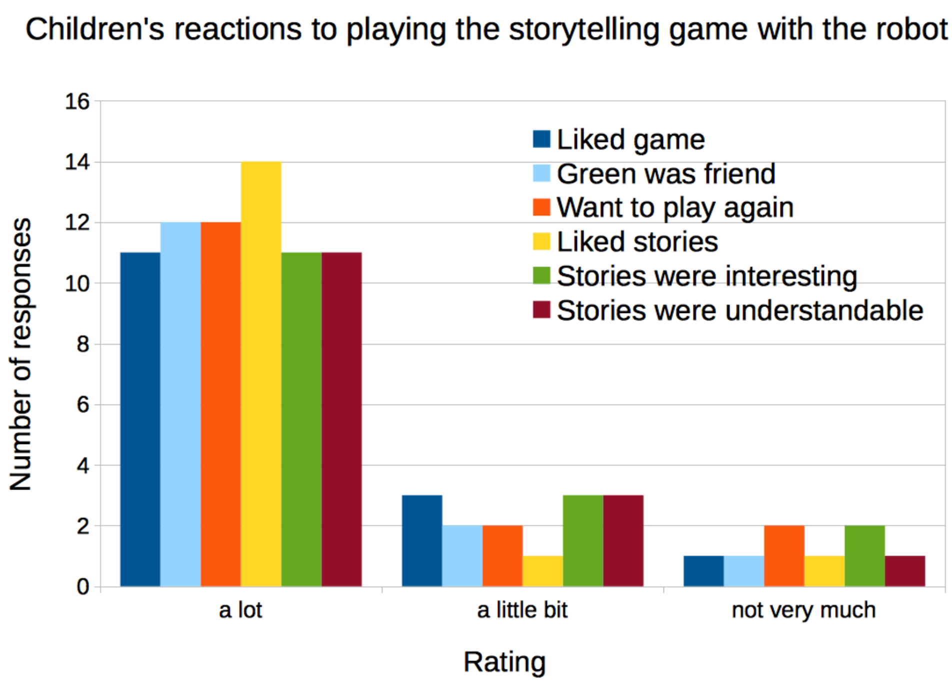

Figure 7: During the final interview after Session 8 with the robot, the majority of children reported that they liked the story game a lot, that the stories were interesting and understandable, that Green was their friend, and that they would like to play the game again.

Perhaps because of the robot's perceived perceptual and psychological abilities, the majority of children indicated that they thought the robot was more like a person than like an iPad or tablet when the robot answers a question (68.8%), thinks about something (68.8%), teaches them something (81.3%), or is interested in something (87.5%). They were split on two questions, with about half the children saying the robot was more like a person when it remembers something (43.8%), or when it tells them something they did not know (56.3%). Children's justifications that Green was like a person included “because people do that,” “because iPads don't think,” and “because people answer questions.”

When asked the robot's gender, 52.9% said Green was a boy and 41.2% said Green was a girl. One child did not answer. For example, one child said that Green was a boy because “dragons are boys.” The robot's gender matched their own for 75% of the children who responded. For the mismatches, three girls thought Green was a boy and one boy thought Green was a girl. Although this question regarding the robot's gender was asked of children during the final interview, most children had considered the robot's gender previously and had been using either feminine or masculine pronouns to refer to Green. Some children had asked the experimenter if Green was a boy or a girl; the experimenter responded by saying, “What do you think?” to encourage children to make their own decisions. Recall that the experimenter always referred to Green in a non-gendered way.

Finally, a majority of children said they liked playing the storytelling game with Green (93.8%) and would want to play the game with Green again (87.5%). A majority also liked the stories Green told, insofar as they said that the stories Green told were either very interesting (68.8%) or somewhat interesting (18.8%), and that the stories were easy to understand (68.8%). In addition, 87.5% said they would want to play the story game with their other friends sometime.

4.3.2 PLS-5 language ability

Children's scores on the PLS-5 subset were overall quite high: Out of a maximum score of 9 on the PLS-5 AC questions, no children scored below 7. The mean score was 8.56 (SD=0.73) and the median score was 9. The PLS-5 EC questions had more variation, but scores were still generally quite high. Of a maximum score of 9, the mean score was 6.89 (SD=1.76) and the median score was 7. Children's mean combined score (of a maximum of 18) was 15.4 (SD=1.94); the median combined score was 16. There was no significant difference in scores across children from different preschools, suggesting that their overall language abilities were comparable.

4.3.3 Vocabulary measures

Children were tested on each of the 24 target vocabulary words four times, as described in Section 4.2.9. At time 1 (before the robot used the words), children knew a mean of 13.4 words (SD=3.62). They improved over time (time 2 M=14.9, SD=2.94; time 3 M=15.9, SD=2.87), with time 4 scores (M=18.9, SD=2.84) significantly higher than time 1 scores, $t(14)=7.21$, $p<.001$. Figure 8 depicts this progression of scores over time. As can be seen, there appears to be improvement from time 2 (immediate recall) to time 3 (delayed recall). I expected to see the opposite pattern, with time 3 scores lower due to some forgetting over time. This result could be explained from observation of children's behavior. Since time 1 and time 2 words were presented at the beginning and the end of the same session, some children would simply repeat the (possibly incorrect) answers they gave the first time. As such, it may be that one week later, when they no longer remembered their previous answers, they would feel more free to select the correct answer if they knew it, reflecting their actual knowledge of the words.

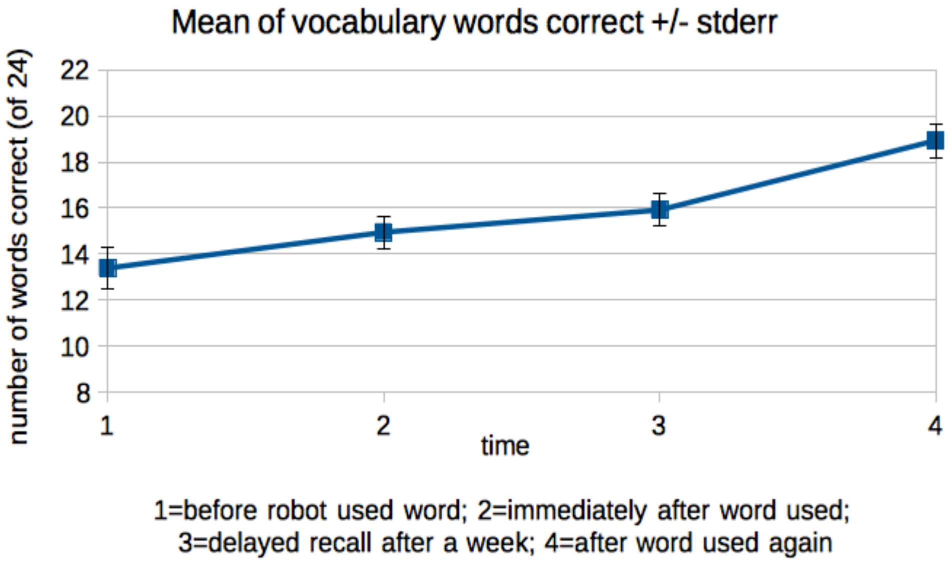

Figure 8: The mean number of vocabulary words children got correct during testing at four points in time, +/- standard error: (1) before the robot used the word; (2) immediately after the robot used the word in a story; (3) delayed recall, about a week later; and (4) after the robot used the word again, in a later session.

An examination of the data split by preschool revealed that at all time points, children from Preschool 1 knew significantly more of the target vocabulary words than children from Preschool 2, as shown graphically in Figure 9. The mean scores at each preschool at each point and the results of independent samples t-tests showing that the scores are significantly different are listed in Table 6. This result may have been seen because the children at Preschool 1 were slightly older, and thus may have had more knowledge at the outset of the study. However, as noted earlier, children's PLS-5 scores were not significantly different between preschools, so while children at Preschool 1 knew more words, their overall language abilities were comparable to children at Preschool 2.

	Time 1		**Time 2**		**Time 3**		**Time 4**	
	Mean	***(SD)***	**Mean**	***SD***	**Mean**	***SD***	**Mean**	***SD***
Preschool 1	15.7	*2.96*	16.7	*2.59*	18.1	*1.94*	20.6	*1.30*
Preschool 2	10.8	*2.24*	12.9	*1.84*	13.5	*1.41*	17.0	*2.94*
t(15)	3.81		3.44		5.53		2.95	
p	0.002		0.004		<.001		0.011	

Table 6: This table shows the mean vocabulary score from children at each preschool at each time point, and the results of the independent samples t-test that shows the scores at the two preschools are significantly different. Children from Preschool 1 knew more words than children from Preschool 2.

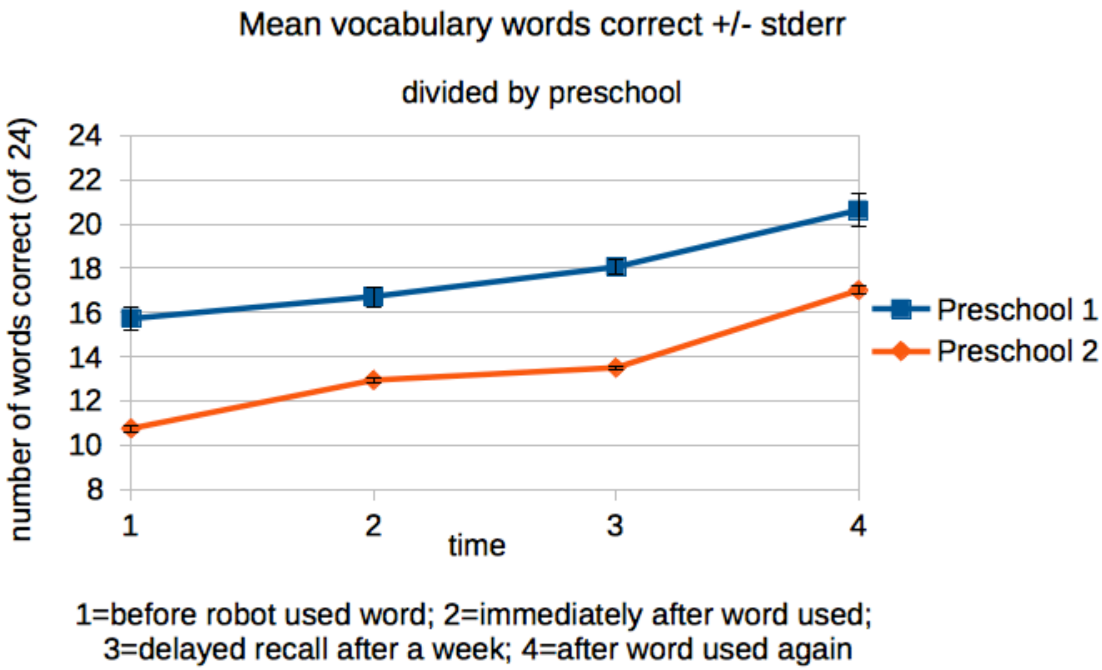

Figure 9: This graph shows the mean number of vocabulary words that children got correct, divided by which preschool attended, during testing at four points in time: (1) before the robot used the word in a story; (2) immediately after the robot had used the word in a story; (3) delayed recall, about a week later; and (4) after the robot used the word in a story again, in a later session. These results show that children at Preschool 1 had higher vocabulary scores at all times than children at Preschool 2.

Examining the data split by gender and age revealed no differences. However, the majority of children (76.4%) were the same age (5 years) and the number of children in the study was small (N=17). If the sample size was larger, I would expect to see some differences due to age, since children rapidly acquire new words during their preschool years. Six-year-olds would be expected to know more words than four-year-olds, for example.

Next, I examined whether children who told stories aloud learned more vocabulary words from time 1 to time 4 than children who told silent stories or no stories (for more discussion regarding this division, see Section 4.3.6). There were no significant differences, which indicates that actively telling a story to the robot was not necessary for learning words from the robot. Even children who did not speak aloud learned words.

I also wanted to know whether children's scores were predicted by initial language ability (low or high) or the robot's story level (easy or hard). I found that categorizing children by condition (adaptive or non-adaptive) was less useful than categorizing them by language ability and robot story level, due to the fact that children with low ability in the non-adaptive condition were treated the same as children with low ability in the adaptive condition: they both heard easy stories from the robot, as that was the robot's default level. Only children with high ability in the adaptive condition heard hard stories. As such, children were classified into three groups: easy-low, easy-high, and hard-high. The vocabulary scores for each group are plotted in Figure 10.

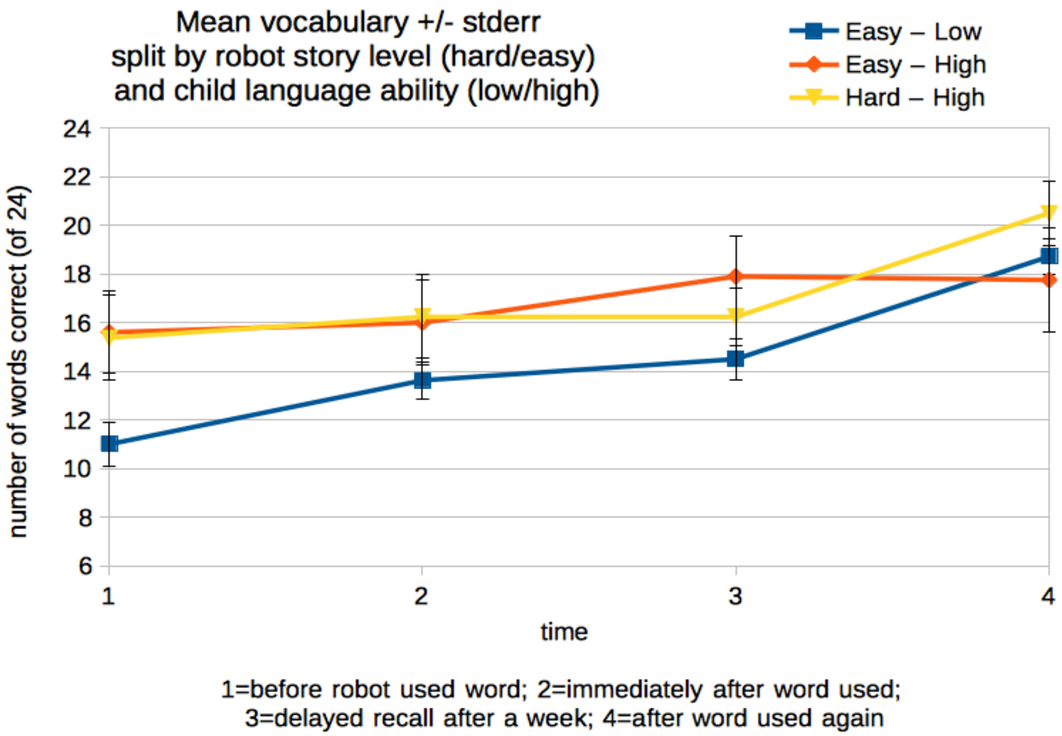

Figure 10: The mean vocabulary scores by group at four points in time, +/- standard error. Note that the first time the robot uses the target words (measured at time 2, delayed recall at time 3) was during the first four sessions (introducing a few words each session). During the first four sessions, the robot always told easy stories; for the second four sessions, the robot adapted, telling harder stories to the children in the hard-high group, and continuing to tell easier stories to the children in the easy-low group. The robot did not adapt for children in the easy-high group. The second time the robot used the target words was during the second 4 sessions, (measured at time 4). As the graph shows, from time 3 to time 4, the orange and yellow lines diverge, indicating that higher ability children to whom the robot told harder stories learned more words than those to whom the robot told easier stories. However, this trend was not significant.

The effect of group on the number of vocabulary words learned was examined. Mean words learned (of 24 words total, with a mean starting knowledge of 13.4 (*SD*=3.62) words known) was determined for each of the three groups by subtracting the number of words known at time 1 from the number of words known at time 4: easy-low = 7 participants (*M*=7.93, *SD*=1.34), easy-high = 4 participants (*M*=2.50, *SD*=2.08), hard-high = 4 participants (*M*=5.13, *SD*=3.28). Mean words learned from time 1 to time 4 differed significantly among these three groups, $F(2,14)=8.27$, $p=.006$. The eta-squared value was 0.579, which indicates that 57.9% of the variance in the sample was explained by which group the child was in. However, this should be interpreted cautiously, as the eta-squared measure can overestimate the variance explained, especially with small sample sizes.

Post-hoc comparisons with the Bonferroni test indicated that the easy-low group's mean words learned score was significantly higher than the easy-high group. There was no significant difference between the easy-high group and hard-high group, though the mean words learned by the hard-high group was greater, trending in the expected direction. The mean words learned by the hard-high group and the easy-low group was not significantly different. These results indicate that when the robot adapted to the child's language level (i.e., telling easy stories to children of lower ability and hard stories to children of higher ability) the children learned more words on average than when the robot did *not* “level up” with the child (i.e., telling easy stories to children of higher ability).

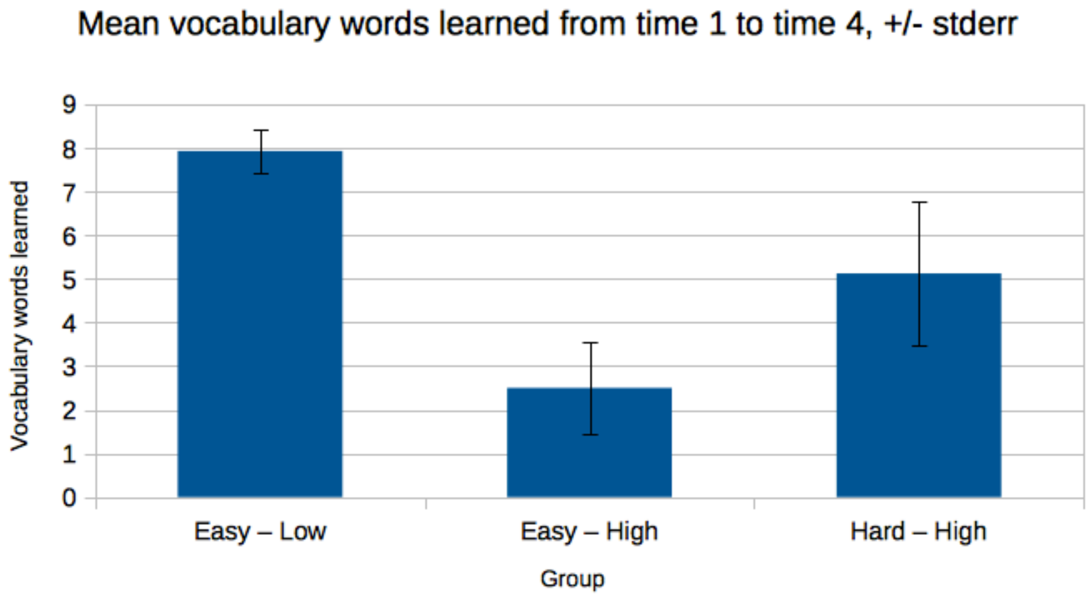

Figure 11: This graph shows the mean vocabulary words learned from time 1 to time 4, with 95% confidence intervals. The easy-low group learned significantly more words than the easy-high group. There were no significant differences between the other groups, though the trends were in the expected direction.

The 95% confidence interval for the mean of each group is shown in Figure 11. The confidence interval for the easy-low group does not overlap the interval for the easy-high group, which reflects the significant difference found between those means. However, the intervals among the other groups overlap substantially. In large studies, this can indicate that the population means for the those groups are not likely to differ. However, the N in this study was quite small, and so these results should be interpreted cautiously. It may be that the hard-easy and hard-high groups would differ significantly if more children were tested.

More analysis could still be done regarding the vocabulary that children learned. Specifically, it may be worthwhile to delve deeper into the context surrounding the specific words in relation to whether or not the child learned them. For example, was the child looking at the robot, at the tablet, or somewhere else when the robot used the target word? Was the child paying attention? How audible was the robot's speech? It may be possible, through examining this context, to tie whether the child learned words, and which words were learned to what the child was doing with the robot and how the child construed the robot.

4.3.4 Tablet actions

Across all sessions, children used a mean of 6.79 *drag-begin* actions (*SD*=7.08), a mean of 372.2 *drag-move* actions (*SD*=358.3), a mean of 2.85 *tap* actions (*SD*=3.56), and a mean of 8.68 *finger down* actions (*SD*=9.98). The amount of actions children performed varied widely, as reflected by the high standard deviations in those overall counts. Some children performed no actions at all – either because they did not tell any stories, or because they told stories but without touching the screen. Some children performed many actions (*drag-begin* max=71, *tap* max=370). One child, notably, tapped the screen repeatedly one session in what has been called "woodpecker finger," but because this behavior was rarely seen otherwise, that child's tap and finger-down counts were removed from the means reported above, as they were clearly outliers.

Next, I looked between groups. Mean drag-begin, drag-move, and tap actions for each group, across session, are shown in Figures 12, 13, 14, and 15. Averaging across sessions, the effect of group on the number of each action performed was examined. The means and standard deviations for each action by group, averaged across all sessions, are listed in Table 7. Among the groups, significant differences were seen among the three groups for tap actions, $F(21,2)=8.17$, $p=.002$, eta-squared=.438; finger-down actions $F(21,2)=19.7$, $p<.001$, eta-squared=.652; drag-begin actions $F(21,2)=12.41$, $p<.001$, eta-squared=.542; and drag-move actions, $F(21,2)=7.05$, $p=.005$, eta-square=.402. The eta-squared values indicate that approximately half the variance in the sample was explained by which group the child was in, though, as noted earlier, this should be interpreted cautiously due to the small sample size.

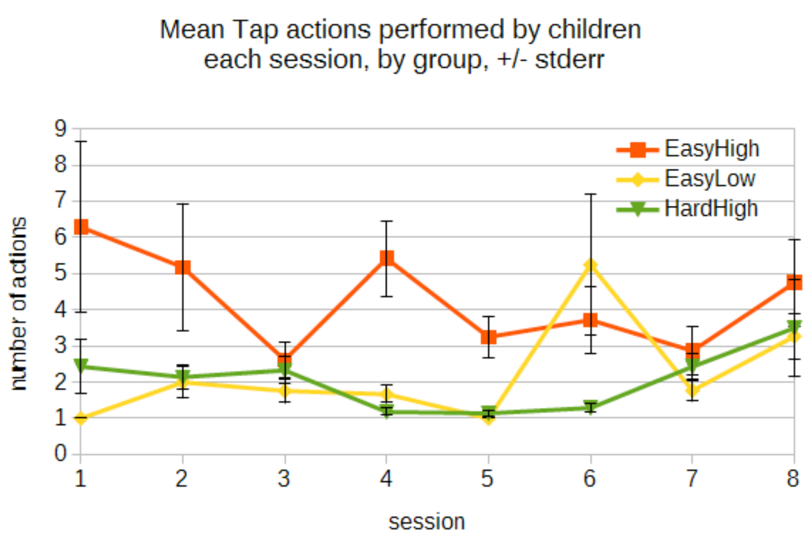

Figure 12: Mean tap actions made by children on the tablet each session, +/- standard error, split by group.

Post-hoc comparisons with the Bonferroni test indicated that the easy-high group's mean actions performed were significantly higher than both the easy-low and hard-high groups for all the actions, but that the easy-low and hard-high groups did not significantly differ on any of the actions (Table 7). These results indicate that when the robot adapted to the child's language level, children performed significantly fewer actions on the tablet than when the robot did not adapt. This may be indicative of boredom on the part of the children in the non-adaptive condition – the behavior needs to be examined in the context of the videos of children's behavior as well as the content of stories children told, which are discussed in the next section. In addition, Figures 12 through 15 show that over time, children used slightly fewer actions overall. This behavior should be examined in the context of what children were doing while they were using the tablet.

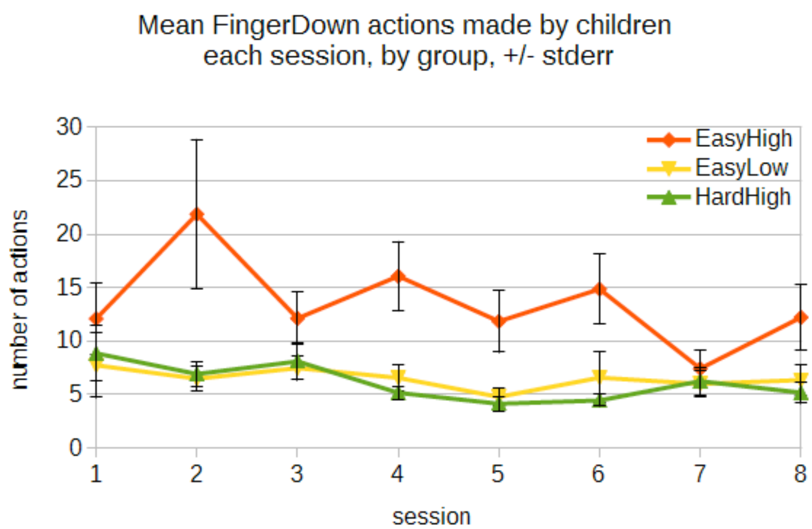

Figure 13: Mean finger-down actions made by children on the tablet each session, +/- standard error, split by group.

I should note that further analysis of the tablet data is ongoing. For example, actions counts by object (e.g., dragon, butterfly, ball) were recorded, but at present, the analysis is over actions on all objects on the screen each session. Further work could be done to analyze actions by story, based on which objects were present in each story scene. This could, for example, help gauge children's interest and engagement in individual story scenes. Additionally, further work could be done in analyzing the change in actions used over time.

Mean actions performed on the tablet, averaged across all sessions

	Tap	**Finger-down**	**Drag-begin**	**Drag-move**
	Mean *(SD)*	**Mean *(SD)***	**Mean *(SD)***	**Mean *(SD)***
Easy-Low	2.22 *(1.42)*	6.52 *(.904)*	5.39 *(.790)*	306.0 *(57.0)*
Easy-High	4.26 *(1.34)*	13.6 *(4.20)*	9.54 *(2.99)*	499.0 *(173.3)*
Hard-High	2.06 *(.814)*	6.15 *(1.71)*	6.66 *(2.84)*	314.5 *(84.9)*

Table 7: The mean actions performed by children on the tablet by group, averaged across all sessions. The actions are listed in columns, while the rows show the scores for each group. The easy-high group performed significantly more of each action than both the easy-low and hard-high groups. The easy-low and hard-high groups did not differ significantly.

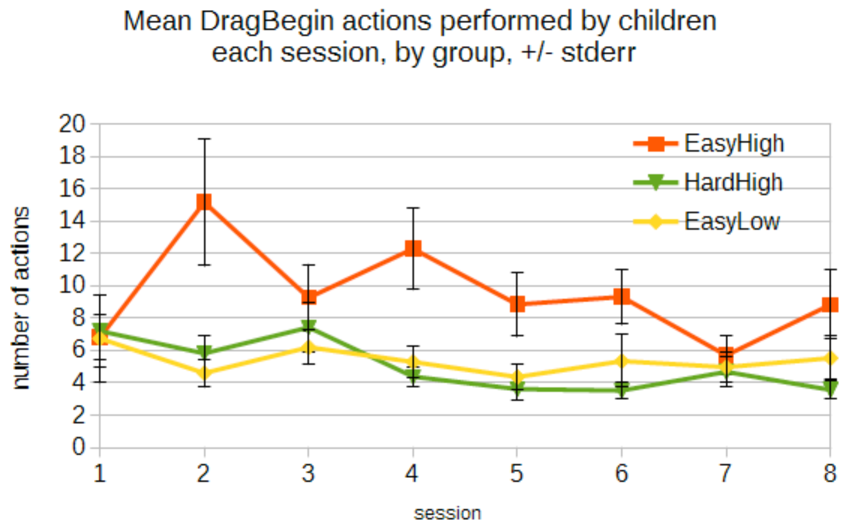

Figure 14: Mean drag-begin actions made by children on the tablet each session, +/- standard error, split by group.

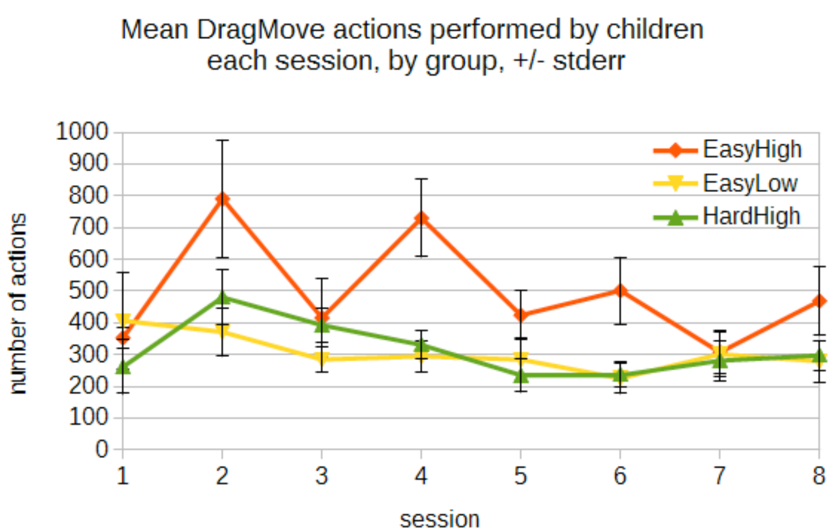

Figure 15: Mean drag-move actions made by children on the tablet each session +/- stderr, split by group.

4.3.5 Overall language

Children uttered a mean of 159.8 words (*SD*=142.7, *max*=634, *min*=0) during each interaction session with the robot (including the introductory chat, story game, and closing). They said a mean of 36.3 utterances each session (*SD*=28.6, *max*=130, *min*=0), with a mean of 15.4 words per minute (*SD*=12.3, *max*=44.7, *min*=0). The large standard deviations are indicative of the range of children's behavior: some children were quiet and shy (one barely said a word, even to the experimenter), while others were boisterous and talkative. The mean MLU (mean length of utterances) in morphemes was 3.79 (*SD*=2.04, *max*=9.4, *min*=0). This is in the general range one would expect for normal children aged 4 (Brown, 1973; Rice et al., 2010), though slightly low for children of age 5. This may be because some children rarely spoke, drawing the mean down.

The mean type-token ratio was 0.60 (*SD*=0.22, *max*=1, *min*=0). This is a slightly higher ratio than some studies of preschool-age children have found (e.g., Watkins et al., 1995). However, the number of words and utterances per child may have been too small (i.e., less than 200 words) to have adequate reliability (Hess, Sefton, and Landry, 1986), which may be why the ratio found here is higher than is typical for this age group.

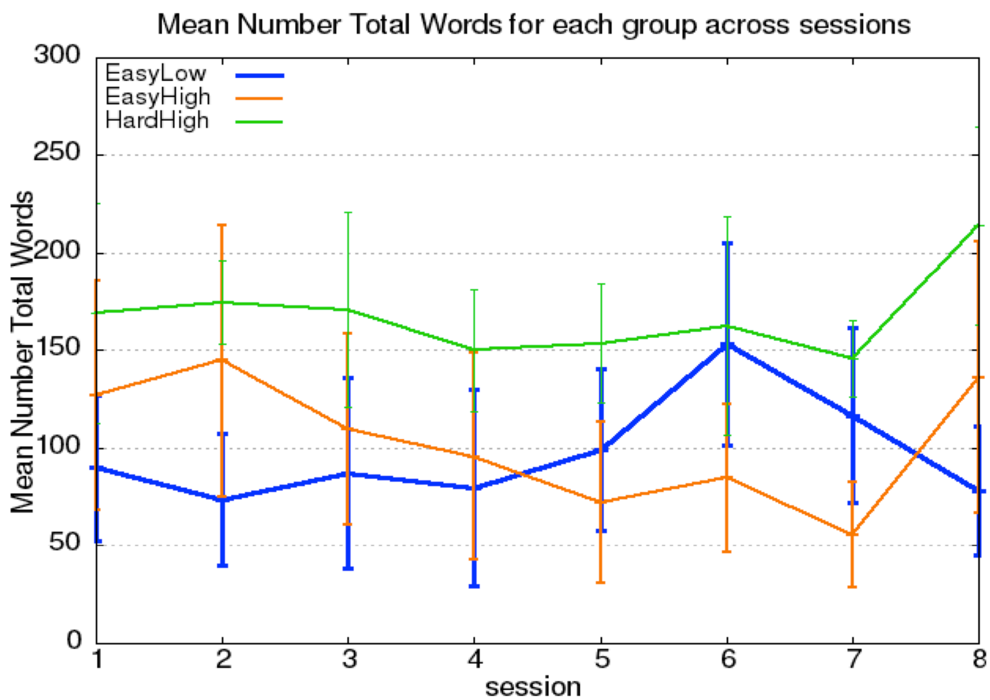

Figure 16: This graph shows the mean number of Total Words spoken by each group across sessions, +/- standard error. Initially, the easy-low and hard-high groups are significantly different, but during sessions 5-8, the easy-low group increases the number of words spoken, while the easy-high group begins to talk less.

In analyzing children's overall language use each session, I compared the mean of the first four sessions for each participant (the baseline) to the mean of the last four sessions (when the robot's adaptation came into play). This allowed me to see whether adaptation affected children's language. To this end, multiple aspects of language were analyzed. First, a one-way anova revealed that the *total words* spoken by each group, averaged over sessions 1 through 4 were significantly different, $F(2,65)=3.15$, $p=.049$, eta-squared=.088. Post-hoc comparisons with the Bonferroni correction indicated that the easy-low group spoke fewer words total (*M*=82.6, *SD*=116.0) than the hard-high group (*M*=166.2, *SD*=76.0), but neither of those groups was significantly different from the easy-high group (*M*=119.4, *SD*=120.4). This is as expected: children with higher language ability generally used more words than children with lower language ability. However, the fact that the easy-high group did not use as many words as the hard-high group indicates that these two groups may not have been entirely equal at the outset of the study. Indeed, all the children in the hard-high group were especially loquacious, while two of the children in the high-easy group seemed more shy and only told silent stories (see Section 4.3.6). This is reflected in the large standard deviations for the groups.

During sessions 5 through 8, again, there were significant differences in the *total words* spoken, $F(2,59)=3.28$, $p=.045$, eta-squared=0.10. Post-hoc comparisons indicated the difference was between the hard high group (*M*=169.0, *SD*=80.0) and the easy-high group (*M*=84.7, *SD*=91.7), but not between those and the easy-low group (*M*=113.1, *SD*=111.2). This result suggests that after the robot adapted, children in the adaptive condition maintained or increased their level of talking, but children in the non-adaptive condition in fact talked less. Figure 16 shows the mean number of *total words* for each group across sessions.

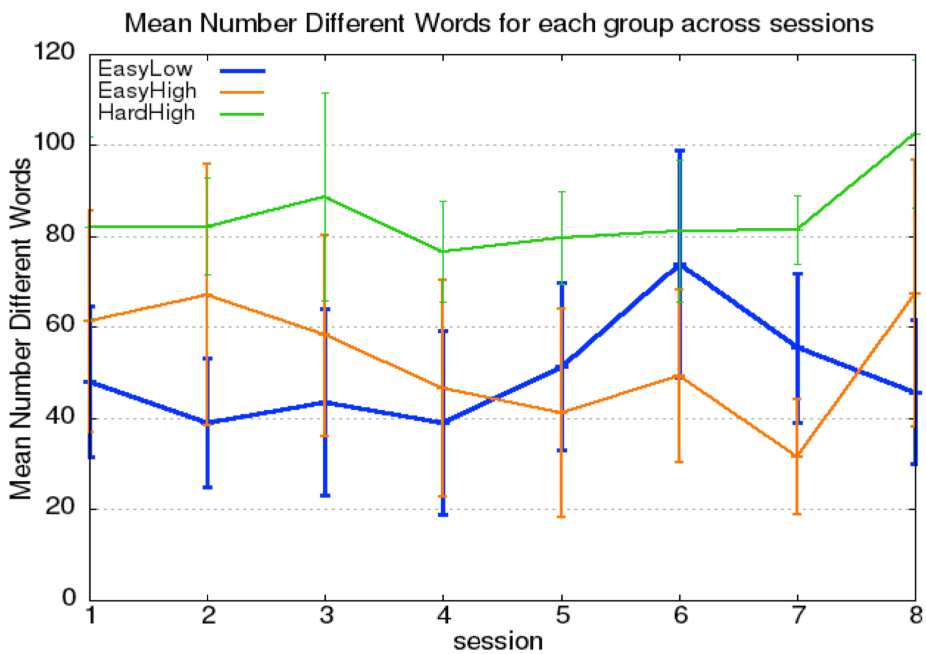

Figure 17: This graph shows the mean number of different words used each session, for each group, +/- standard error. During sessions 1-4, the hard-high group used significantly more words than the easy-low group. During sessions 5-8, after adaptation, the hard-high and easy-high groups diverged, with the hard-high group using significantly more different words than both the other groups.

A similar pattern was revealed when comparing the total number of *different words* during sessions 1-4 versus sessions 5-8. During sessions 1-4, there were significant differences between groups, $F(2,65)=4.02$, $p=.022$, eta-squared=.110. Post-hoc tests indicated that the easy-low (*M*=42.5, *SD*=48.7) and hard-high (*M*=82.4, *SD*=30.7) were significantly different from each other, but that neither was significantly different from the easy-high group (*M*=58.4, *SD*=51.5). During sessions 5-8, a different significant pattern emerged, *F*(2,59)=4.09, *p*=0.022, eta-squared=.121. The hard-high group (*M*=86.3, *SD*=25.0) used significantly more different words than both the easy-low (*M*=57.1, *SD*=48.9) and easy-high (*M*=46.4, *SD*=43.6) groups. This suggests that children of high language ability will maintain the diversity of words used with a robot that adapts to tell harder stories, but not if the robot continues as a “younger” peer. Figure 17 shows the mean number of *different words* used by each group, across the eight sessions.

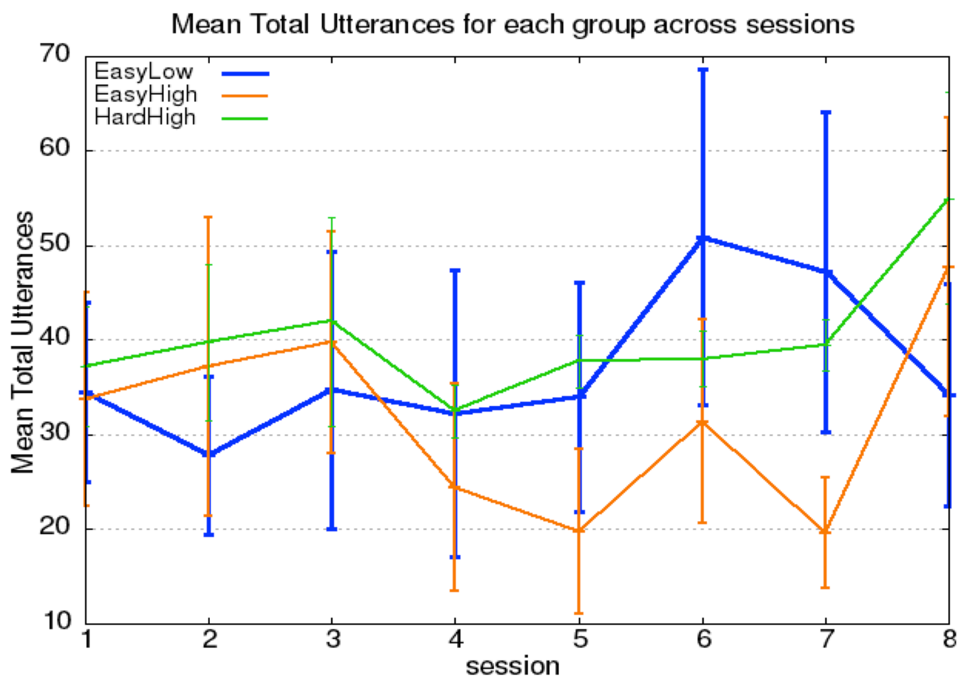

Figure 18: This graph shows the mean Total Utterances for each group across sessions, +/- standard error. The easy-high and hard-high groups start out speaking about the same number of utterances during sessions 1-4, but later, the two groups diverge. This suggests that the robot's adaptation is helpful in encouraging speech.

Related to this result, I examined several of the structural indices of language associated with literate language, including adverbs, elaborated noun phrases (containing, e.g., determiners and adjectives), and coordinators/conjunctions (Curenton & Justice, 2004). During sessions 1-4, there were no significant differences in the number of adverbs used by each group. However, during sessions 5-8, a one-way anova revealed significant differences, *F*(2,65)=4.03, *p*=.022, eta-squared=.120. Post-hoc tests with the Bonferroni correction indicated that the hard-high group (*M*=10.9, *SD*=7.10) used significantly more adverbs than the easy-high group (*M*=4.26, *SD*=5.76). This is due to a decrease in the adverbs used by the easy-high children (during sessions 1-4, *M*=8.05, *SD*=8.33). During sessions 1-4, there was a significant difference in the number of determiners used by group, *F*(2,65)=, *p*=.003, eta-squared=.164, and in the number of coordinators used by group, *F*(2,65)=5.23, *p*=.007, eta-squared=.139. Post-hoc tests indicated that the easy-low group used significantly fewer determiners (*M*=8.19, *SD*=12.9) and coordinators (*M*=4.66, *SD*=8.44) than the hard-high group (determiners *M*=23.4, *SD*=11.7, coordinators *M*=12.3, *SD*=7.05). This difference, however, was not present during sessions 5-8. Children in the easy-low group increased the number of determiners they used over time. A significant difference was seen during sessions 5-8, however, $F(2,59)=3.93$, $p=.025$, eta-squared=.118. Post-hoc tests revealed that the high-hard group used more coordinators than the high-easy group, due a decrease in the number of coordinators the high-easy group used.

When comparing the difference in *total utterances* from session 1 to session 8 for each group to see whether there was an overall amount of learning, an independent-samples t-test revealed that the hard-high group had significantly more *total utterances* during session 8 than in session 1, $t(30)=2.80$, $p=.008$. The other groups did not have significant differences. Figure 18 shows the mean *total utterances* for each group across sessions. This result suggests that children with high language ability talked more after playing with the robot that told harder stories.

Many more analyses could be performed. For example, the pattern of changes across all sessions could be examined. More measures than were discussed here could be analyzed, such as the use of different kinds of words (e.g., adjectives, nouns), or a more detailed accounting of sentence structure.

4.3.6 Stories

In total, given the number of children, sessions, stories per session, and missed sessions, 262 stories could have been collected. However, three children told no stories. Five children primarily told silent stories, in which they spent time dragging characters on the tablet and sometimes murmuring to themselves, but not speaking aloud. Nine children told stories aloud every session. Of those, 141 (53.8% of the total possible) were twenty words or longer. The mean word count of these 141 stories was 93.8 (*SD*=55.1, *max*=269, *min*=23). The mean Flesch-Kincaid Grade level across the stories was 1.1 (*SD*=0.95, *min*=0.0, *max*=3.7), which reflects the use of shorter sentences and simpler words with fewer syllables. The means for the Coh-Metrix measures were as follows: deep cohesion *M*=63.5, *SD*=37.9; narrativity *M*=83.4, *SD*=16.6; referential cohesion *M*=50.6, *SD*=26.5; syntactic simplicity *M*=92.1, *SD*=13.1; and word concreteness *M*=68.3, *SD*=31.6. These means are all comparable to the values obtained during the pilot study (Section 3.3.2.g). The narrativity and syntactic simplicity scores in particular point to highly narrative stories, with simple sentence structures and easier words.

Examining children's stories averaged across all sessions by group, there was a significant difference in the FKGL between the groups, $F(2,77)=7.85$, $p<.001$, eta-squared=.169. Post-hoc comparisons with the Bonferroni correction indicated that the easy-low group (*M*=0.59, *SD*=0.66) had a lower FKGL than the hard-high group (*M*=1.36, *SD*=.81). The easy-high group was not significantly different (*M*=1.08, *SD*=0.79). The mean FKGL for each group across sessions is plotted in Figure 19. No other significant differences in children's stories across all sessions by group were found. This may be in part because, as noted earlier (Section 3.3.2.g) Coh-Metrix may not give accurate or informative results when analyzing small samples of text, and children's stories were generally quite short.

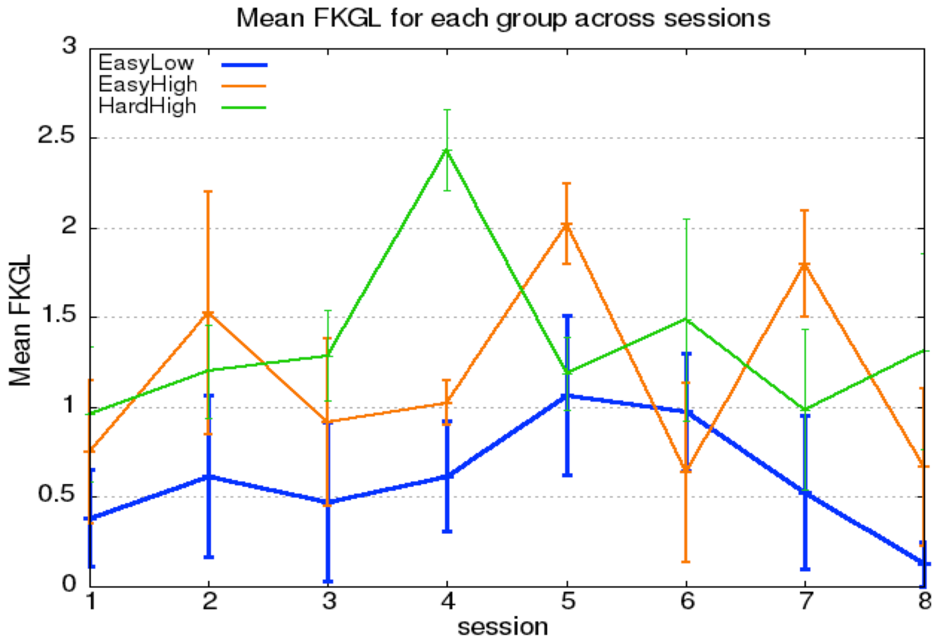

Figure 19: The mean Flesch-Kincaid Grade Level (FKGL) for each group, across all sessions, +/- standard error. When averaged across sessions, the hard-high group had significantly higher FKGL scores than the easy-low group. The easy-high group did not significantly differ, though as can be seen here, it is often higher as well.

A great deal more analyses could be performed. Foremost would be a count of the total time spent telling stories, whether silent or aloud. In addition, examining similarities in the content between the robot's and children's stories would be worthwhile. Qualitative assessment suggested that some children borrowed elements from the robot's stories, using names, words, and themes that the robot used. In addition, the overall pattern across sessions could be examined, as well as changes from the first to last session. It may be that aspects of the robot's stories (e.g., deep cohesion scores, word count, content, characters, theme) influenced various attributes of the stories children told. Finally, the occurrence of literate language, such as decontextualized language, spatial expressions, temporal expressions, and quoted speech could be coded in children's stories and totaled, as was done by Ryokai et al. (2003), since these measures may differ across groups as well.

Qualitatively, children covered a range of themes in their stories. Sometimes, children borrowed elements from the robot's stories, such as names for characters, activities that the characters performed, or patterns of speech. For example, in one of the robot's stories about the playground scene, one of the characters was named Miach. One child continued to use this name:

> *One time there were three friends, XX, Micah and Isabella. Micah liked going on the swings. Isabella liked going on the slide. One time they made a new friend, Daisy. She liked ball. One time she hid behind a bush until nobody saw her. Then both of the kids that were playing, approached and hid. Then, Micah slid down the slide and saw her. She stepped out but landed on the top of the brick tower. So then, they both came down together. The end.*

Several children retold versions of the robot's stories. For example, after the robot told a story in the forest scene about three animals that played hide-and-seek together, one child told the following story:

> *Once upon a time there was a squirrel named, Squirrel, a turtle named Turtle and a rabbit named Rabbit. That particular day They played hide and seek. Squirrel hid in the mud. Turtle hid in the trees while bunny counted. "One, two, three, four". "Found you"! "Found you, Turtle". "My turn". XX behind a tree. Squirrel found Turtle. and then they played again and again. The end.*

After the robot told a story about a dragon that just wanted to play with the bird and butterfly in the meadow scene (not eat them), one child told this story:

> *I think XX was XX about butterfly. Once there was a butterfly and a bird. Once there was a dragon and they loved to play. One day a dragon came by. "Run", said the bird. "Run", said the dragon. He flew higher and higher. But the dragon breathed fire. He almost caught him. He climbed higher and higher and higher and higher and higher. He almost went to the top. But the dragon twisted around and around and around. "I want a friend", said the dragon. "Really"? "You're not gonna eat me"? "Yes", said the dragon. So they played and played and played. The end.*

Another child copied the way the robot always started its stories. The robot would say, “I'm going to tell a story about ____” and then start the story. This child told the following story about the playground scene. This story also incorporated elements of conflict between the kids on the playground, similar to the robot's story:

> *I'll tell a story about a playground and some kids. One day some kids were swinging on the swings. They were going back and forth and back and forth. Their friend was playing on the slide. And then he found a ball. He started to throw it up and down, up and down, and up and down. And then, his friends came and said, "We want to play". And then she, "I'm playing by myself". "No thanks". And then they got mad. And they took the ball. But then the other girl found another ball and worried that the other girl might find it. So she took it away, too. And then, she said, "Bring those balls". And then the other one said, "Bring those balls back". And they went, "OK". And they threw them down. Wait. Oh come on. It's not moving. Wait it's OK. The end.*

Other children made up their own stories about the scenes on the tablet. For example:

> *(Dinosaur scene)*
> *Once upon a time there was a dinosaur named Lollipop and a dinosaur named Candycane and a dinosaur named Cookie. This one will be Cookie. Candycane. This one is named Lollipop. This one is named Candycane. And they all lived in the pool. The end.*

(Castle scene)
Once upon a time there was a prince who had a pet crocodile. He was friends with the princess and liked to give her rides on the crocodile. So they went for a ride every day she visited. One day, the Croc said, "I am tired". "I wanna run away". So he ran away. " But we can't have any more playdates". "That's sad". "Let's go talk to the croc". "La la la la". So they searched and searched for days and weeks. And one day the prince found him and said, "Over here". And then he said, "Come back here". So they rode on the croc all the way back to the castle. The end.

(House scene)
Once upon a time a little girl was XX all day long. And she had a puppy. And she had a cat. the puppy chased the cat all the time. And she had a ball all the time. And she had boxes. The boxes stayed where they were all the time. One day the dog hid under a chair. He was hoping to pounce up and get the cat. for the first time she climbed up on the couch. And he tripped on a ball. then the cat jumped out of the window and was never seen again. The end. So the dog XX because his XX. And he jumped out of the window. so he was never seen again. And the girl XX rest of her life like she promised herself. Your turn? Mhm.

One child, however, often ignored the tablet entirely when she made up her own stories. For example:

Once upon a time there was two XX, a brother and a sister. They went to XX picking flowers. But when they pick flowers sometimes they go up to the flower monsters house. And in the night he likes to sneak and grab other peoples flowers. but they want magic, the monster and the boy. And the girl does not have the magic. And the only way to get the magic is to get a flower that has seven colors on the petals. and the monster was not nice. And if he got the flower that has seven colors he would do all the bad things that he can do. And then one day they got the flower. But then the monster saw that. and then they a flower that had seven colors. But it was actually a poisonous one for the monster because the monster was bad. but except for eating it he got the poisonous power to get XX. so they ran as fast as they can. But the monster had such slow feet that he could only walk because his legs weren't feeling well. So he could only walk. And then his legs started breaking apart. And then he died but it was actually good because he's actually bad. The end.

Sometimes, she did incorporate elements of the robot's story into her own, even though she did not use the tablet. For example, after the robot told a story in the Mars scene about an alien family who celebrated a birthday party, she told the following story:

Once upon a time there was a family of astronauts. And they loved to play in parks. But one park was their best park that they loved to play in. But there was a bad guy that lived near it. And it liked to bake aliens and then eat them. So they figure outed a way to do that. But they could not. But then the baby got a solution to that. And he said, "Maybe we could find another playground that I would like to play in, so we don't have to play in that one". So *they checked at some playgrounds. And then they found one. And then they started playing in it. The end.*

4.3.7 Video & Behavior

Video analysis and coding is a time-consuming endeavor, and as such, is still ongoing. To initiate the video analysis, I held a two-hour workshop in which a group of colleagues and I observed a small selection of clips, about 30min in total, taken from the videos of children's sessions (over 1200 minutes in total). We discussed the behaviors children engaged in, compiling a list of questions that we wanted to answer through analysis of the videos, as well as a list of behaviors that may provide clues into how the children construed the robot, which is provided at the end of this section.

The primary questions of interest pertained to how children related to the robot, and how the children construed the robot. What role did the robot take for the children? Was the robot viewed as a trustworthy resource, a teacher, a friend or peer? Did the role the robot take change over time? How did the children build a relation with the robot over time? How did they treat the robot? How consistent were children's actions, and under what conditions would they switch from treating the robot one way to treating it another way?

First, we could look at children's initial encounters with the robot. How can the robot engage children initially? Did children initiate conversation with the robot on entry, or did they wait until the robot spoke first? Did this behavior change over time?

During the first session, one child moved the table at the start of the robot's story – perhaps in awe of the things on the screen moving on their own. How did children respond to the robot moving objects on the tablet screen without touching them? Were any confused? In a recent, unpublished study conducted by other members of my lab group, in which children played a game on a tablet with a similar robot, some children found it confusing when the robot had arms that did not move, and were also confused as to how the robot could affect the tablet without touching it. In this study, however, children did not vocally express confusion. Were there other indicators (e.g., facial expressions) that indicated confusion? Did they think of the tablet as a movie that the robot narrated? Did they simply accept that the robot has some "magic" powers, by which it could remotely move objects on a screen? Furthermore, the characters on the tablet may have acted like props (see Bretherton, 1984 on children's pretend play with props). Do children treat these props differently? For example, they may treat a prop as an extension of their own body (e.g., when a child picks up an object and says, "This is an airplane, bsshhh!" while waving it around in the air), or they may choreograph and narrate the actions of characters.

Figure 20: One child moved the robot and table so she could sit side-by-side with the robot. This photo is from her last session with the robot.

One child moved the robot and table so that she could sit side-by-side with the robot, rather than across the play table from it. There may be some related work on body language and proxemics that could bring insight into how this child perceived the robot: Sitting at a ninety-degree angle or side-by-side has been seen as psychologically closer than sitting across the table (Haase & DiMattia, 1970; Lott & Sommer, 1967; Sommer, 1959). The robot may be viewed as an equal, someone close. Over the sessions, the child moved from sitting across the robot to sitting side-by-side with the robot, which could indicate a change in how she construed the robot and in how comfortable she felt with the robot. There was a second child who also turned the table around on the robot's turn, so that the pictures on the table would not be upside-down for the robot. This may be indicative of thinking of the robot as an other, like oneself, who might want to see the pictures right-side-up.

One child repeatedly misunderstood the robot's final question each day ("Will you play with me again later?"), answering, "No, I'm not coming back!" She would go on to justify her response by saying, "I have to go to class" and "I have to go home." She would insist, "Green, I'm not coming back!" The robot, in turn, would respond by saying, "Aw, maybe we will anyway," or making a sad facial expression. However, either of these responses led the child to continue insisting that she would not return. Perhaps she was expecting a more emotional response from the robot. This suggests that children's *departure behaviors*, during the final conversation and good-bye phase of the interaction, may be worthwhile to analyze further. In addition, this clip suggests that it may be interesting to see whether any other children are looking for a more emotional response from the robot. When children were telling their own stories, the robot tended to respond with small backchanneling behaviors – "Mmhm!", "Ooh!", "Oh no!" – but not with big responses or hugely emotional responses. Children's interview answers regarding whether the robot could be happy (100% said it could) or sad (58.8% said it could) also give insight on this: the robot's emotional range could be extended, as it was a predominantly cheerful, happy robot. This child also called the robot by name – did others use the robot's name?

During all the children's interactions with the robot, the experimenter was present in the room. It would be worthwhile, when analyzing the video, to keep the broader context of the *room* in mind – the robot, but also the experimenter, the camera, and so forth. Indeed, during the first session, children tended to look back at the experimenter for confirmation or approval more often than in later sessions. In later sessions, they tended to look back less, only involving the experimenter in the robot interaction when something about the robot or game was not working as expected (e.g., if the pictures on the tablet would not move).

Finally, the voice the robot had was praised as being very good (it was recorded by a human), very child-like and friendly, with the right intonation and not “robotic” at all. How was this voice perceived by children? Is it possible for a voice to be “too good,” in that the “perfection” of the voice took away something from the robot being a robot? Could it be that the robot was changed into a character that was more like a puppet than like a separate entity? How does the voice affect construal of the robot more generally? It could be that children associated the robot more with a puppet character, playing pretend, rather than like interacting with another child. Furthermore, it may be worthwhile to look at what voices children use in their stories – their own? Made up voices? Do they change voices as they change which characters are speaking? For example, Bers et al. (1998) had children who were in the hospital play with a puppet. The children liked different voices for the puppet in relation to their own condition – more or less like their own voice, depending on whether they wanted to distance themselves from what the puppet was feeling.

Behaviors to code

- gaze (e.g., looking at robot, tablet, experimenter, or away)
- answering questions or responding to robot's speech
- initiating speech with robot
- listening to story
- looking at tablet
- touching tablet
- touching robot (e.g., hugging, touching its feet)
- location of the child in relation to the robot
  - did child move the robot or turn the table around?
  - were they seated at a table or on the floor?
- child's story content:
  - same or different names for characters than those the robot used
  - same or different story themes than the robot's
  - same words that robot used
- age, gender, preferences (e.g., one child especially liked princess stories)
- initial greeting behavior (e.g., initiating conversation, waiting for robot)
- departure behavior (e.g., hugging, talking, just going)

- welcome behavior, does it change over sessions.
- what the child was doing when the robot said the target words
- addressing the robot by name

Figure 21: A child creates a story on the tablet. Here, the robot gives nonverbal feedback in the form of a smile.

4.4 Discussion

4.4.1 Language and learning

The main question examined in this work was whether playing with a robot that adapted to speak as a slightly older peer relative to a child could improve (i) the number of new words children learned from the robot's stories, and (ii) the length and complexity of the stories children told. The results presented above support these hypotheses: children in the adaptive conditions maintained or increased the amount and diversity of the language they used during interactions with the robot. While children in all conditions learned new vocabulary words, created new stories during the game, and enjoyed playing with the robot, children who played with a robot that adapted to their language ability improved more than children who played with a robot that did not adapt.

This result is supported in particular by one pattern that emerged from the analysis of children's overall language: a divergence between the easy-high and hard-high groups after the robot's adaptation. For multiple measures, these two groups are initially not different – which is as expected, since both groups consisted of high language ability children. However, after the robot adapted, "leveling up" with the child, the hard-high group generally exhibited more total speech and more diversity of speech. In fact, in some cases, when the robot did not adapt, children actually talked *less*. For example, children in the high-hard group used more adverbs, determiners, and coordinators than children in the high-easy group during sessions 5-8.

Similarly, I saw a convergence between the easy-low and hard-high groups. These groups initially started out different – as expected, since the groups differed in their language ability. However, as time went on, children in both these groups talked more and used more diverse words than children in the non-adaptive group.

Why did the adaptive robot have this effect? Research from psychology and education fields has argued that a learner will learn best when in the zone of proximal development (Vygotsky, 1978), when enough but not too much challenge is presented. This is also characterized by the theory of flow, in which the optimal state of learning is achieved when challenge or difficulty and ability are balanced – too much challenge, the learner becomes frustrated; too little, and the learner grows bored (Csikszentmihalyi, 1991). When the robot "leveled up," it was increasing the challenge. This may explain why children in the adaptive condition maintained or increased the sophistication of their language. But it does not explain why children who played with the non-adaptive robot gradually *decreased* the amount and diversity of words used. Perhaps this drop-off was seen because the children were no longer engaged – they grew bored; the game did not present sufficient challenge. However, the children's responses during the final interview indicated that this was ostensibly not the case. They still reported that they liked playing with Green, that Green's stories were interesting and easy to understand, and that they would like to play the story game with Green again sometime.

One possible explanation comes from one of my original hypotheses. I had expected that children of high language ability who played with the non-adaptive robot (i.e., children in the high-easy group) might exhibit teaching or mentoring behavior, as children might with a younger sibling or peer. While the transcripts of children's interactions have not yet been coded for this kind of behavior, the decrease in the amount and diversity of words used may be indicative of children using simpler speech to accommodate their "younger" robot friend's level of ability. Children adapt their language as a function of who they're talking to. In a classic study by Shatz and Gelman (1973), four-year-old children told adults, peers, and two-year-olds about a toy. When speaking to a younger child, the children adjusted their language to use shorter, simpler utterances than when they spoke to a peer or an adult. If children are also doing this with the robot, then it may be that they perceive the robot as younger than themselves, rather than as a peer.

Another finding was that children learned multiple kinds of words through the robot's stories – nouns, adjectives, verbs. Most prior work with robotic learning companions has focused on teaching just one aspect of language, such as nouns (e.g., Freed, 2012; Movellan et al., 2009; Movellan et al., 2014) or verbs (e.g., Tanaka & Matsuzoe, 2012). But here, I found that children readily picked up any kind of new words during the stories. This is similar to how children learn words in everyday life: by hearing the words used in context. A few words were harder for all the children to learn than others – primarily *spruce* and *approach* – but the remainder were learned by at least some of the children.

The results reflect the richness of the interaction setting between the robot and child. The setting was not just a computer teaching a child to tell better stories – it was a subtle, relational interaction, in which both the robot and the child were participants. The tablet, containing the scenes and characters for their stories, was the mediator. It created a context in which stories could be constructed. The tablet was critical. It was a starting place for children's stories. The children could build on stories the robot told, constructing their own, new narratives. When the tablet failed – e.g., if the story characters did not move or were hard to drag – then some children ended their stories earlier. The tablet "breaking" also broke children out of the play-world. But when children were immersed in the play-world, they told varied, rich stories with the robot.

One concern is that the children who participated in this study were more intelligent, on average, than the typical preschooler, which may influence how generalizable these results are. The children attended good preschools; they knew, on average, half the target vocabulary word at the study's outset, and had high scores on the PLS-5. However, their mean MLU and type-token ratios were approximately the same as other children their age (Brown, 1973; Rice et al., 2010; Watkins et al., 1995; Hess et al., 1986), suggesting that they were not developmentally ahead of other children. This suggests that the results may generalize beyond these two preschools – young children will happily create stories together with a robot. The language the children use will be influenced by the robot's language level. Children who play with a peer robot that speaks at or above their level will maintain or increase the level of their own language, while children who play with a "younger" robot may "talk down" to it. All children, regardless, could learn new words from the robot through the game.

Many open questions remain regarding what factors could help children's language improve more each session. For example, perhaps it would help if the robot leveled up a little every session, advancing the complexity of its stories gradually, rather than switching levels just once. Or perhaps the robot do more to adapt to an individual child's needs – such as adjusting not only the story's complexity, but also which new words are introduced and how many new words are introduced based on the child's present knowledge. More suggestions are presented in the Future Work section (5.2).

4.4.2 Long-term interactions with robots

The second goal of this work was to investigate long-term interactions between children and robots. Very few studies have been performed in which children played with a robot multiple times over weeks or months. While prior work has shown that within just *one* session with a robot, children will engage in nuanced social interactions with the robot and enjoy playing with the robot very much (e.g., Kahn et al., 2012; Breazeal et al., in review), little work has investigated how children's engagement in, construal of, and interactions with a robot will change over *repeated* encounters as the novelty of the robot wears off. What kind of relationship ensues, if any? And, beyond how the child *relates* to the robot, how is the robot construed by the child? The results here suggest that not only will children continue to enjoy playing games with a robot, they may also develop substantial and meaningful relationships with robots, engaging them as social and moral others. Not only did children indicate that they liked playing the storytelling game and would like to play it again sometime, they were also quick to name Green their friend, and they talked about Green the way they talked about their other friends. During the last session, two children asked whether the robot would be attending the preschool's end-of-school picnic, the same way they would inquire whether a friend would be there. Another child began sitting beside the robot rather than across from it, indicating greater psychological closeness and comfort over repeated sessions, and would give the robot hugs before she left each day.

However, this is definitely not to say that all children will develop meaningful relationships with all robots. Several factors are at play. The first factor is the robot's speech and behavior. Prior research investigating long-term interactions with robots has suggested that *variation and adaptation* in the robot's speech and behavior are necessary to maintain engagement over time and to build a long-term relationship (e.g., Bickmore et al., 2010; Kidd & Breazeal, 2008; Lee et al., 2012). The robot has to change through time. The robot has to have its own story that it can share. The fact that children still enjoyed playing with the robot during session 8 is likely related to the fact that the robot's dialogue and stories varied each session. There was something new to look forward to, to discover. Notably, one child who adamantly refused to tell his own stories was still very eager to participate in the study. He said he always liked to hear Green's new stories. Another child who specifically said she would rather play a *different* game with Green still liked to come play and tell stories anyway, because Green was her friend and that was the game Green played.

With regards to the robot's dialogue, it is worthwhile to recognize how well the robot managed to communicate and to maintain engagement, given the scripted interactions available. There were a number of moments when it was clear that the robot ought to have had different speech options available and more flexibility in its responses. For example, sometimes children requested the robot to tell a story first, or would ask if they could go first. The robot had a script in which it would ask either "Will you tell a story first today?" or claim the first turn by announcing, "I'll go first today!" As such, the robot did not respond to children's requests (though sometimes, it *seemed* like the robot did, if the script for the session happened to be one that aligned with the child's desires). The robot could also have done more to encourage children to talk outside of the story game, during the introduction and closing. The robot would share a new fact or story about its "day," but often did not sufficiently prompt or invite the child to respond in kind. Finally, children often say unexpected things which cannot be anticipated ahead of time, to which the robot could only respond generally. All these limitations could be partially overcome through revision of the scripts used, based on the new knowledge acquired through this study of children's probable speech and behavior. Autonomous generation of varied dialogue – though a very hard problem – could also help.

Another factor is children's own speech and behavior. Some children were much more open to the new experience of playing with a robot than others. One child, for example, was very quiet and barely said a word to either the robot or the experimenter. It was unclear whether the robot actually made her uncomfortable, or whether she was just shy. This behavior highlights how, as when children make friends with some children but not others, children may more readily make friends with some robots, but not others. The personalities and behavior of both the child and the robot will influence the relationship that develops.

The third factor at play in the relationships children developed with the robot was how children actually construed the robot. Their perceptions of the robot as a character and social other came partly from the robot's personality, as revealed through its dialogue and stories. Their perceptions came partly from the robot's voice and appearance. They also came from the way the experimenter interacted with the robot, and the language used by the experimenter to talk to or about the robot. And they came from the language children themselves used with the robot. As Susan Engel (1995) put it, "language constructs the experience it describes rather than merely reflecting that experience." (For more detail on the role language plays in constructing our relations with robots, see, e.g., Coeckelberg, 2011a, 2011b, 2011c).

Understanding children's construal of this robot (let alone robots in general) is a difficult task. It is apparent that children considered Green to be a social other: they responded to Green with social behaviors and attributed social attributes to Green. All the children could understand the intentions or feelings the robot displayed, insofar as they stated that yes, “Green likes the story game.” And, although conceivably the children could be just *pretending* that Green is a social other, evidence presented by a growing number of researchers suggests this is not the case – children actually do believe the robots are social entities (Clark, 2008; Kahn et al., 2012; Severson & Carlson, 2010). But to what extent was the robot considered a moral agent? Alive, or a machine, as having biological, psychological, or mental capacities, and so forth? And how do these various qualities affect how much a child, for example, trusts the robot as a resource for exploring relational issues, as an informant for learning, or simply as a conversationalist?

A few researchers have delved into these questions. Robots fall in a strange category, straddling the space between alive, animate beings and inanimate objects. Children sometimes describe robots as being “in between” living and non-living (Kahn et al., 2011; Kahn et al., 2012; Severson & Carlson, 2010). Some work even suggests that a person's conception of a robot as more alive and social versus more like a machine can be manipulated just by introducing the robot in a more (or less) social way (Kory & Kleinberger, 2014).

In one study, Jipson and Gelman (2007) showed a series of short videos and photographs of objects (including a robotic dog) to children ages 3, 4, and 5, as well as to adults. The objects varied in whether or not they were alive, had faces, or showed autonomous behavior. Participants answered a series of questions pertaining to each object's biological, psychological, perceptual, and artifact-related properties. They found that children rarely attributed biological properties or aliveness to the robot dog, but did still attribute psychological and perceptual abilities to it. I asked these same questions of children during the final interview. My results were similar: the majority of children said Green could think, be happy, see, and feel. A minority said Green could eat or grow. However, approximately half said Green was very alive or kind of alive, and the number of children who attributed biological, psychological, and perceptual properties to the robot was generally higher than the number of 5-year-old children who did so in Jipson and Gelman's (2007) study. This could be for several reasons. First, Green spoke and displayed facial expressions, which are both behaviors associated with humans, rather than animals or machines. Children's understanding of the kinds of agents that can speak and smile may have influenced what other properties they would expect such agents to have. Green was also an odd, fantastical dragon character, rather than a mechanical version of a familiar animal (a robot dog). This nature as an already-strange character may have let children believe that the robot could have other, out-of-the-ordinary abilities, while in the case of the robot dog, they could understand that the robot dog was not a *real* dog. In addition, children actually *interacted* with Green – they did not just watch a video.

In an unpublished study, I asked children whether they thought a DragonBot like Green was more like a person or more like an iPad both before and after interacting with the robot. I asked these questions again in the present study. The percentage of children in both studies who said the robot was more like a person was roughly equivalent, but only when comparing children's responses from *after* they interacted with the robot – not *before.* This suggests that studies in which children are asked about robots via just a picture or short video (such as Jipson & Gelman, 2007) may obtain a skewed view of children's perceptions of robots. Interaction can change one's perceptions. Children's construal of a robot may change across sessions, or even from before to after a session. This is important because robots *are for* interaction. Studying how children construe them outside of the context of interaction may say something about children's preconceptions of what robots are and what robots can do, but less about how children actually construe a robot with which they interact. We do not yet know how consistent children are in treating the robot as they do, or under what conditions they switch from treating the robot one way to treating it another way, or what aspects of context matter. Indeed, it is an empirical question whether the results found in these various studies regarding children's construal of robots will generalize to other situations and other robots. I expect that many of the results would not – children's construal of robots will be dependent on factors such as the robot's appearance, voice, and behavior; the framing or introduction of the robot; the language used by the child, the robot, and the experimenter; and even the task the child engages in with the robot. That said, I suspect there will be some commonalities in how children approach any kind of robot, as robots do share some of the same strange, animate-but-not-alive properties. Perhaps there will be a continuum filled by different robots, moving from the most animate and lifelike on one end to the more machine-like at the other.

Many questions still remain regarding children's construal of Green and how children built relationships with the robot over time. Some of these questions may be answered through further video analysis (described in Section 4.3.7). Many will require further research to be done.

5 Conclusion

5.1 Contributions

5.1.1 Technical & Design

The technical and design contributions of this thesis included, foremost, the iterative development of a storytelling game that the robot could play, with pilot testing with children at the Boston Museum of Science. This included creation of eight story scenes on the tablet, along with a repertoire of two different stories for each of those eight scenes, at both easy and hard levels, for the robot to tell to children. Following this, I created scripts for a long-term interaction with the robot, with each session containing varied content in the introduction, stories, and closing.

I also made multiple improvements to the teleoperater interface for controlling the robot, which reduce the operator's cognitive load and increase the repeatability of the robot's behavior. These improvements included (a) porting the teleoperator interface to run on a laptop computer from the original android tablet-based interface; (b) integrating the robot's look-at behaviors into the commands for triggering speech and mouth movements, so that the robot's gaze could be scripted and would stay consistent across participants; (c) automatically loading and advancing the robot's dialogue script, so that the teleoperator only had to select among the four current options (rather than navigate a sea of buttons for triggering speech). In addition, multiple perl scripts for automatically calculating statistics for the tablet logs and speech transcripts were written.

5.1.2 Psychology, education, & human-robot interaction

This work builds on a growing body of literature on robotic learning/teaching companions. I investigated two primary questions regarding (i) the potential of a robot companion to support children's long-term oral language development through play, and (ii) children's construal of and relationships with a robot during a long-term interaction.

I addressed the social, interactive nature of language learning through a storytelling game, mediated by a tablet, that the child and robot played together. The robot acted as a peer and a guide, not as a teacher, within a relational, dialogic context. The results showed that children will learn new words from the robot through the robot's stories. Playing with a robot that adapted the complexity of its stories to match the child's language ability improved children's language learning outcomes. This result supports prominent learning theories that argue that children will learn most when in the zone of proximal development, when the challenge presented matches their ability (e.g., Vygotsky, 1978).

The child-robot interaction can add rich variation to how children learn. This work demonstrates that a robot can maintain engagement across many sessions with a child, while leading an educational play activity. By showing that a robot can support the development of oral language, this work paves the way for future technologies that support not only oral language but also other precursors to literacy, such as joint pointing, scribbling, and understanding the performative power of words (Ackermann, 2002). This work also gives insight into the relationships children develop with robots over repeated encounters, children's construal of robots more generally, and what factors influence how children interact with a robot. Understanding how children construe robots, how robots influence children’s language use, and in what ways robots can support long-term oral language development will inform the design of future learning technologies.

5.2 Future work

5.2.1 Data analysis

Further data analysis is ongoing, as discussed earlier. More insights can likely be gained from delving deeper into the tablet data (Section 4.3.4), the language and story data (Sections 4.3.5 and 4.3.6), and the videos (Section 4.3.7).

5.2.2 Automation

At present, the robot used was not fully autonomous. Speech, animations, stories, and behavior are primarily scripted. Given the difficulties of natural language processing, a human will be in the loop to select among suggested actions and dialogue options, and trigger them at the appropriate times. As such, the system could be made more autonomous in several ways.

The first step is to automatically recognize children’s speech to create a transcript that can be analyzed. With this capability in place, children’s stories could be automatically analyzed for features such as key words and the complexity measures mentioned earlier. This could then allow dynamic adaptation by the robot to the child in real-time. The complexity of stories, specific keywords, or even the content of stories could be adapted to better match the child’s ability and interest. I have perl scripts already that will automatically calculate some relevant statistics from a text transcript. In addition, at present, only two levels (easy and hard) of each story were available. In the future, the robot could have 3-4 versions of each story available, at increasing levels of complexity, to adapt at a finer level to each child's ability. In addition, at present speech variations were scripted for the robot; they were not dynamically generated. Future work could include autonomous generation of varied dialogue. Finally, it may be worthwhile to explore other avenues of adaptation beyond language: perhaps the robot could adapt to factors such as the child's personality, communication style, learning style, and interests.

The robot's nonverbal behavior could also be automated. For example, one improvement pertains to the robot's gaze behavior. Several researchers have modeled human-like gaze behavior in communicative, storytelling settings (e.g., Cassell et al., 1998, Mutlu et al., 2006). Cassell et al. (1998) used parameters from coding a professional storyteller's gaze; their gaze model was based on turn-taking and on the structure of the information within one turn. However, children's gaze may be different during this storytelling game. The video data collected during pilot testing, as well as from the long-term study, could be coded to determine children's general gaze behavior for this game, such as how much time is spent generally spent looking down at the game and up at the child's face, so that the robot could glance down at the game and up at the child at appropriate intervals. Another gaze improvement could be to run a face detection algorithm on the video from the robot's camera feed, so that the robot's gaze could be directed to look at the child's face, rather than just looking up in the general direction of where the child usually sits. Other nonverbal behavior could also be automated: For example, children's emotions could be detected during the play session from modalities such as linguistic content, qualities of the voice, or facial expressions. With this information, a model could be developed for the robot to automatically react appropriately, e.g., with surprise, laughter, or support.

Data collected during this study could be used in the future to build a model of how the robot should act overall, based on how a skilled robot operator acted. Particularly interesting would be a model of turn-taking during conversation, in which when the robot should speak is predicted – that is, the timing and rhythm of the conversation. Additional data could be collected of children playing the storytelling game together or with teachers or parents – rather than with a robot – in order to learn how a robot should act if it wants to act in different roles, mimicking the behavior of another child, a teacher, or a parent.

5.2.3 Context & psychological

The context in which an interaction places, as well as the role the robots plays, could significantly impact how children view the robot and how engaged they are in the game. How the robot is introduced, for example, could influence how talkative and social a person is with the robot (Kory & Kleinberger, 2014). Children's own condition may lead them to prefer different voices for an interaction character during storytelling (Bers et al., 1998). As such, it may be worthwhile to explore context more fully. The form of the robot may matter, its shape, its appearance, or its movement. What if the robot takes on a different personality – an extrovert versus a shy introvert, for example? Would some children be more comfortable with a robot that has a particular personality? The robot's personality could also be developed further through its dialogue. For example, the robot could reveal a deeper character through the storytelling game, in the same way that children reveal more about themselves through their stories. Stories help construct the self – not just for ourselves, but to others (Engel, 1995). The robot might always include certain characters in its stories, use particular funny noises or expressions, have its stories follow a set of themes, or introduce other quirks that are dependent on its personality.

What happens if the robot's voice is artificial versus recorded by a person? Considerable effort went into recording each of the robot's possible utterances for the storytelling interaction. The ability to use a text-to-speech system or a synthesized voice would greatly increase the number of possible utterances, and would make it easier to expand the game. However, qualitative observation of several children who interacted both with (a) the robot that played the storytelling game with a human-like voice, and (b) a robot that had a synthetic voice (albeit playing a different game), showed that children may have a preference for the robot with the more alive, human-like voice. It is an empirical question whether the voice would create differences in learning outcomes, especially for children who are exposed to just one robot (and thus who do not have the opportunity to directly compare the different robot voices).

There was an open question of how children built a relationship with the robot over time. How does the robot's role influence the relationship? For example, the robot could play the role of a didactic teacher instead of a peer. Children's interactions with a robot-as-teacher may be very different from their interactions with a robot-as-friend/peer. Perhaps they would not like playing with the robot as much. There may even be differences if the robot plays the role of a younger peer, versus an older peer, prompting the child to take on more, or less, teaching/mentoring of the "younger" friend. A robot playing a different role would provide a different type of support or presence – how might this inform how children go about telling their stories?

Another avenue to explore is children's trust of the robot. It may be that child would view the robot as a more credible as a teacher on some subjects or in some circumstances more than others. For example, Breazeal et al. (in review) found that children trusted a robot that displayed more socially contingent behavior more as an informant than a robot displaying non-contingent behavior. Children monitor the reliability of informants and trust the information they provide accordingly (Harris, 2007; Koenig & Harris, 2007). Do children treat the robot as an informant during play? Would be the robot be viewed as more credible when telling stories about particular subject matter – such as aliens or strange creatures, about which the robot may have cause to be knowledgeable since it is a strange creature itself, rather than stories about human children or human activities?

Finally, this work focused on oral language, but the robot could also support other precursors to literacy, such as scribbling, joint pointing, the positioning of a children and fellow narrator side-by-side, the power of words, and more (Ackermann, 2002). For example, what if the robot was physically positioned differently in relation to the child? In the long-term study, one child repeatedly moved the robot to be situated beside her, rather than across the table from her. This positioning could be deliberately manipulated to determine whether seating the child and robot side-by-side changes how children perceive the robot as a friend and companion.

These are only a few suggestions regarding the directions of future study. Much more work is needed to attain our goal of understanding how robots, their language, their behavior, and their context influence the relationships children develop with them during long-term interactions.

6 References

Ackermann, E. (2002). Language Games, Digital Writing, Emerging Literacies: Enhancing kids' natural gifts as narrators and notators. In A. Dimitracopoulou (Ed.), *Proceedings of the 3rd Hellenic Conference on Information and Communication Technologies in Education (*pp. 31-38). University of Aegean, Rhodes, Greece.

Bainbridge, W. A., Hart, J. W., Kim, E. S., & Scassellati, B. (2011). The benefits of interactions with physically present robots over video-displayed agents. *International Journal of Social Robotics, 3*(1), 41-52.

Bers, M. U., Ackermann, E., Cassell, J., Donegan, B., Gonzalez-Heydrich, J., DeMaso, D. R., Strohecker, C., Lualdi, S., Bromley, D., & Karlin, J. (1998). Interactive storytelling environments: Coping with cardiac illness at boston's children's hospital. In C-M. Karat & A. Lund (Eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 603-610). New York, NY: ACM.

Bickmore, T., Schulman, D., & Yin, L. (2010). Maintaining engagement in long-term interventions with relational agents. *Applied Artificial Intelligence, 24*(6), 648-666.

Biemiller, A. (2010). *Words worth teaching: Closing the vocabulary gap* McGraw-Hill SRA.

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

Breazeal, C., Harris, P., DeSteno, D., Kory, J., Dickens, L., & Jeong, S. (in review). Young children treat robots as informants. *Topics in Cognitive Science*.

Bretherton, I. (1984). *Symbolic play: The development of social understanding*. Academic Press.

Brown, R. (1973). *A first language: The early stages*. London: George Allen & Unwin.

Cassell, J. (2004). Towards a model of technology and literacy development: Story listening systems. *Journal of Applied Developmental Psychology, 25*(1), 75-105.

Cassell, J., & Ryokai, K. (2001). Making space for voice: Technologies to support children's fantasy and storytelling. *Personal and Ubiquitous Computing, 5*(3), 169-190.

Cassell, J., Torres, O., & Prevost, S. (1998). Turn taking vs. discourse structure: How best to model multimodal conversation. Wilks (Ed.), *Machine Conversations*. Kluwer: the Hague.

Castellano, G., Leite, I., Pereira, A., Martinho, C., Paiva, A., & McOwan, P. W. (2013). Multimodal affect modeling and recognition for empathic robot companions. *International Journal of Humanoid Robotics, 10*(01).

Chang, C., Lee, J., Chao, P., Wang, C., & Chen, G. (2010). Exploring the possibility of using humanoid robots as instructional tools for teaching a second language in primary school. *Educational Technology & Society, 13*(2), 13-24.

Chartrand, T. L., & van Baaren, R. (2009). Human mimicry. *Advances in Experimental Social Psychology, 41*, 219-274.

Chen, G., & Wang, C. (2011). A survey on storytelling with robots. In M. Chang, W-Y. Hwang, M-P. Chen, & W. Müller (Eds.), *Proceedings of the 6th International Conference on E-learning and Games: Edutainment technologies: Educational games and virtual reality/augmented reality applications*. (pp. 450-456). Berlin Springer-Verlag.

Clark, H. H. (2008). Talking as if. In T. Fong, K. Dautenhahn, M. Scheutz, & Y. Demiris (Eds.), *Proceedings of the 3rd ACM/IEEE International Conference on Human–Robot Interaction* (pp. 393–394). New York, NY: Association for Computing Machinery.

Coeckelbergh, M. (2011a). Humans, animals, and robots: A phenomenological approach to human-robot relations. *International Journal of Social Robotics, 3(2),* 197-204.

Coeckelbergh, M. (2011b). You, robot: On the linguistic construction of artificial others. *AI & Society, 26*(1), 61-69.

Corriveau, K. H., Harris, P. L., Meins, E., Fernyhough, C., Arnott, B., Elliott, L., Liddle, B., Hearn, A., Vittorini, L., & De Rosnay, M. (2009). Young children's trust in their mother's claims: Longitudinal links with attachment security in infancy. *Child Development, 80*(3), 750-761.

Csikszentmihalyi, M. (1991). *Flow: The psychology of optimal experience* (Vol. 41). New York: Harper Perennial.

Curenton, S. M., Craig, M. J., & Flanigan, N. (2008). Use of decontextualized talk across story contexts: How oral storytelling and emergent reading can scaffold children's development. *Early education and development, 19(1),* 161-187.

Curenton, S. M., & Justice, L. M. (2004). African American and Caucasian preschoolers' use of decontextualized language: literate language features in oral narratives. *Language, Speech, and Hearing Services in Schools, 35(3)*, 240-253.

D'Mello, S., & Graesser, A. (2012). AutoTutor and affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems, 2*(4), 23.

Dunn, L. M., & Dunn, L. M. (2007). *Peabody picture vocabulary test, 4th edition*. Pearson Assessments.

Duranti, A., & Goodwin, C. (1992). *Rethinking context: Language as an interactive phenomenon*. Cambridge University Press.

Engel, S. (1995). *The stories children tell: Making sense of the narratives of childhood*. Macmillan.

Fish, M., & Pinkerman, B. (2003). Language skills in low-SES rural appalachian children: Normative development and individual differences, infancy to preschool. *Journal of Applied Developmental Psychology, 23*(5), 539-565.

Freed, N. A. (2012). *"This is the fluffy robot that only speaks french": Language use between preschoolers, their families, and a social robot while sharing virtual toys*. (Master's Thesis, Massachusetts Institute of Technology).

Gaiman, N. (2013, Oct 15). Neil Gaiman: Why our future depends on libraries, reading and daydreaming. *The Guardian*.

Golinkoff, R. M., Hirsh-Pasek, K. A., & Singer, D. G. (2006). Play= learning: A challenge for parents and educators. In D.G. Singer, R.M. Golinkoff, and K. Hirsh-Pasek (Eds.), *Play= Learning: How Play Motivates and Enhances Children's Cognitive and Social-Emotional Growth*, pp. 3-12. New York, NY: Oxford University Press.

Haase, R. F., & DiMattia, D. J. (1970). Proxemic behavior: Counselor, administrator, and client preference for seating arrangement in dyadic interaction. *Journal of Counseling Psychology, 17(4)*, 319.

Harris, P. L. (2007). Trust. *Developmental Science, 10*, 135-138.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young* *american children*. ERIC.

Hess, C. W., Sefton, K. M., & Landry, R. G. (1986). Sample size and type-token ratios for oral language of preschool children. *Journal of Speech, Language, and Hearing Research,* 29(1), 129-134.

Jipson, J. L., & Gelman, S. A. (2007). Robots and rodents: Children's inferences about living and nonliving kinds. *Child Development,* 78(6), 1675-1688.

Kahn, P. H., Kanda, T., Ishiguro, H., Freier, N. G., Severson, R. L., Gill, B. T., Ruckert, J., & Shen, S. (2012). "Robovie, you'll have to go into the closet now": Children's social and moral relationships with a humanoid robot. *Developmental Psychology,* 48(2), 303.

Kahn, P. H., Reichert, A. L., Gary, H. E., Kanda, T., Ishiguro, H., Shen, S., Ruckert, J., & Gill, B. (2011). The new ontological category hypothesis in human-robot interaction. In A. Billard, P. Kahn, J. A. Adams, & G. Trafton, *Proceedings of the 6th International Conference on Human-Robot Interaction* (pp. 159-160). ACM.

Kasap, Z., & Magnenat-Thalmann, N. (2012). Building long-term relationships with virtual and robotic characters: The role of remembering. *The Visual Computer, 28*(1), 87-97.

Kidd, C. D., & Breazeal, C. (2008). Robots at home: Understanding long-term human-robot interaction. In R. Chatila, J-P. Merlet, & C. Laugier (Eds.), *Proceedings of the 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems,* 3230-3235.

Koenig, M. A., & Harris, P. L. (2007). The basis of epistemic trust: Reliable testimony or reliable sources?.*Episteme, 4(3)*, 264-284.

Kory, J., & Breazeal, C. (2014). Storytelling with Robots: Learning Companions for Preschool Children's Language Development. In P. A. Vargas & R. Aylett (Eds.), *Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE: Washington, DC.

Kory, J., & Kleinberger, R. (2014). Social agent or machine? The framing of a robot affects people's interactions and expressivity. *2nd Workshop on Applications for Emotional Robots held in conjunction with the 9th ACM/IEEE International Conference on Human-Robot Interaction*.

Kory, J. M., Jeong, S., & Breazeal, C. L. (2013). Robotic learning companions for early language development. In J. Epps, F. Chen, S. Oviatt, & K. Mase (Eds.), *Proceedings of the 15th ACM on International Conference on Multimodal Interaction* (pp. 71-72). New York, NY: ACM.

Kuhl, P. K. (2007). Is speech learning 'gated'by the social brain? *Developmental Science, 10*(1), 110-120.

Lee, M. K., Forlizzi, J., Kiesler, S., Rybski, P., Antanitis, J., & Savetsila, S. (2012). Personalization in HRI: A longitudinal field experiment. In H. Yanco, A. Steinfeld, V. Evers, & O. C. Jenkins (Eds.), *Proceedings of the 7th ACM/IEEE International Conference on Human-Robot Interaction* (pp. 319-326). New York, NY: ACM.

Lott, D. F., & Sommer, R. (1967). Seating arrangements and status. *Journal of Personality and Social Psychology, 7(1)*, 90.

McNamara, D. S., Louwerse, M. M., Cai, Z. & Graesser, A. (2013). Coh-metrix version 3.0. Retrieved 10/23, 2013, Retrieved from http://cohmetrix.com

Meltzoff, A. N., Kuhl, P. K., Movellan, J., & Sejnowski, T. J. (2009). Foundations for a new science of learning. *Science, 325*(5938), 284-288.

Miller, J. F., & Chapman, R. S. (2012). Systematic Analysis of Language Transcripts (2012 Version) [Computer software]. Madison: University of Wisconsin—Madison, Waisman Center. Language Analysis Laboratory.

Movellan, J., Eckhardt, M., Virnes, M., & Rodriguez, A. (2009). Sociable robot improves toddler vocabulary skills. In F. Michaud, M. Scheutz, P. Hinds, & B. Scassellati (Eds.), *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction,* 307-308.

Movellan, J., Malmir, M., & Forester, D. (2014). HRI as a tool to monitor socio-emotional development in early childhood education. *2nd Workshop on Applications for Emotional Robots held in conjunction with the 9th ACM/IEEE International Conference on Human-Robot Interaction*.

Mutlu, B., Forlizzi, J., & Hodgins, J. (2006). A storytelling robot: Modeling and evaluation of human-like gaze behavior. In G. Sandini & A. Billard (Eds.), *Proceedings of the 6th IEEE-RAS International Conference on Humanoid Robots* (pp. 518-523).

Naigles, L. R., & Mayeux, L. (2001). Television as incidental language teacher. *Handbook of Children and the Media*, 135-152.

Nicolopoulou, A. (1993). Play, cognitive development, and the social world: Piaget, Vygotsky, and beyond. *Human Development*, *36*(1), 1-23.

Páez, M. M., Tabors, P. O., & López, L. M. (2007). Dual language and literacy development of spanish-speaking preschool children. *Journal of Applied Developmental Psychology, 28*(2), 85-102.

Rice, M. L., Smolik, F., Perpich, D., Thompson, T., Rytting, N., & Blossom, M. (2010). Mean length of utterance levels in 6-month intervals for children 3 to 9 years with and without language impairments. *Journal of Speech, Language, and Hearing Research, 53(2),* 333-349.

Ryokai, K., Vaucelle, C., & Cassell, J. (2003). Virtual peers as partners in storytelling and literacy learning. *Journal of Computer Assisted Learning, 19*(2), 195-208.

Sage, K. D., & Baldwin, D. (2010). Social gating and pedagogy: Mechanisms for learning and implications for robotics. *Neural Networks, 23*(8), 1091-1098.

Setapen, A. M. (2012). *Creating robotic characters for long-term interaction*. (Master's Thesis, Massachusetts Institute of Technology).

Severson, R. L., & Carlson, S. M. (2010). Behaving as or behaving as if? Children's conceptions of personified robots and the emergence of a new ontological category. *Neural Networks, 23(8),* 1099-1103.

Shatz, M., & Gelman, R. (1973). The development of communication skills: Modifications in the speech of young children as a function of listener. *Monographs of the society for research in child development, 38(5),* 1-38.

Siegler, R., & Crowley, K. (1991). The microgenetic method: A direct means for studying cognitive development. *American Psychologist, 46,* 606-620.

Snow, C. E., Porche, M. V., Tabors, P. O., & Harris, S. R. (2007). *Is literacy enough? pathways to academic success for adolescents*. ERIC.

Sommer, R. (1959). Studies in personal space. *Sociometry*, 247-260.

Tanaka, F., & Matsuzoe, S. (2012). Children teach a care-receiving robot to promote their learning: Field experiments in a classroom for vocabulary learning. *Journal of Human-Robot Interaction, 1*(1).

Valdesolo, P., & DeSteno, D. (2011). Synchrony and the social tuning of compassion. *Emotion-APA, 11*(2), 262.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Watkins, R. V., Kelly, D. J., Harbers, H. M., & Hollis, W. (1995). Measuring children's lexical diversity: Differentiating typical and impaired language learners. *Journal of Speech, Language, and Hearing Research, 38(6),* 1349-1355.

Wiltermuth, S. S., & Heath, C. (2009). Synchrony and cooperation. *Psychological Science, 20*(1), 1-5.

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2011). Preschool language scales, 5th edition. (PLS-5). *www.pearsonclinical.com*.