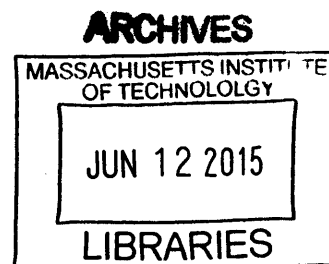


GlobalLit



A Platform for Collecting, Analyzing, and Reacting to Children's Usage Data on Tablet Computers

David S. Nuñez

B.A. Computer Science, Rice University, 2000

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

June 2015

© 2015 Massachusetts Institute of Technology. All rights reserved.

Author: Signature redacted

David S. Nuñez
Program in Media Arts and Sciences
School of Architecture and Planning
August 8, 2014

Certified By: Signature redacted

Cynthia Breazeal, Ph.D
Associate Professor of Media Arts and Sciences
Thesis Supervisor

Accepted By: Signature redacted

Patricia Maes, Ph.D
Professor of Media Arts and Sciences
Interim Academic Head, Program in Media Arts and Sciences

GlobalLit

A Platform for Collecting, Analyzing, and Reacting to Children's Usage Data on Tablet Computers

David S. Nuñez

B.A. Computer Science, Rice University, 2000

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 at the
Massachusetts Institute of Technology**

Abstract

Mobile devices, like tablet computers, are potentially potent tools for delivering language learning to children otherwise unreachable at any reasonable scale in a variety of home situations. The ubiquity and distribution of these devices make them ideal research tools, as well, and provide an opportunity to collect field data from children using tablets and ultimately about child-driven learning. This thesis describes a system, GlobalLit, through which researchers and educators can collect, analyze and react to data from various deployments around the world. The system enables geo-shifted data collection, normalized among heterogeneous deployment configurations around the world. Time-Shifted analysis allows us to understand the life-cycle of tablet use and exploration models in short-term engagements alongside long-term multi-year studies. Probes can be examined at a variety of scales (ex. Regions, Deployments, Tablet Groups (Conditions), Individuals). The system supports a Live mode in which data streams from the tablet in real time. The thesis presents background on field research, data collection frameworks, and the current state of tablets and applications applied to children's literacy and education. The thesis explains the design and architecture of the GlobalLit system, including the software infrastructure used to enable the platform. Two case studies illustrate example applications for the platform demonstrating observations made possible by GlobalLit on data sets from deployments around the world and showing how the GlobalLit system can be used to collect high fidelity data in a live, laboratory study. Through the presentation and discussion of the system and potential applications, this thesis attempts to show the potential value of systems like this for research and intervention for child literacy.

Thesis Supervisor: Cynthia Breazeal

Title: Associate Professor of Media Arts and Sciences

GlobalLit

A Platform for Collecting, Analyzing, and Reacting
to Children's Usage Data on Tablet Computers

David S. Nuñez

Thesis Reader: _____

Signature redacted

Tinsley Galyean, Ph.D

Research Affiliate, MIT Media Lab
Program Co-Director, The Dalai Lama Center for Ethics and Transformative Values

GlobalLit

A Platform for Collecting, Analyzing, and Reacting
to Children's Usage Data on Tablet Computers

David S. Nuñez

Signature redacted

Thesis Reader:

Robin Morris, Ph.D

Associate Provost for Strategic Initiatives and Innovation
Regents Professor of Psychology
Georgia State University

GlobalLit

A Platform for Collecting, Analyzing, and Reacting
to Children's Usage Data on Tablet Computers

David S. Nuñez

Thesis Reader:

Signature redacted

August 20, 2014

Maryanne Wolf, Ph.D

John DiBaggio Professor of Citizenship and Public Service
Director, Center for Reading and Language Research
Tufts University

Acknowledgements

Lauren, I am grateful for you, my loving partner. You encouraged me to take the chance to go back to school, and you supported and soothed me through the all-nighters with unwavering patience, care, and candy treats. Without you there is no way I would have started and certainly no way I could have finished, and for that I can only say, "thank you, and I love you." As far as I am concerned, we accomplished this together, and it is ours to share.

I thank you, **Dad**, for being my inspiring role model for my entire life, most especially how you recovered and got back up after you fell. You taught me to eat my sardines and work hard to overcome life's challenges.

Mom, know that I felt your prayers all the way from Texas, and I was always comforted to know that you would always love me no matter what. Thank you for being my wind.

My brothers, from you I learned to value creativity and to laugh until I cry and can't breathe with my small throat. This skill has made my life fun and filled with wonderful, ridiculous moments.

Cynthia, the couple of years I spent in your group following your vision and watching you think *big*, on all fronts, have truly been life changing. I will cherish the experience for the rest of my career. Thank you for taking the chance on me.

Tinsley, thank you for pushing me at just the right times. In particular, I truly appreciated your words of support that one day last December when work was piling on. I am excited to see what heights you will take the Global Literacy Project in the upcoming years.

Maryanne, you have served as a continuing reminder that this is about people and not just about playing around with shiny toys. I am fortunate to have had a brief glimpse of your experience and passion for understanding the reading brain.

Robin, I admire your drive to keep us on track despite all of the mistakes, last-minute fixes, and sometimes messy methodology. With the patience of Job, your willingness to help figure this out as we went allowed us to iterate on a platform that just keeps getting stronger with every deployment. Despite its warts, I am proud of the work, and I have you to thank for that.

Polly, thank you for making me feel instantly and forever more at home in PRG. "Always a Robot..."

Angela, thanks for inviting me take part in your research; I had no idea just what I was getting into, but I am so glad you were there to root me on! **Brad**, thank you for all the gallows humor and unforgettable times running experiments this summer - you were a worthy adversary and officemate. **Goren**, your enthusiasm and insight are truly inspiring; it made me feel like we have something incredibly valuable here.

Gibbs, I truly enjoyed our brief shared office time this summer. You were my unwitting muse for this thesis; as I was writing, I kept asking myself, "Will Gibbs appreciate this?" You have an enormous job ahead of you, but I cannot think of a more capable, responsible, and creative problem solver than you to pull it off. You can do it, and I'll be rooting for you!

To **students, faculty, staff and fellows at the MIT Media Lab**, I've had the precious opportunity to be staggered by your talent; walking among you has been an unbelievable dream and privilege. There is no amount of chocolate that will express my gratitude to you, **Linda**, for your shepherding and advocating. Thank you to **Tod and my new Opera colleagues** for giving me refuge and such welcoming support in the homestretch.

Most especially, thank you to my **colleagues in the Personal Robot Group**. It turns out that *you* were the rock stars, this whole time. I will be forever indebted to you for sharing your passion, wisdom, Mexican Martinis and so much laughter. I will miss these past two years, but I am looking forward to watching you flourish and am excited for when our paths will most certainly cross again.

To **David Crouse** and **Christy Fordham** in Roanoke, thank you for embracing disruptive approaches; you came along for the ride and grabbed the wheel to steer us along the way. Your dedication and love for your community and students is awesome.

Finally, I thank the unnamed hundreds of children, parents, and educators who have participated in this research. When I think about the scope of this project, I am reminded at once how far we've come and how much further there is to go. I'd like you to know we are learning so much with your help, and your contribution really will make a difference to people all around the world.

Thank you.

Table of Contents

List of Figures	9
List of Tables	10
1 Introduction	11
2 Background	14
2.1 Field Research	14
2.2 Literacy Tablets and Apps	15
2.3 Data Collection Framework	17
3 GlobalLit System Architecture Goals	19
3.1 Goal: Geo-shifted data normalized across heterogeneous deployment configurations.	20
3.1.1 A Note about Heterogeneous Deployments	20
3.2 Goal: Time-Shifted analysis facilitates study of life-cycle of tablet use and exploration models.	23
3.3 Goal: Probes available for examination at a variety of scales (ex. Regions, Deployments, Tablet Groups (Conditions), Individuals).	23
4 GlobalLit System Description	24
4.1 GlobalLit Remote	24
4.1.1 Funf Open Sensing Framework	24
4.1.2 FunfBackgroundCollector	29
4.1.3 FunfFileMover	30
4.2 GlobalLit Live	32
4.2.1 GlobalLit Apps	36
4.2.2 Deployment, Server Side Processing, and GlobalLit Admin	42
5 Case Study - GlobalLit Remote	47
5.1 Pilot Deployments	47
5.2 Hypotheses	48
5.2.1 Study Deployments	48
5.2.2 Tablet Content	50
5.3 Data Cleaning	52

5.4	Results	54
5.4.1	Dwell Time	54
5.4.2	Roanoke Full Day	54
5.4.3	Roanoke Half Day	55
5.4.4	Roanoke Home	56
5.4.5	Cross Deployment Dwell Time Analysis	58
5.4.6	Most Used Apps	59
5.5	Discussion	65
6	Case Study - GlobalLit Live	66
6.1	Introduction and Background	66
6.1.1	Dialogic Reading	66
6.1.2	Dialogic reading with Apps	67
6.2	Experiment	68
6.2.1	Hypotheses	69
6.2.2	Study Task & System	70
6.2.3	Method	75
6.2.4	Results	84
6.3	Discussion	97
7	Conclusion	100
7.1	Goal: Geo-Shifted data is normalized across the globe. . .	100
7.2	Goal: Time-Shifted analysis allows us to understand life-cycle of tablet use and exploration models.	101
7.3	Goal: Probes can be examined at a variety of scales (ex. Regions, Deployments, Tablet Groups (Conditions), Individuals)	101
7.4	Contributions and Outlook	102
8	References	103
	Appendix A Sample Probe Data	110
	Appendix B Apps on Tablets	116
	Appendix C Crowd-Sourced Questions	121
	Appendix D Colophon	123

List of Figures

1	Heterogeneous Connectivity	21
2	FunFileMover User Interface	33
3	Tinkerbook Scene Composition in Unity3D.	37
4	Matching app with lower case letter matching.	39
5	Matching app with semantic categories.	40
6	User interface for Launcher app.	41
7	Settings Mode for Launcher.	42
8	Web-based administrative interface for GlobalLit Platform	43
9	Class diagram and data model for GlobalLit Admin.	44
10	Roanoke deployment duplication in process.	51
11	Roanoke tablets ready to deploy.	51
12	Dwell time - Roanoke FULL DAY.	55
13	Dwell time - Roanoke HALF DAY.	56
14	Dwell time - Roanoke HOME.	57
15	Dwell time - UGANDA.	58
16	Cross Deployment Dwell Time	59
17	Top ten apps - Roanoke WHOLE DAY	61
18	Top ten apps - Roanoke HALF DAY	62
19	Top ten apps - Roanoke HOME	63
20	Top ten apps - Roanoke UGANDA	64
21	Sample Tinkerbook interactions include Iconimation hints and word-graphic highlighting.	71
22	Tinkerbook with prompting character.	71
23	Tinkerbook with Prompting Scenes Highlighted.	72
24	Annotated video frame from parent-child interaction.	78
25	Percentage of coded utterances per category across all conditions.	84
26	Percentage of coded utterances per category by condition.	85
27	Total seconds of reading sessions.	88
28	Engagement Related Utterances	88
29	Behavior Related Utterances	90
30	Full Utterances	90
31	Single Word Utterances	92
32	Novel Word Utterances	92
33	Lexical Diversity	94
34	Turn taking exchanges.	94
35	Number of Questions	96
36	Distancing Prompts	96

List of Tables

1	Description of database filename with metadata embedded in naming scheme.	29
2	Parameters sent to GlobalLit server during administrative ping. .	32
3	Characteristics and Demographics of Roanoke	49
4	Number of Subjects per Condition	51
5	Expert Prompts for Tinkerbook by Scene	73
6	Number of Subjects per Condition	76
7	Categories for Coding	79
8	Words Included in Tinkerbook	81
9	Summary of Measurements	83

1 Introduction

In the US, by the age of 4, a child in a middle-class family in the US will have heard 42 million words. A child in a welfare family will have only heard 13 million words. This gap in vocabulary at such an early age is one of the primary differences between kids who learn to read easily and kids who struggle in school (Hart & Risley, 2003).

Children that have had limited opportunity to experience reading and language ahead of their first exposure to a classroom are at a significant learning deficit. Compared to their peers who have access to an enriching, stimulating home reading life, these students are more likely to struggle in school, and this is especially true for low-income families (Forget-Dubois et al., 2009). There is an incredible need to improve pre-literacy skills among children entering schools, and the circumstances of a child's home life can drastically affect her school readiness (La Paro & Pianta, 2000).

In far too many cases, children simply do not have enough access to language in their homes, and this presents a challenge for any intervention in these unstructured environments. Mobile devices, like tablet computers, are potentially potent tools for delivering language learning to children otherwise unreachable at any reasonable scale in a variety of home situations (McManis, 2012). Unfortunately, the current selection of available literacy-related educational software is sparse, and good examples are hard to find among an overwhelming sea of poorly designed and even counter-productive "educational" apps (Guernsey, Levine, Chiong, & Severns, 2012). Nevertheless, children are consuming media from mobile devices at an ever-increasing pace (Gutnick et al., 2011).

Parents must sort through 20,000 poorly labeled and uncategorized educational games in the Apple App Store (Shuler, Levine, & Ree, 2012). They hope they can choose apps that can educate their kids on devices like the iPad, but they are fighting a losing battle against this onslaught of content and marketing from so many app developers.

Before coming to graduate school at the MIT Media Lab, I was part of this problem. I ran a game development studio for a number of years that developed children's apps and games for mobile devices. I helped produce games for PBS Kids, worked with award winning book publishers, and even built a project for JK Rowling, arguably one of the most impactful authors of children's media in the last decade.

I can honestly say that most of the people I worked with loved to make joyful experiences for kids. We all intended to create beautiful, high-quality

educational apps.

I can also tell you, in my experience, that it is disturbingly rare for the average educational app producer to spend the time and resources necessary to measure anything about the actual effectiveness of their apps. Most care just about basic sales numbers or rudimentary analytics.

We do not have a comprehensive idea what kids are doing with these mobile devices, much less whether they are learning anything from them.

What if we knew apps they open and for how long? What if we could see how they progress through an app, exactly? Where are they playing these games? Are they playing with their parents or are they alone? How does their use change over time? Ultimately, are they achieving learning objectives or just silently poking at screens?

Suppose educators could **collect** detailed, minute-by-minute, information back from these devices, **analyze** the data with the guidance of education and literacy experts, and then **react** to these findings. Could they use this data implement efficient, low-cost interventions? Could they deliver on the promise of educating diverse populations of children where traditional structures of learning, like schools, human tutors, or highly fluent households are unavailable?

Such a framework could allow expert educators to categorize and prescribe third-party educational apps using an objective, needs-based methodology.

This thesis describes the initial prototype efforts from 2011-2014 for conducting this kind of research. It leverages low-cost tablet computers to collect, analyze, and react to data about child-driven learning as a technology platform contribution. This research proceeded in collaboration with the Global Literacy Project, a joint venture between MIT, the Dalai Lama Center for Ethics and Transformative Values, Tufts University, and Georgia State University.

The Global Literacy Project is tackling a moon shot challenge of bringing literacy to 100 million children. This ground shift of empowerment could help build empathy among cultures, reduce poverty, and improve the health and prosperity of people all around the globe (Wolf, Gottwald, Galyean, Morris, & Breazeal, 2014).

I view the contributions of my thesis as a modest exploratory step in a long series of upcoming work. It explores technology as a facilitator for literacy development in circumstances where teachers and parents can not guide very young students through important stages in learning to read.

The first section of this document presents background and recent work on field research, data collection frameworks, and the current state of tablets and applications applied to children's literacy and education.

Next, the thesis will provide motivation to approach some of the challenges involved in field research with children and tablets while illustrating some of the potential ways research might benefit from a system.

Then it will describe the design considerations and architecture behind the GlobalLit system, a technology framework for data gathering, processing, and tablet administration.

Two case studies, highlighting different modalities for the GlobalLit system, follow the system description. In the first example, example deployment scenarios for this system demonstrate data collection and analysis from deployments around the world. It will demonstrate how this data can be used to make observations about tablet usage in the field. The case study will describe the unique challenges in places like Ethiopia, Alabama, and the Boston Museum of Science, including what makes data collection difficult in each case.

In the second case study, the GlobalLit system is used to collect high fidelity data in a live, laboratory study of parent-child interactions around a single app on the tablet.

Finally, the thesis will discuss how well the system fulfilled its architecture goals and suggest improvements and future directions so that it could be further used as a model for scientists and educators to engage in research of this nature (either through using the system as an experimental platform, itself, or to construct their own deployments using lessons from this work).

2 Background

2.1 Field Research

From 1947 through 1972, Dr. Roger Garlock Barker conducted a large-scale, long-term observational study on the children in the small town of Osaloosa, Kansas. He set up the Midwest Psychological Field Station through which he and his research team could collect detailed empirical observations about the community (Roger G Barker & Wright, 1954).

In the early days of psychological research, as his contemporaries like B.F. Skinner were conducting pioneering research on rats in labs, Barker argued that studies of human nature should involve actual human beings. It should take place where these people naturally lived and worked in their daily lives. Understanding the environmental conditions, or what he called the “behavior settings,” was critical in understanding social dynamics and child development.

To explore this idea, he and his researchers embedded themselves in the Kansas community over several years and documented seemingly minute observations about individual children and their interaction with their families, peers, and schools.

Capturing a variety of data probes like noting what a family was eating or marking down the facial expressions of parents as they asked their child questions, the team made tens of thousands of minute-by-minute observations such as:

7:01. Raymond picked up a sock and began tugging and pulling it on his left foot. ...

7:07. Raymond turned to his dresser and rummaged around among the things on it until he obtained a candy Easter egg for his dog (R G Barker & Wright, 1951).

His approach was to collect as much data as possible, first. Then he and his team would code, aggregate, and organize the notes into coherent episodes. He would finally sort out the meaning in post-hoc analysis.

Some of Barker’s peers questioned his methodology and wondered how he could engage in rigorous science with so many confounding factors outside of a lab. For example, the very presence of researchers in homes and classrooms could influence behavior. Critics were especially concerned about his approach (i.e. observing subjects without first defining a rigorous set of

hypotheses). Practically, sorting through this mountain of data required much stamina and was open to researcher bias and errors, if nothing else. (Sabar, 2014)

Nevertheless, the magnitude of data collection was staggering and unprecedented. Using this repository as the foundation for a number of investigations, he was able to propose a new set of ideas in the field he coined, Environmental Psychology (Roger G Barker, 1965). He was even awarded a Distinguished Scientific Contribution Award from the American Psychological Association simply for obtaining such a rich set of data about these children.

Today, field studies of this scope are rare due to cost and logistical practicalities. Recent examples include Deb Roy's Human Speechome Project, where he recorded the spoken words of his children in his home over the course of several years to study language development (D. Roy et al., 2006). Also, The University of California, Los Angeles Center on Everyday Lives of Families, put video cameras in homes for a week to learn about family dynamics (Repetti, Wang, & Saxbe, 2009).

2.2 Literacy Tablets and Apps

Increasingly, tablet computers are used by children in classrooms institutionally (Twining, Cook, Ralston, & Selwood, 2005) and with support software that facilitates mobile devices for time and place shifted learning. A child can engage in reading activities on a tablet anytime and anywhere (Fletcher, Tobias, & Wisher, 2007). Children can use tablets filled with educational software, successfully in places like Kenyan libraries (Kinyanjui, 2014), Sri Lankan pre-schools (Priyankara et al., 2013), or homes in New Zealand (Fleer & others, 2013).

Initial studies have shown promise in tablet usage at home improving the school readiness for at-risk preschoolers (Gonzalez & Fryer, n.d.).

Governments are supporting nationwide deployments of tablets to schoolchildren. Programs are commencing in Brazil (Ally, 2014) and devices have been deployed in Thailand (Viriyapong & Harfield, 2013).

Tablets afford new opportunities for spontaneous play and encourage social learning experiences. For example, when children use tablets at home, it often includes playing and interacting with parents and siblings. (Verenikina & Kervin, 2011). Children seem to be very comfortable with tablet interfaces and readily take to the form factor of a tablet (Plowman, Mcpake, & Stephen,

2008), tolerating small technical glitches or interface difficulties (Couse & Chen, 2010).

When kids do have access to mobile devices in their homes, they use them for about 48 minutes every day. However, parents must sort through 20,000 poorly labeled and categorized educational games in the Apple App Store in hopes their selection of apps can provide a reasonable educational experience for their children (Chiong & Shuler, 2010). It is no surprise that of all the digital media available to children in homes (including television and video games), parents consider mobile phones to be the least appropriate for learning (L. Takeuchi, 2011).

Researchers are proposing rubrics through which experts can evaluate apps can for educational appropriateness and engagement (Gonzalez, 2014, Ruttkay, Bényei, & Sárközi (2014), Tsai & You (2013)).

Of the apps categorized as educational in the Apple app store, only about 5% are related to literacy (Shuler et al., 2012). Collaborators at Tufts University conducted an initial, unpublished review of available literacy apps in the Android app store. They confirmed that only a very small set of apps could be considered even barely adequate in both pedagogy and engagement (Wolf & Gottwald, 2013).

Even worse, third-party literacy apps on these tablets do not sufficiently provide coverage for all the skill categories required to nourish a developing reading brain. Guernsey et al. surveyed offerings available in both the Android and Apple app stores searching for literacy-related educational apps. They found that apps that claimed to teach reading skills seemed to focus overly on the earliest of skills (letters and sounds, phonics/ word recognition). They also did not seem to address comprehension and grammar skills required for reading (Guernsey et al., 2012).

Various comprehensive tablet projects have attempted to scaffold a child's literacy development using structured, ever-evolving literacy approaches. For example, Nell is an attempt to implement a story-telling framework that grows with a child and can adapt to his improving skills (Ananian, Ball, & Stone, 2012). TinkRBook (Chang & Resnick, 2011) is a system for encouraging interactive play with words, language, and comprehension. However, these examples are limited in their scope and don't provide feedback into a larger ecosystem of multiple learning apps on a tablet.

Goodwin and Highfield conducted an analysis of the pedagogical design of apps in the Apple App store. They showed that 75% of the "top" apps in the education category would be classified as instructive (linear progression or "lesson" based), rather than constructive (free-form, creative activities) apps. The app stores tend to categorize open-ended apps in game and enter-

tainment categories, despite their educational value. The authors suggested that both app developers and parents have preconceived notions about what makes a good learning experience (ex. Drills, linear progression through skills) (Goodwin & Highfield, 2012). These skills might not map to more creative and engaging apps.

In addition, teachers are unsure about the effectiveness of these devices in concert with their class activities. They often have a pessimistic attitude regarding the value of the technology, and this creates a confirmation bias that negatively impacts the acceptance of tablets in schools (Ifenthaler & Schweinbenz, 2013). Deploying tablets in schools presents a unique set of logistical challenges, as well (Ali, 2013), especially if the school implements strategies for allowing children to take home the tablets for after-hours reinforcement (Walling, 2014).

However, when it comes to literacy, simply increasing exposure to tablets does seem to help. In an Australian study of pre-schoolers ($N = 109$), those with access to tablet devices had higher letter sound and name writing skills. The researchers speculate that the quality of those experiences matters as there did not seem to be an effect related strictly to quantity of time spent on tablet. The researchers speculate that home-life factors, like communicative parents or siblings' engagement, were more important (M. M. Neumann & Neumann, 2014). Since this study relied on self-reporting from parents, it is difficult to quantify which activities the children were doing with the devices.

There exists some work to understand how digital devices, like desktop computers, can influence early literacy skills (Hisrich & Blanchard, 2009), but tablets are too new a technology to have many long-term examples to reference.

Neumann & Neumann propose a framework through which to examine tablets and emergent literacy development, provide guidelines for literacy app design, and suggest a fertile set of research questions which should be explored with tablets (M. M. Neumann & Neumann, 2014). In particular, they emphasize the lack of understanding about how children might integrate tablet education in their homes and day-to-day life alongside traditional schools.

2.3 Data Collection Framework

Tablets afford a possibility to collect large amounts of data regarding how children use mobile devices. Compared to Barker's labor-intensive work, the

means of data collection with tablet computers can be relatively inexpensive, fairly automatic, and widely dispersed geographically.

The government of Thailand is endeavoring to distribute tablets to all children entering their first year of school as part of a “One Tablet Per Child” project. Two of the major challenges have been not being able to provide support for the teachers and not being able to evaluate the learning outcomes of the activity (Viriyapong & Harfield, 2013). They have piloted the use of real-time monitoring that combines data mining techniques with adult observations of the children to give teachers more insight into the individual and aggregated class learning (Harfield, Jormanainen, Rungrattanaubol, & Pattaranit, 2013). While their results seem promising, their system still requires a complete infrastructure of a school and strict curriculum, a very small selection of custom app activities, and significant teacher training and oversight.

There is also a potential opportunity to develop a framework, with which any 3rd-party literacy app can integrate, that provides data-driven app suggestions based on individual child’s needs. An offline, theory-based scaffolding, like RAVE-O (Wolf, Miller, & Donnelly, 2000), could be augmented via content delivered on the tablet.

An effective system might employ user-interface and social persuasion techniques (Mintz & Aagaard, 2012) to guide a child’s exploration across a greater variety of apps. If the system can access assessments of a child’s literacy skills as an input to its suggestion algorithms, it could also encourage the child to engage with apps that could deliver appropriate content developmentally. A fully instrumented system could generate feedback directly to the child or his learning community (including peers, parents, and teachers) about the experiences across the apps.

The system described in the remainder of this thesis, called GlobalLit, is an attempt to provide an infrastructure for research, both in the field and in labs, about child-driven literacy learning using tablet computers.

3 GlobalLit System Architecture Goals

This chapter describes the technical architecture and design considerations for the various subsystems that comprise the GlobalLit system. It discusses some of the unique challenges faced in deployments along with potential remediation techniques.

The main purpose of the GlobalLit system is to provide infrastructure for scientific inquiry about the way children use mobile technology. It is a tool to study how children develop literacy skills with or around mobile devices. Ultimately this technology might facilitate interventions and support for individuals around the world to engage in literacy development.

First, the system must facilitate researchers and child development experts to generate **collections** of useful information about the target population, regardless of where the subjects and researchers physically reside. Then, the system must provide this data in a structured format, with a reasonable expectation of fidelity and accuracy; researchers or educators will produce **analysis** of this data to draw conclusions from deployments. Finally, the system should provide a feedback path for researchers to create conditional studies or for educators to provide interventions. It should facilitate modifying the tablets remotely, with these **reactions** based on sound observations and grounded in theoretical scaffolding.

To support the magnitude and diversity of experiments and potential interventions, in a variety of contexts, architecture goals for such a system include:

- Geo-shifted data normalized across heterogeneous deployment configurations.
- Time-Shifted analysis facilitates study of life-cycle of tablet use and exploration models.
- Probes available for examination at a variety of scales (ex. Regions, Deployments, Tablet Groups (Conditions), Individuals).

The implementation of GlobalLit, as described in this thesis, represents a substantial and useful first attempt at achieving these goals, but it did not successfully address all aspects of an ideal system. Below is a description of each goal along with their challenges.

3.1 Goal: Geo-shifted data normalized across heterogeneous deployment configurations.

The system is intended to be deployed all around the world. Each site will present wildly unique constraints, goals, and opportunities. The system must maintain its core aim of data collection, analysis and reaction, despite unpredictable situations.

The system should:

1. Provide normalized and structured data to facilitate cross-site studies can.
2. Handle data collection from structured environments like schools and more unpredictable sites like homes.
3. Function robustly enough for rustic deployments with no guarantee of network connectivity or even reliable electricity, buffering data and provided alternate transmission paths.
4. Provide a path to scalability of a large number of sites and tablets.
5. Allow for third-party developers to create apps that work seamlessly with the system.
6. Respond to the cultural sensitivities of various regions. For example, the initial content mix on the tablet should be adjusted to reflect the value system of the target community. The types and volume of data collected must meet the privacy expectations of families using the tablets.
7. Allow for technical and non-technical local support personnel to maintain and administer deployments.

3.1.1 A Note about Heterogeneous Deployments

Multiple deployment configurations should be supported simultaneously to achieve the design goal of flexible, geographically diverse data collection. A “deployment” is a collection of tablets with GlobalLit software, sent to a given site, with an explicit start date.

We can categorize deployments into (at least) four types:

- **Disconnected-Remote Tablet Groups** - Deployment assumes no or very infrequent internet connectivity on site. Data records to the tablet; manual intervention for collection of data from tablets must occur on a reasonably frequent basis (i.e. “Sneaker Net”). For example, a technician could swap out sd card disks from tablets, aggregating the data at a central, better-connected workstation that can then upload the data to the

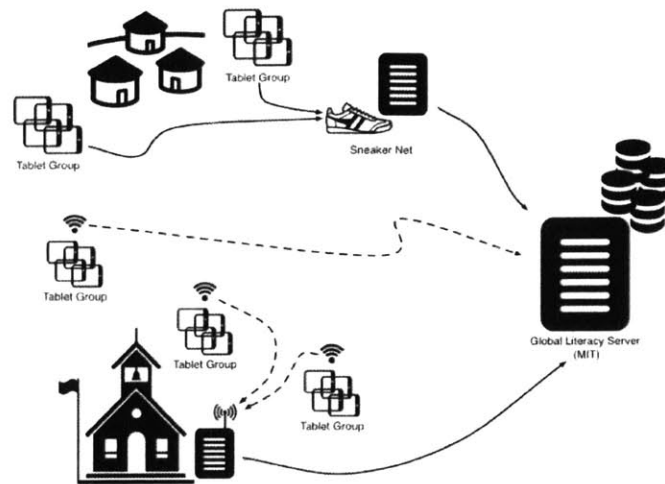


Figure 1: Heterogeneous Connectivity

server. Alternatively, flash drives or external drives can be shipped to the research regularly for uploading and processing.

- **Directly-Connected-Remote Tablet Group** - Tablets have embedded WiFi or cellular connectivity and regularly make contact with a central server to upload data and receive updates. This usage scenario is an ideal case for sites with adequate network bandwidth; the frequency of uploads can be throttled to make best use of the uplink.
- **Indirectly-Connected-Remote Tablet Group** - A deployment can scale to a large size or may have strong, but intermittent network connectivity (for example, in countries with poor internet backbone). In these cases, it becomes impractical for the tablets to attempt to make a connection to the GlobalLit servers directly. Instead, the tablets can make a high-speed connection to a local server cache and this, in turn, trickles data through a more stable connection. Deployment protocols might encourage distributed tablets to return to a central location on a regular basis; for example, children take tablets for at-home use but charging cords could remain at the school.
- **Directly-Connected-Live Tablet Group** - For some studies, it may be necessary to collect data in real time as children are interacting with apps on the device. For instance, we may want to test the design of

a recommendation engine, which requires more robust processing beyond what a tablet could provide alone. The tablet could establish communication with a server to provide feedback in real time. GlobalLit provides facilities for this configuration as well.

Each deployment might assign tablets to different groups corresponding to experimental conditions. GlobalLit, currently, does not implicitly support this categorization, but as long as the researcher maintains a tablet ID to condition mapping, this data can be properly analyzed.

For each of the initial deployments in the past few years, the batches of tablets involved identical Android hardware and software at each site. However, there was no consistency of hardware among the different deployments. Also, even at the same site, software was not always identical between tablets.

Indeed, in order for GlobalLit to scale to millions of children, the system must support a heterogeneous mix of technology, as well. Different sites will have different access to tablets (or cell phones, even). The software should install and work seamlessly across different hardware.

Currently, however, system configuration is tied very closely to specific hardware models. Even devices from the same manufacturer and model line, but with minute hardware differences require a reimplementations of much of the software stack and deployment duplication recipes.

The lack of systematic technology upgrades and management becomes problematic if GlobalLit is expected to scale beyond the capacity of one or two graduate students to prepare tablets for deployment. Instead, consistent deployment recipes and labor should be shared among deployment partners.

There are two major areas currently dependent on hardware architectures.

1. Access to certain low-level features varies wildly among Android tablets. In the 2011 version of Motorola XOOM, for instance, access to the external sd card was disabled, so the deployment required workarounds to support copying data to those drives. Other idiosyncratic differences, like the procedure to modify networking settings, the sequence of key presses required to enable safe mode, or even the expected location of application files are not predictable on each new device encountered. Often the lowest cost tablets are poorly supported and documented.
2. Duplication of tablets can be a very tedious process. The first deployment of 41 tablets to Ethiopia took an entire week to copy, (including Christmas Eve!). Now, we can duplicate forty tablets in a few hours. In order to achieve this rapid duplication, without access to the manufacturer's source images or kernel-level compilation, deployments re-

lied on the community of Android hardware enthusiasts to provide “rooted” images and “recovery mode” modifications for a device. These custom builds of Android software are very hardware specific; for example, a recovery mode for a United States version of a Samsung Tab III may not necessarily work on the European version of the same tablet.

Future versions of the GlobalLit system should relax the dependency on device root access and provide multiple, user-friendly methods for duplicating the software (for example, as user-installable applications in the Google Play Store).

3.2 Goal: Time-Shifted analysis facilitates study of life-cycle of tablet use and exploration models.

The system should:

1. Support long-term deployments and studies (months and years).
2. Maintain flexibility to adjust data probes, remotely, as the site conditions and experimental design require.
3. Provide mechanisms to refresh content and change mixture of data probes over time.
4. Report problems and maintenance issues to deployment managers so that deployments never atrophy from lack of use over time.

3.3 Goal: Probes available for examination at a variety of scales (ex. Regions, Deployments, Tablet Groups (Conditions), Individuals).

The system should:

1. Provide for adjustable resolutions of data collection. For example, in year-long deployments, interactions measured per millisecond may not be as important as interactions measured by day.
2. Organize data by location or conditional groups.
3. Provide mechanisms for individual identification of users.
4. Provide an infrastructure for lab studies where connected tablets participate in the collect-analyze-react cycle in near real-time.

4 GlobalLit System Description

The GlobalLit system consists of a collection of software to manage the collection migration of data from sites or studies for processing on a central server. The tablets, themselves, host the majority of the software, but the system also refers to server-side processing and administrative software installed to host study data.

4.1 GlobalLit Remote

The GlobalLit Remote system refers to the data collection and management suite built to run, asynchronously, on Android-based tablets and mobile devices as part of deployments. GlobalLit Remote is completely encapsulated to run on a device in isolation. It is expected to be functional in deployments with high-latency network connections or where data collection does not need to be real-time. For example, in some deployments with no internet connections on-site, GlobalLit Remote is used to batch data for manual extraction on a monthly basis.

4.1.1 Funf Open Sensing Framework

The backbone of the data collection component of this system relies on the Funf Open Sensing Framework, a project developed and launched at the MIT Media Lab (Aharony, Pan, Ip, Khayal, & Pentland, 2011).

“The Funf Open Sensing Framework is an extensible sensing and data processing framework for mobile devices, supported and maintained by Behavio. The core concept is to provide an open source, reusable set of functionalities, enabling the collection, uploading, and configuration of a wide range of data signals accessible via mobile phones.” (“funf | Open Sensing Framework,” n.d.)

Funf maintains a collection of software that runs in the background on an Android device without user intervention. It manages a configurable collection of data probes that report either on periodic (data collected at a set frequency, on a timer) or episodic (data collected in response to some event) basis.

The background process stores probe data as entries in a temporary sqlite3 database file. Funf copies the database to either an archive directory on the device's internal storage or directly onto a removable sd card if that is the configured behavior.

Then these archive files are copied off of the device (either through manual copying or uploading to a server). A copy of these database files exists in a flexible backup location on the device, as well, that is pruned of old files when the size of stored files reaches certain capacity thresholds.

4.1.1.1 A Note about Timestamps

Each Funf data recording should have an associated Timestamp that, ideally, corresponds to the exact time of the data reading (measured in milliseconds from the epoch date of January 1, 1970). The device's internal clock provides this time. If a tablet can make an internet connection, it can usually set its own time using a Network Time Protocol server on the network. This behavior has been problematic in a number of cases where the clock's accuracy cannot be assured:

- In Ethiopia, there was no electricity in the villages. Children would use the tablets until they completely ran out of battery power; they likely continued to press the power button, draining reserve power. The clock lost its setting as a result and had no way of resetting. To mitigate this, GlobalLit contains software that periodically checks the current set time of a device and determines if the clock seems off using the heuristic:
 - Is the current clock date before or after the deployment year by more than one year?
 - Is the date more than one week before the last known good timestamp?
 - Has it been more than two weeks since the last GPS time check? If the system needs to reset the clock, it turns on the device's internal GPS chip and communicates with GPS satellites, which continuously broadcast a timestamp. The software adjusts the tablet clock to match the GPS time.
- Some probes were set to use seconds instead of milliseconds from epoch, causing range errors in recordings. Post-processing converted timestamps to milliseconds when necessary (i.e. if the digit length of the timestamp is too short, multiply by 1000).

- Since tablets are all running off of their individual clock, it is possible that tablets on the same deployment might have discrepancies in their timestamps, so reconstructing cross-tablet chains of events becomes difficult. The problem is somewhat mitigated when tablets can connect to a network, updating clocks automatically using services like NTP. A future approach might be to have the tablets use a peer-to-peer protocol to propagate impulse messages to introduce tags in tablet data streams that can be used to synchronize clocks.

As a failsafe, each tablet also maintains an internal counter that is incremented by one every time the tablet archives its local data. Thus, the counter preserves the order of data collected from a given tablet, even if the timestamps are not completely trustworthy.

In addition, the system provides for remote configuration of type and frequency of data collection via a configuration file hosted on a server. The tablets regularly query a given URL (the URL and frequency of checking, themselves configurable) to download a configuration file that provides updates to the data probe mix. This feature allows researchers, for example, to throttle a particular probe to gather data at different frequencies per experiment protocol.

4.1.1.2 Funf Data Probes

The deployments have used a variety of built-in Funf data probes and a number more of custom probes written as part of the thesis project. Not all probes proved useful for each deployment; some configurations retired certain probes due to bandwidth consumption and other constraints.

4.1.1.2.1 Periodic Probes Periodic probes continuously collect data in the background on a timer. The frequency of collection is configurable per tablet and deployment. The data collection contains the following probes:

- **HardwareInfoProbe** - (built-in) A description of the device's hardware, manufacturer, and other device parameters related to hardware sent every few days.
- **BatteryProbe** - (built-in) The status of the charge and health of the device battery sent hourly.
- **RunningApplicationsProbe** - (built-in) A snapshot of the currently front-running application and the last five applications that have had user focus. This probe turned out to generate disproportionately large data set (lots of

text per reading). Also, children would often switch apps much faster than this probe could reasonably. Ultimately, deployments excluded this probe in favor of the custom LauncherApp probe.

Because periodic probes are should run on a schedule, as long as the device is functioning correctly, we can treat these probes as a makeshift “heartbeat” for the tablet. As long as we see this data coming through to the server regularly, we have hope that the tablet is operational enough to maintain basic GlobalLit functionality. To put it another way, when analyzing data from a deployment, we might highlight gaps in the periodic probe stream. The holes in the data are potential exceptions to resolve through data scrubbing (ex. a tablet is malfunctioning, or a teacher chose not to pass them out one day).

4.1.1.2.2 Episodic Probes Episodic probes record data when an event occurs to trigger the probe. The majority of probes we use are in this category.

Some episodic probes appear to be periodic because the event that triggers them is, itself, on a timer. However, as the timer is external to Funf, there are no guarantees made about the regularity of the probe, even if Funf is running perfectly. The data “heartbeat” cannot rely on these probes:

- `FileMoverService` - Triggered by log messages from the file uploading and remote administration app, `FunfFileMover`, executes an action.
- `GpsDateFixService` - The system regularly uses GPS satellites to validate that the internal clock seems reasonably accurate.
- `bgcollector.MainPipeline` - Triggered by log entries from the Funf system as it regularly archives and uploads data files

The remaining probes record data in response to non-deterministic user or device events:

- `ScreenProbe` - (built-in) Triggered when the screen turns on or off.
- `LauncherApp`, `LauncherAppPaused` - Messages associated with our custom launcher as a means of recording what apps children were opening.
- `RecorderService` - Triggered when the camera was used to record a photo of the user.
- `TinkerText`, `tinkerbook`, `worldliteracyapp`, `Matching` - App-specific data from various custom apps built for particular studies or as part of the world literacy app class projects taught at MIT.

Listing 4.1: Code snippet to be used by third part app developers to insert data into GlobalLit system.

```
// Uses actions and keys defined in
// edu.mit.media.funf.storage.DatabaseService
// edu.mit.media.funf.storage.NameValueDatabaseService

public void sendLog(String name, String value) {
    Intent i = new Intent();
    i.setAction(DatabaseService.ACTION_RECORD); //
    Bundle b = new Bundle();
    b.putString(DatabaseService.DATABASE_NAME_KEY, "mainPipeline");
    b.putLong(NameValueDatabaseService.TIMESTAMP_KEY, System.
        currentTimeMillis()/1000);
    b.putString(NameValueDatabaseService.NAME_KEY, name);
    b.putString(NameValueDatabaseService.VALUE_KEY, value);
    i.putExtras(b);
    sendBroadcast(i);
}
```

Listing 4.2: Sample probe data from a FingerDown event in Tinkerbook.

```
{"type": "FingerDown", "x": 16.4, "y": 35.25}
```

4.1.1.3 Third-Party Data Hooks

Any third party app developer could send data into the GlobalLit system simply by using the Android Broadcast Intent messaging system to broadcast a message. A sample code snippet is in Listing 4.1:

The value, in this case, can be any arbitrary string. Bandwidth considerations aside, a structured format, like JSON, can encode sophisticated data.

For example, Tinkerbook publishes data that looks like Listing 4.2.

Samples of each probe's data can be found in Appendix A.

4.1.1.4 Funf Sqlite Databases

The data is stored and transmitted as a collection of Sqlite3 databases. Each file maintains a simple name-value table that consolidates all probes into the same structure, as in Listing 4.3.

Listing 4.3: SQL command for simple schema description.

```
sqlite> .schema data
CREATE TABLE data (_id INTEGER primary key autoincrement, name TEXT,
    timestamp INTEGER, value TEXT);
```

It is up to each probe to publish the value of the probe in whatever format is most prudent. The built-in probes tend to format data as JSON, for example.

These individual readings are called “Funf Data Rows,” where each row has a timestamp and the name of a corresponding probe.

4.1.1.4.1 Filename The individual database file names are modified to provide extra metadata for upstream processing. An example filename and description of the annotated metadata appears in Table 1.

Table 1: Description of database filename with metadata embedded in naming scheme. Sqlite database filename and components using example filename 1384035555-00000106_1700618844400517_7fbd3be-60cc-4ab1-8324-d653c21a8430_1384035614_mainPipeline.db

Component	Description
1384035555	Timestamp of upload.
00000106	Counter that gets incremented every time the tablet uploads a file. This value preserves the order of data received when the tablet’s clock is not trustworthy.
1700618844400517	Actual hardware serial number from device
7fbd3be-60cc-4ab1-8324-d653c21a8430	Funf generated id for the tablet
1384035614	Timestamp of time file was saved on device
mainPipeline	Name of Funf “channel” (Although the Global Literacy Project deployments did not use this feature, the system allows for different streams of data, compiled as separate repositories).

4.1.2 FunfBackgroundCollector

An app called FunfBackgroundCollector is the GlobalLit instantiation of the Funf system. It runs in the background on the device and is responsible for the collection and uploading of the data collected.

To support the various modes of connectivity, GlobalLit enhances the behavior of Funf in some crucial ways:

1. **GlobalLit disambiguates the tablet identity in the incoming data by**

collecting and transmitting hardware serial numbers for individual tablets.

Funf's original design was to collect anonymous data from adult subjects. It maintained an emphasis on privacy and obfuscation so that individuals would be willing to install the Funf system on their own devices.

To facilitate this privacy, Funf generates its own arbitrary, unique identifier that is used in the data store to identify a tablet. This Funf-generated number does not, in any way, relate to the actual Android device serial id on the device. In addition, this arbitrary serial number would often be reset or change over time (particularly if the Funf pipeline reinstalls as part of an upgrade).

However, for the studies that GlobalLit should facilitate, we desire unique and consistent tablet identifiers. For this reason, Funf was modified to augment its data collection with the hardware serial number. This serial number gets concatenated in the filename of the database sent to the server.

2. GlobalLit provides additional metadata alongside data upload.

GlobalLit modifies Funf to maintain an ordinal counter for each data file store to help mitigate timeline ambiguities.

3. Allows for Funf configuration check and upload disabling in cases where internet connectivity is unexpected.

By default, Funf does not provide a facility for ignoring the periodic configuration check. The thesis project includes alterations of the configuration system to allow for the specification of negative values for the configuration frequency. Setting this value to a negative number is equivalent to disabling the check.

4.1.3 FunfFileMover

FunfFileMover is a custom app that exports a background process to manage administrative tasks, provide extra data file management, and provide an administrative interface to accomplish simple tasks on the device.

When the tablet first starts up, it launches this app in the background and starts two periodic processes on a timer loop:

4.1.3.1 Process: Archive, Backup, Upload / Move Data Files

1. If the tablet connects to a power source, run powered actions. For example, in some deployments, we desired to reset the tablet's icons when the device was back in a classroom, charging.
2. Check to see that the tablet is set to allow data collection (in some deployments, we permitted the parent to determine whether or not certain data would be saved and collected). If data collection is not allowed, end this loop.
3. If the tablet is in a deployment with connected configuration (i.e. internet is available):
 - Execute a Funf archive and store file.
 - Save the SQLite database in the correct directory.
 - Upload files to the data server.
 - Copy successfully uploaded files to the backup directory.
 - Update Funf configuration.
4. If the tablet is in a disconnected configuration (i.e. manual extraction of data)
 - Execute a Funf archive (to save the SQLite database in the archive directory)
 - Copy all files in the archive to the backup directory.
5. If applicable, copy all files in the backup directory to an external sd card.
6. Move any audio, video, or special text files into the /archive folder (these will be uploaded on the next Funf server connection (or moved to the external sd card if applicable)).

This app also provided for some system-level tweaks to improve the viability of deployments. For example, it monitors and disables the Android status bar to prevent children from circumventing tablet security to modify settings. It also triggers the GPS timestamp fix.

4.1.3.2 Process: Ping Administrative Server

On a separate timing loop, the system connects to a GlobalLit server to transmits its serial id, version, label, hardware model identifier, any status message set by software, and its current app manifest. The metadata sent with the administrative ping is described in Table 2.

Table 2: Parameters sent to GlobalLit server during administrative ping.

Metadata	Purpose	Location
serial id	unique identifier for the device	derived from Android hardware
version	identifier for version of the GlobalLit software on device	text file '/sdcard/version.txt'
name	human-readable label for device	text file '/sdcard/label.txt'
model	model identification for Android hardware	derived from Android hardware
message	status message communicated to server for administration	text file '/sdcard/message.txt'
manifest	the app manifest of apps displayed for child in launcher app	text file '/sdcard/manifest.txt'

The tablet sends these parameters via a POST request to the server. This administrative server communication also serves as the mechanism through which commands may be executed on the tablets remotely. As a response to the tablet's ping, the server returns either:

- -1 for no command
- a shell command the tablet should execute

This is useful, for instance, to alter the app manifest to display a different set of apps for the child, to download and update system apps, or to run troubleshooting scripts and post the results in /sd card/message.txt.

4.1.3.3 FunFileMover User Interface

FunFileMover also provides a user interface through which the system displays information about the tablet (ex. version, label, serial number). It also provides for manual triggering of the data archive and upload process.

4.2 GlobalLit Live

The GlobalLit Remote system is a powerful tool for long-term, offline data collection and processing across a variety of conditions in situ. For studies with more predictable setups, sometimes it is useful to measure and react to

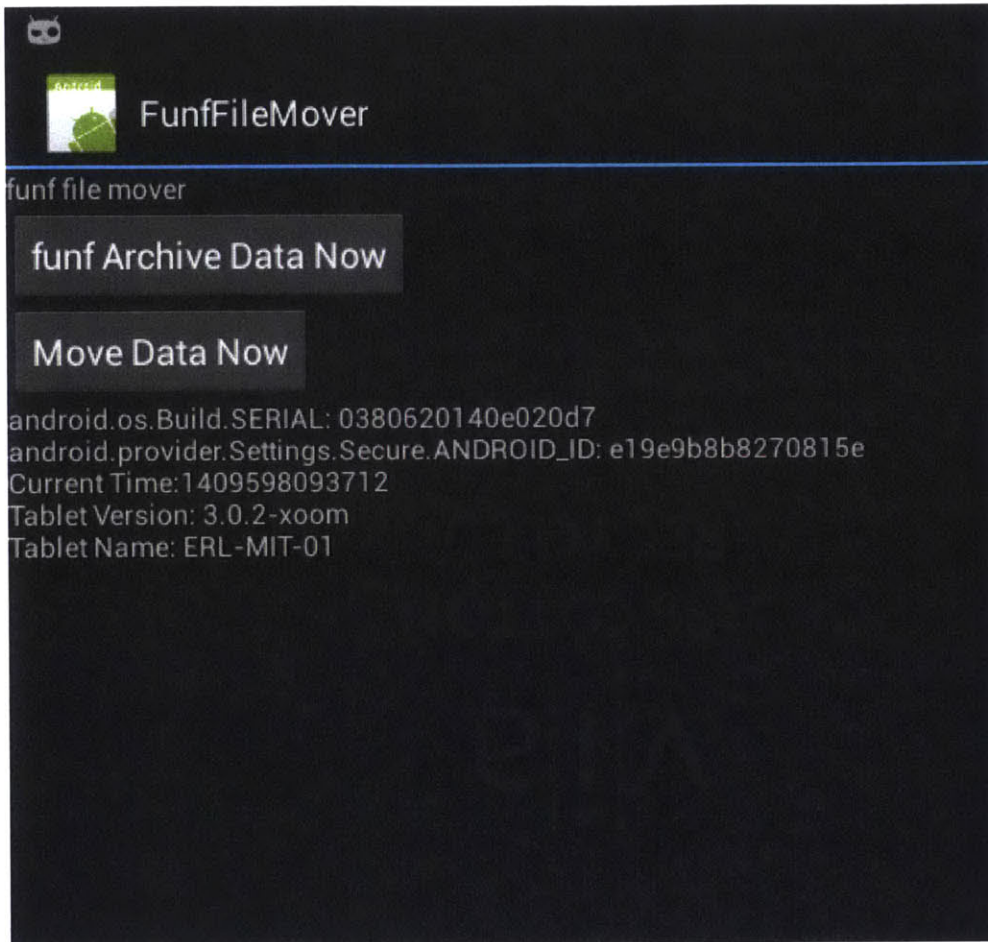


Figure 2: FunfFileMover User Interface

tablet data synchronously. The GlobalLit platform also facilitates live experiments, where data collects in real-time. This data collection can feature high fidelity probes streamed from the tablet with minimal latency. This support for immediate data collection and interaction with external systems allows a fused view of user interaction events with additional probes and sensors. The extra probes could include those that are external from the tablet (ex. motion capture systems or microphones).

As a system for data collection during live studies and experiences, this system is most effective when there is robust networking between the tablet and a host computer. However, GlobalLitLive uses the same within-tablet

messaging system as the rest of the GlobalLit platform. Data collected in the live scenario can also be queued to propagate to the data collection servers for further management and analysis.

The system for live data collection facilitates fusing additional live probes with the timestamp aligned asynchronous probes. So it can collect data from human subjects experiments or interventions where the resolution and frequency of data can modulate as dependent on specific user behavior. This flexibility allows for switching between aggregate, long-term analysis with real-time monitoring and feedback loops.

In addition, this integrated view of real-time user interaction events with additional sensors (ex. microphone streaming), can act as a diagnostic and learning aide. A remote teleoperator could observe the user's progress through literacy apps and trigger additional reading or mentoring prompts, or an expert coach could influence the reading experience on tablets in multiple homes and classrooms.

For example, as a child is tapping on a tablet, we can align finger input or app-specific states tied to literacy goals with a video and audio stream. If the system analysis triggers the need for a remote coach to engage in direct intervention, GlobalLitLive can be enabled. A feature like this would allow for altering of app behavior alongside more asynchronous mentoring where feedback is aggregated and deployed to groups of tablets over a longer term. For example, the mix of apps on a tablet could develop alongside the literacy skills deployed over the course of a semester. Individual apps could be assessed and altered, in real-time, for tutoring scenarios. Switching between live and remote mentoring scenarios would likely be imperceptible from a child's app experience.

Additionally, this lays a pipeline for creating versions of the tablet that are responsive to real-time modification from algorithmic or cloud driven computer systems. Already, the GlobalLitLive system can extract audio data from a given interaction for analysis using a crowdsourced transcription service. This thesis will describe a case study where this methodology is used to provide seamless spoken language analysis during a parent-child interaction with the tablet.

The system also includes a visualization and annotation tool for reviewing GlobalLitLive sessions. Using this software, a researcher can observe, in real time, a child using a tablet and make observations about the various probes.

4.2.0.4 GlobalLit Live Architecture

For live data collection, a host computer is required to serve as the central clocking coordinator, to receive probe data, to process and analyze incoming data, and to send messages back to the tablets.

This requirement may be prohibitive for field studies where no central system is deployable. However, a host system can provide the important added benefit of allowing for multiple tablets and devices to coordinate their data collection. The devices can synchronize along a consistent, shared timeline with additional off-device sensing.

Recall that timestamp ordering is **not** guaranteed in GlobalLitRemote. However, unlike distributed tablets in the field, where each device maintains an internal clock, the tablets using the GlobalLitLive system will all publish their messages to the same, central node. This connectivity implies a persistent network connection to a central computer through which tablets can coordinate their clocks using a system like chrony ("chrony," n.d.).

Therefore if Tablet A publishes a message with timestamp ta and Tablet B publishes an event with timestamp tb with $ta > tb$ measured in real time, then we will expect to see this order reflected in the collected data with a reasonable fidelity. In ad-hoc testing, this seems to be sufficient to maintain timestamp integrity with a ± 0.01 second accuracy.

GlobalLitLive uses globally broadcasted messages, so if the experiment requires addressing of individual tablets, a protocol should be introduced where devices publish a unique identification under which messages are subscribed and published. Currently this is supported via unique node naming provided through ROS, but identification numbers are also sent as metadata when multiple tablets publish on the same topic.

The GlobalLitLive system relies on the Android Broadcast Intent system. Event driven messages derive from a `com.android.Intent` object with a correctly formatted payload bundle. When apps can rely on this standard messaging system, they can provide a consistent, standardized source of data in the live context.

The GlobalLitLive package includes an Android service called `ROSIntentHandler`. The service is a background process that listens for GlobalLit data probe messages and relays this to the GlobalLit system with a timestamp with a timestamp localized to the `rosclock` clock.

Note that this mechanism will introduce a level of indirection between the app data probe and the processing system. The message will first be constructed and transmitted via the Android messaging system and then be processed and transmitted via GlobalLitLiv over a ROS message transport

layer. Thus, streaming data, like video, is not currently supported by GlobalLitLive tablet handlers. Videos would need a different handling since the latency inherent in the broadcast intent system would be prohibitive to high-frequency messages. Nevertheless, this data *could* be collected in real-time under a scheme in which a probe would publish the URI of a streaming TCP/IP socket, for instance.

4.2.1 GlobalLit Apps

As part of the GlobalLit platform, the thesis project includes implementation of applications that use the data collection subsystem or otherwise drive the child's experience. These serve as example approaches for third-party developers to engage with the comprehensive system.

4.2.1.1 Tinkerbook Engine

The Tinkerbook Engine is a collection of reusable components to support an interactive storytelling experience on tablets. Based on Angela Chang's Ph.D work (Chang & Resnick, 2011), the source code was completely rewritten as a high-level, deeply configurable modules on top of the Unity3D ("Unity3d," n.d.) game engine. This iteration contains additional features for data collection and prompting, and has the new name, "Tinkerbook". Figure 3 shows a typical scene construction using this engine.

This generalized framework is a platform for development of additional Tinkerbook-driven apps using this same technology, but with a variety of different content and characters. For example, third-party app developers could design a themed Tinkerbook that could focus on STEM subjects or specific early literacy skills.

4.2.1.1.1 Tinkerbook Engine Game Components

- GameManager - high-level director for gameplay and interaction flow. Handles user input (ex. taps), global game state, and manages scene changes.
- SceneManager - organizes
- Scene - represents a "page" in the interactive storybook. Tinkerbook Scenes functionally map to Unity3D's notion of scenes as a way to delineate distinct gameplay units. Tinkerbook scenes are a collection of Stanzas, TinkerGraphics, TinkerTexts, and TinkerPrompts.

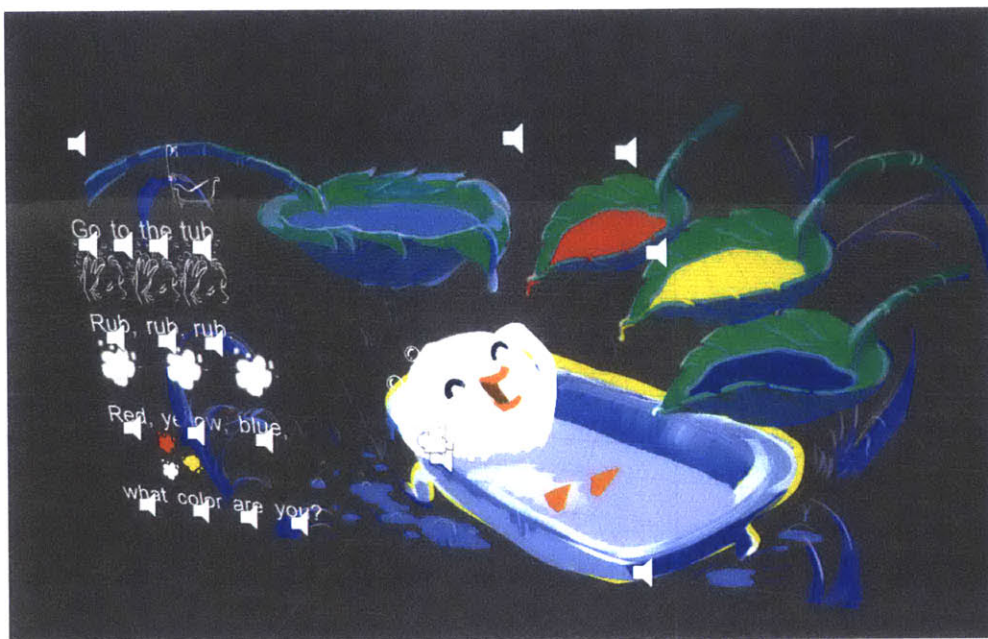


Figure 3: Tinkerbook Scene Composition in Unity3D.

- Stanza - Stanzas handle voice over playback if the user triggers a reading for the entire sentence or stanza.
- TinkerText - A TinkerText is a word or short phrase that pairs with a TinkerGraphic. It usually has an associated audio (ex. voiceover pronunciation of the word) and visual hint. Tinkerbook employs Iconimations, or short, looping images. These hints appear above words in the story to indicate their meaning (Chang & Resnick, 2011)).
- TinkerGraphic - A TinkerGraphic contains a graphic or animation that pairs with a TinkerText as a visual representation of a word or phrase.
- TinkerPrompt - A graphic or animation that serves as a narrative guide, providing mentoring prompts as the child engages in the story.

To create a new Tinkerbook, a game developer would storyboard multiple pages or scenes in the app. The storyboard defines the way scenes transition to each other and how the text, animation, sound and illustration would fit together on the “page.”

The developer would then commission illustration, animation, and audio (music, sound effects, voiceover) to incorporate into the engine. By defining the page order and flow in the SceneManager, navigation is automatically

negotiated between pages.

Scenes that require special interaction (for example, the color-mixing mini-game experience in Baby D (Chang & Resnick, 2011)) can be built using the game engine features that Unity3D provides.

4.2.1.2 Example Assessment App: Matching

A comprehensive platform for delivering literacy intervention must support assessment tools to review the progress of a child's development. Measurements derived from games, tests, and other assessment apps at various points during the child's tablet could act as literacy probes. They could provide learning data using the same carrier infrastructure as other tablet probes.

The thesis software includes an app as an example mechanism for providing assessments, including the feedback loop, as a matching game (i.e. child attempts to "put together" different concepts). No claims can be made about the efficacy of this matching app, as its validity remains untested, and no live deployments have included the app. However, one can imagine apps like this deployed into the framework to provide more insight into a child's development.

The matching app can provide different collections or categories of matching targets:

- upper case to upper case letters
- lower case to lower case letters
- upper case to lower case letters (and vice versa)
- semantic categories (ex. vehicles, food, nature)

Example screenshots from this app can be in Figures 4 and 5

In the most basic mode, a child is presented with three options and a target on "cards" on the screen. The child might tap one of the three options, and the system records the interaction (both the options presented and the child's choice). No matter what the child taps, the cards disappear and a new set appears. The "correct" answer might simply be the option that matches the target, exactly.

A more advanced level might involve the child matching upper case letters to lower case letters or to match an apple to a banana because they are both fruits.

This app listens for Broadcast Intent messages from the Android system to choose which category set to present to a child and for how many trials. This app could be presented as an interstitial experience as a child is playing

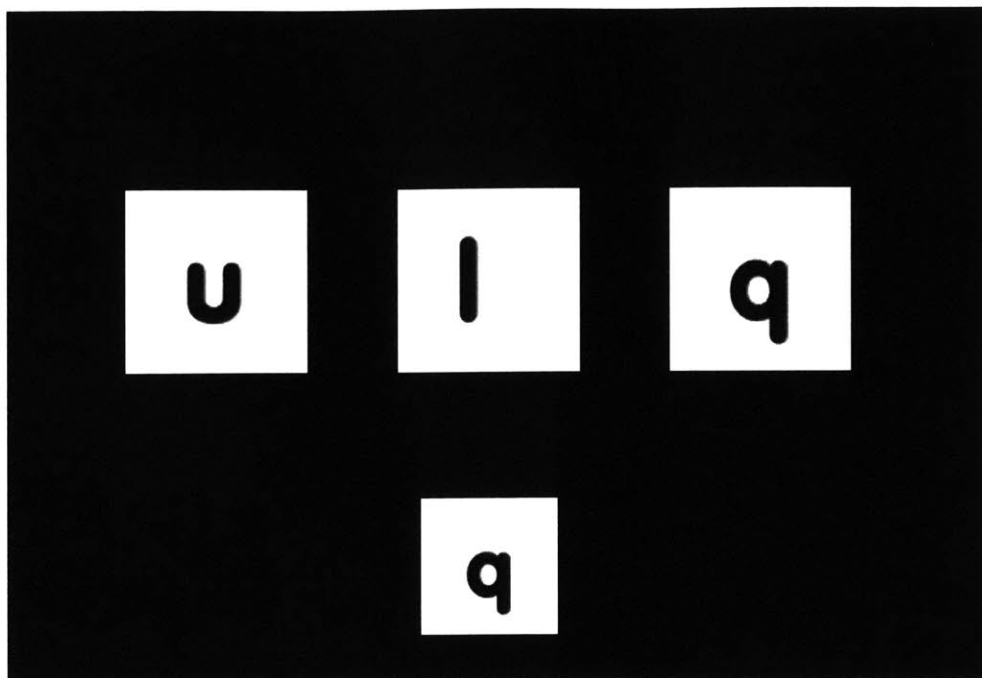


Figure 4: Matching app with lower case letter matching.

with the tablet. For example, it might appear after a child exits a game (every 20 games, for instance). It should also be a quick, reasonably fun interaction so that the child will answer some iterations and remain engaged for further play on the tablet.

Note that in order for this or any other app to serve as a rigorous assessment tool, it must first be validated and compared to traditional assessments. Also, a specific study showing its efficacy as an unsupervised measurement tool should be conducted. However, if apps like this hold potential as usable assessment tools, then future approaches could include assessments throughout deployments to measure and guide a child's attention and learning trajectory.

4.2.1.3 Launcher App

One of the key data probes for studying child driven exploration of the tablet content relies on capturing data regarding which apps children are opening, and when. In a normal Android device, all apps are available for children to interact with (including apps like web browsers and the settings

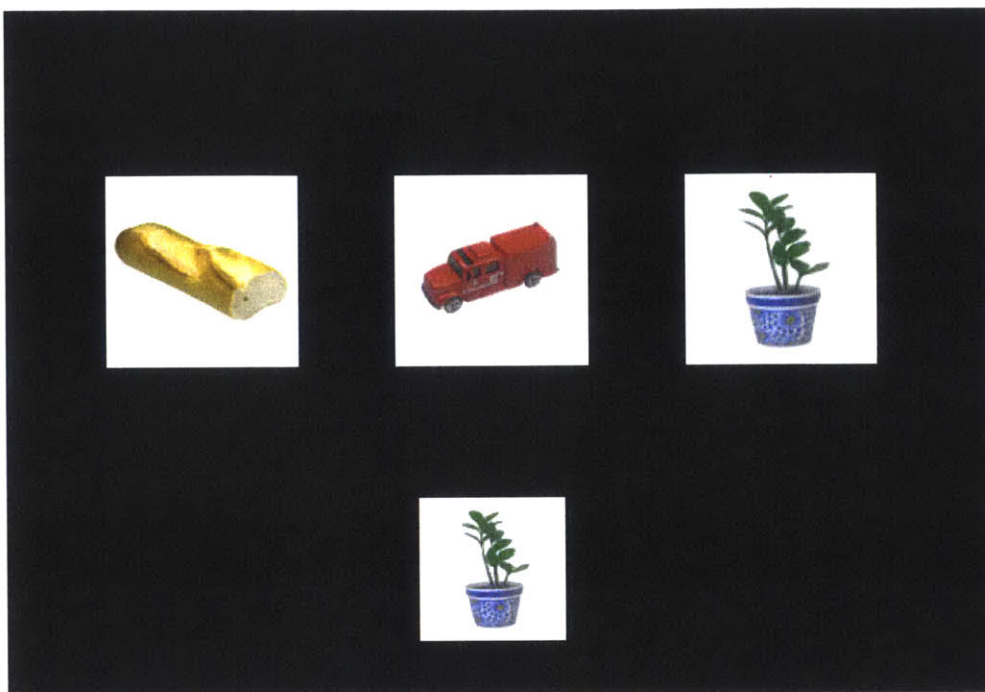


Figure 5: Matching app with semantic categories.

for the tablet). The built-in Funf probe, `RunningApplications`, could potentially provide this kind of data. However, early deployments demonstrated that it produced too much data and that children would be activating apps more frequently than could be captured by this periodic probe that sampled every few minutes.

The initial versions of the tablet allowed unfettered access to these apps, including rearranging the icons on the home screen of the tablet, changing system settings, or even installing new, unauthorized software. Children's natural curiosity and exploration extended to the tablet infrastructure, which would interfere with studies. Third-party protection software that required passwords in order to launch certain apps (Carrot App, 2010) was a mitigation attempt. Children were able to circumvent this protection by guessing the password or figuring out how to put the tablet into a safe mode with disabled app protection.

In order to attempt a more successful capturing of these app launches, deployments tests a series of launcher apps that replaced the default Android launcher (i.e. the application that presents the home screen to the user). The

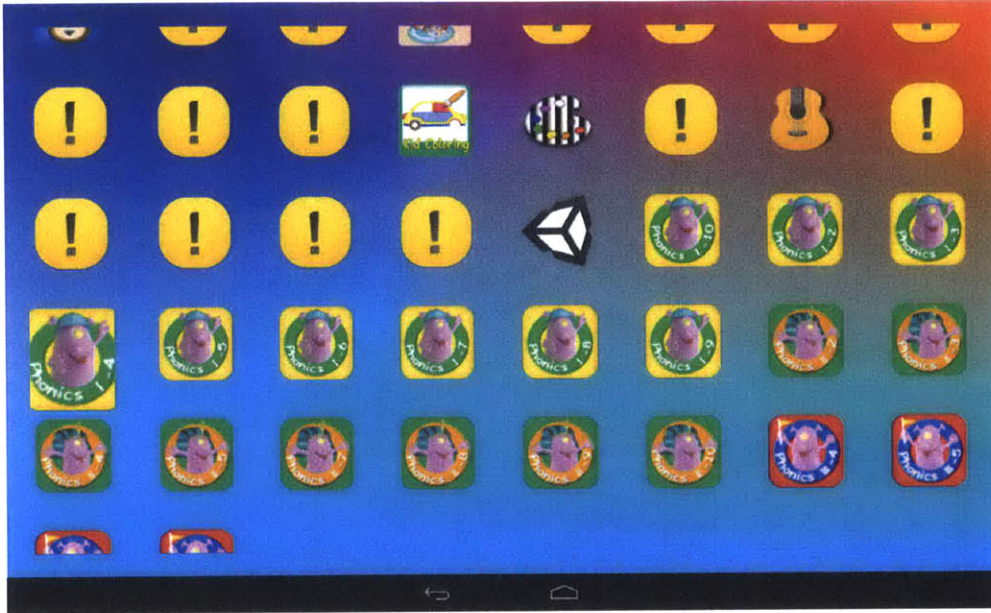


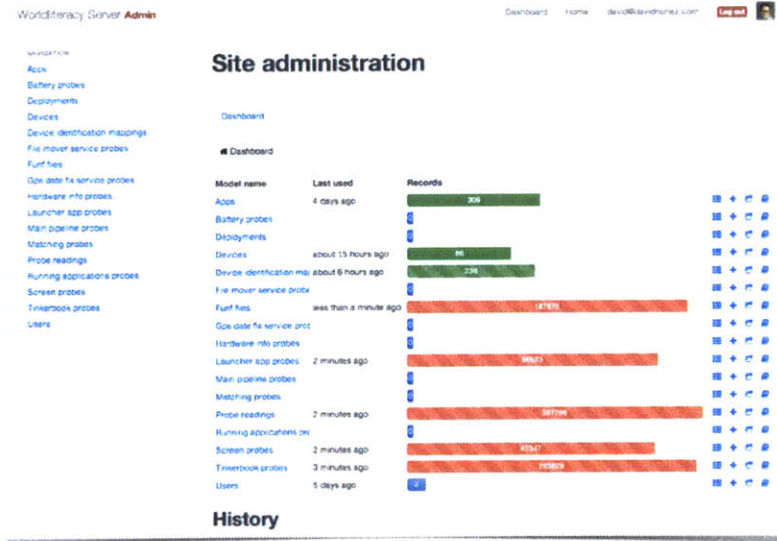
Figure 6: User interface for Launcher app.

GlobalLit launcher presents a field of icons corresponding to apps a child might launch as shown in Figure 6. Every time a child activates an app, a Funf data row is created corresponding to the launch. Likewise, when the launcher app is active (i.e. the frontmost app), another data row is created. From this data, we can try to infer the time a child spends in an app.

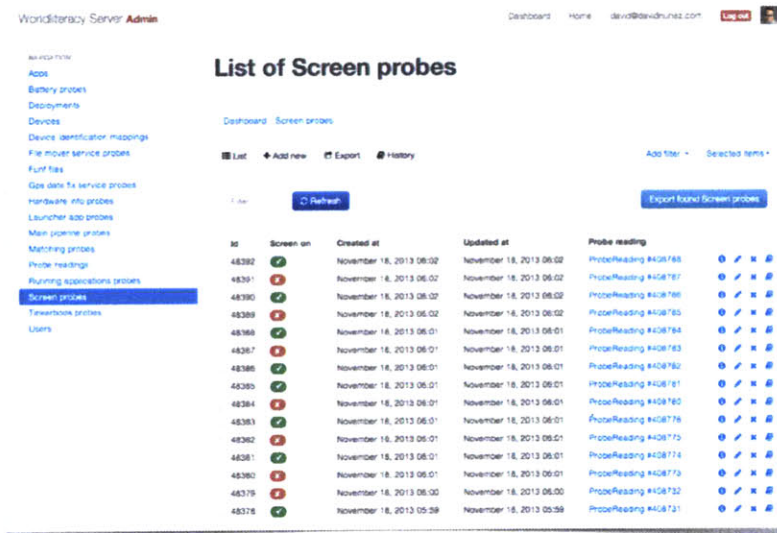
Unfortunately, this mechanism has been problematic. Children in some deployments have been able to close the launcher app and activate apps outside of the launcher, skipping the data capture. Subsequent deployments have continued to iterate on launcher apps that more comprehensively controls the tablet experience.

Deployments and individual tablets might have different requirements for the apps that the launcher presents to a child and the position of the app icons in the launcher. A file mandates this configuration, `apps.json`, that exists on the device’s internal storage (in `/sd card/launcher`), and the configuration is modifiable by external processes.

The composition of apps can change over time, simply by adjusting the manifest. For example, on deployment, the devices could have all apps installed, but only some activated in the manifest. At a designated time, a remote administrative command could be initiated (via `FunfFileMover`) to



(a) Site administration dashboard.



(b) Screen probe listing.

Figure 8: Web-based administrative interface for GlobalLit platform. It is an example of a facility for users to collect, analyze, and react to data collected from tablets in the field.

Listing 4.4: Example administrative command used to update tablet via web interface.

```
# update script 2014-04-10-1300
wget -O - http://worldliteracy.media.mit.edu/apks/2014-01/2014-01-
update.sh | sh
```

4.2.2.1 Deployment Monitoring

The interface allows for assigning tablets to deployments and monitoring of data collection. A site deployment manager can either assign the label and deployment id to the tablet before or after a tablet is deployed. Devices can be searched and sorted by deployment.

When a tablet first connects to the server via the administrative ping, the system registers it as a new device based on its unique serial number. Every time the tablet makes a connection to the server, the record for the device updates with the latest timestamp. This way, the administrative interface can be used to monitor the deployments. For example, an administrator can list the tablets at a deployment and find those that have not recently connected, highlighting potentially faulty tablets.

4.2.2.2 Probe Monitoring

Each probe has an administrative screen in the interface. From this interface, administrators can filter, search, and even modify data. Note that the system logs all changes to data so that a record and accounting exists. These probe tables (or subsets of the collections) may be exported from the interface in a variety of formats (ex. csv, JSON).

4.2.2.3 Administrative Commands

Finally, the interface allows for pushing arbitrary shell commands to the Android devices. This feature facilitates remote diagnostics, updating apps, or performs other low-level maintenance. Listing 4.4 shows an example command script used to download and execute a shell script from the server. Listing 4.5 shows commands that are used in this update to revise system software.

Listing 4.5: Administrative script used to update a tablet by removing, downloading, and adding apps to the tablet. The script also demonstrates the necessary bootstrapping and security workarounds that were necessary to provide these updates.

```
su

mkdir /sdcard/apks
cd /sdcard/apks/
wget http://worldliteracy.media.mit.edu/apks/2013-12-20-1200/com.
    goatella.beginningblends.apk
wget http://worldliteracy.media.mit.edu/apks/2013-12-20-1200/com.
    intellijoy.sightwords.apk
wget http://worldliteracy.media.mit.edu/apks
    /2013-12-20-1200/2013-12-20-1200-mentoring.apk
wget http://worldliteracy.media.mit.edu/apks
    /2013-12-20-1200/2013-12-20-1200-FunFileMover.apk
mkdir /data/tmp
cp /sdcard/apks/com.goatella.beginningblends.apk /data/tmp/
cp /sdcard/apks/com.intellijoy.sightwords.apk /data/tmp/
cp /sdcard/apks/2013-12-20-1200-mentoring.apk /data/tmp/
cp /sdcard/apks/2013-12-20-1200-FunFileMover.apk /data/tmp/

pm install /data/tmp/com.goatella.beginningblends.apk
pm install /data/tmp/com.intellijoy.sightwords.apk

pm uninstall edu.mit.media.prg.funffilemover
pm install /data/tmp/2013-12-20-1200-FunFileMover.apk

pm uninstall edu.mit.media.prg.mentoring_app
pm install /data/tmp/2013-12-20-1200-mentoring.apk
cd /data/system/users/0

echo "3.0.1-2013-12-20-1200" > /sdcard/version.txt
echo "Rebooting now..."
reboot
```

5 Case Study - GlobalLit Remote

The Global Literacy Project has deployed iterations of the GlobalLit software in a variety of conditions and usage scenarios at locations around the world. These endeavors helped to highlight some of the difficulties in conducting field research. This chapter will discuss examples of confounding factors that led to lower quality of data and some attempts to remedy these issues through a case study examining data from a sampling of these deployments.

5.1 Pilot Deployments

In 2011, early versions of tablets were sent to Sierra Leone as an exploratory pilot. We wanted to understand both the contingencies in a remote village and how children might use tablets, especially as the community may not have had much exposure to this technology. A graduate student, originally from Sierra Leone, traveled to these regions and made observations about the usage. Additionally, the devices contained data collection software (“funf | Open Sensing Framework,” n.d.) on the devices. We observed that children needed very little adult instruction to use the tablets; gesture interaction and even how to turn the tablet on in the first place were skills that were either intuitive or shared among children. However, it was clear that a more robust data collection infrastructure would be necessary to facilitate any automated, long-term study (Chang, Nunez, Roberts, Sengeh, & Breazeal, 2013).

In early 2012, we sent 41 Motorola Xoom tablet computers to two villages in Ethiopia in collaboration with One Laptop Per Child and Ethiopian government. The conditions were harsh with no nearby schools, WiFi, or even electricity. Local staff drove to the sites and manually extracted the sd cards from the tablets, aggregated the data onto flash drives which they shipped to the US via diplomatic channels. Updates to the system required an engineer to spend days in the field, manually updating each tablet via portable sd card. The Ethiopia deployment highlighted the need for creating more systematic and robust data collection (Chang, Tilahun, & Breazeal, 2014).

Subsequent deployments to locations in the United States to school districts in Georgia and Alabama allowed us to iterate on the technology more rapidly since the tablets utilized on a more stable internet network. Educators at the schools distributed the devices among kindergarten classrooms.

We had a close collaboration with the teachers, administrators, and parents at these schools; the GlobalLit system continued to evolve until the system was more usable as a science platform. A second deployment in Roanoke, starting in January 2014, is used for the analysis in this chapter.

Also in 2014, the Global Literacy Project sent collections of tablets to South Africa and Uganda using updated server software, still containing much of the core GlobalLit system on the tablets, themselves. The analysis in this chapter will explore the Uganda data.

5.2 Hypotheses

The GlobalLit Remote system allows us to ask questions about individual tablet usage (within-subject), along with questions across multiple deployments in various locations around the world (between-subject). Based on observations reported by teachers and parents from the field in pilot deployments (R. Morris, D. Crouse, personal communication, December 2013), a pattern of usage seemed to emerge. Children would initially rapidly sample from the large collection of apps available on the tablet. Over time, they would converge their attention onto a smaller collection of favorite apps and games, tending to prefer apps with engaging interaction. An attempt to examine deployment data to confirm these observations, statistically, suggests the following hypotheses:

H1: As children have more exposure to a tablet, the time they dwell on open apps increases.

H2: The pattern of increased dwell time will hold consistent between a variety of deployment scenarios.

H3: The apps with the most usage will remain consistent across deployment scenarios.

5.2.1 Study Deployments

The various deployments around the world rarely had equivalent setups to allow for meaningful comparisons (ex. same learning environment, number of children per tablet, or even same version of the software on the devices). In addition, as these deployments were using versions of the software that was changing and improving, there were few opportunities for a longitudinal study. However, there were instances where the software was reasonably stable, and the usage conditions were similar enough to each other to ask some questions across deployments. All subjects were informed of their

rights and parents provided consent as mandated by the MIT Committee on the Use of Humans as Experimental Subjects (COUHES). A summary of the conditions can be found in 4.

5.2.1.1 Roanoke

Preschool instruction can provide an important supplement to at-home language development (B A Wasik, Bond, & Hindman, 2006). Sometimes preschool teachers struggle to provide adequate personalized, individual language education to the children in their classroom (B. Wasik, 2008). Tablet computers might provide some augmentation to preschool instruction or even act as a “virtual preschool” when children do not have access to preschool options in their communities.

The Roanoke school district in Alabama began a new, progressive preschool program for children in the community in January 2014 to study the impact of pre-kindergarten intervention on literacy rates.

Table 3: Characteristics and Demographics of Roanoke

Rural - 39 people per square mile
Total student population 491
80 percent students qualify for free/reduced-cost lunch
40 percent African-American; 58 percent Caucasian
Households low-to-average SES
No previous formal preschool program in school district
Most children entering Kindergarten have not had preschool education

Parents in the community were recruited for their children’s participation in the pre-school pilot program; many parents responded with children in the appropriate ages ($n = 54$, $\bar{x} = 4.82$, $\sigma = 1.025$).

The study provided each child with a tablet (see Figures 10 and 11), and a pre-K lottery process randomly assigned him or her to one of the three conditions in which they participated for a school semester. The study lasted approximately three months, starting on January 15, 2014.

- Children in the *WHOLE DAY* condition attended the pre-school program at the school for the full day. The maximum class size was 18. They had access to the tablets as part of the daily curriculum for 20 minutes, but it was not their only activity. Each child took the tablet home for one night per week.

- Children in the *HALF DAY* condition only attended a half day pre-school program from 7:30 am to 12:00 pm with a class size of 18.
- Children in the *HOME* condition did not attend pre-school but used the tablets at home. All parents were encouraged to participate in weekly sessions with a Pre-K teacher on areas for school readiness skills. Ideally, these sessions would provide the WiFi access necessary for tablet data to synchronize with the server.

Two subjects in the *WHOLE DAY* condition withdrew. One subject in the *HALF DAY* condition was excluded because of an error in data collection. Three subjects in the *HOME* condition were excluded because their tablets were returned damaged, and data was not collected for the full semester.

5.2.1.2 Uganda

The Global Literacy Project partnered with a non-profit organization in Uganda, the Clover Foundation, to bring tablets into an elementary school. The deployment began on February 10, 2014. In this case, children had access to the tablet daily for about an hour. An attempt was made to keep tablets assigned to individual children ($n = 25$).

5.2.2 Tablet Content

The tablets contained a curated selection of educational or age-appropriate experiential apps, as assessed by Global Literacy Project collaborators from Tufts University (Wolf & Gottwald, 2013). The apps address some of the earliest skills needed to obtain pre-literacy competency. However, there can be no claims made regarding the efficacy of these apps without appropriate pre and post testing of literacy. Also, the apps chosen will not comprehensively promote the needs of an emerging reader as described in RAVE-O (Wolf et al., 2000). For example, while the collection contains many general alphabet, phonetic, and vocabulary apps. However, it is deficient in apps that directly address categories important developmental categories (ex. advanced oral language skills like negation and past tense, decoding, and advanced alphabetic principles like diphthongs and inflection). The apps included interactive games and activities alongside passive video content. Most apps were labeled by their developers as educational or designed to promote literacy. Some of these were game-oriented interactions and others were interactive electronic books. The tablets included apps that were more open-ended or focused on creative exploration. For example, apps that simulate a music instrument or



Figure 10: Roanoke deployment duplication in process.



Figure 11: Roanoke tablets ready to deploy.

Table 4: Number of Subjects per Condition in GlobalLit Remote study.

WHOLE DAY	17
HALF DAY	16
HOME	19
UGANDA	25
Total (n)	77

encourage a child to sketch pictures are considered open-ended. The video content came from a variety of third-party party producers and were usually a four to six minute animated short story. A description of all apps found on the tablet can be found in Appendix B. Note that the tablets in the UGANDA condition did not contain passive video content (ex. movies and animated, non-interactive stories) as deployment constraints made it difficult for the duplication to include this heavy content.

5.3 Data Cleaning

After a tablet uploads its data files to the processing server, the system decrypts and unpacks each file into the main database. It iterates through each row in the sqlite3 data file and transcribes each of these lines into the table as a `funf_data_row`. Even before each row expands into probe specific tables, we can run general data file analysis to determine the breadth of data coverage per deployment and attempt to determine data discrepancy. By examining the timestamp the data rows, we can determine which days we successfully received and processed data from each tablet. However, the system does not have an explicit measure of duration for the time any given app is open. Because we are interested in how long children are using the tablet and apps over time, a combination of probes helped to estimate the durations. In addition, the data needed some manipulation and cleaning to be useful for the study.

Three probes could provide the necessary information about app usage.

1. `LauncherApp` - indicates when a given app is launched (Listing 5.1)
2. `LauncherAppPaused` - indicates when the main launcher screen is activated or deactivated (i.e. another app launches). (Listing 5.2)
3. `ScreenProbe` - indicates when the tablet screen turns on or off. (Listing 5.3)

Listing 5.1: `LauncherApp` probe data.

```
name,timestamp,funf_file_id,value,device_id,id,data_uuid
LauncherApp,2013-04-26 14:47:29,1826760,com.zoodles.book.
theelvesandtheshoemaker,43,12891230,67495740-08d5-4432-b4eb-
ab58f4a610d0_4
```

Listing 5.2: `LauncherAppPaused` probe data.

```

name,timestamp,funf_file_id,value,paused,data_uuid,id,device_id
LauncherAppPaused,2014-07-22 07:22:55,1876656,"{"paused": true}",
True,c926a820-a6c6-4e52-a6c3-87133e94cd4a_5,14239263,1

```

Listing 5.3: ScreenProbe probe data.

```

name,timestamp,funf_file_id,value,screen_on,funf_timestamp,data_uuid
,id,device_id
edu.mit.media.funf.probe.builtin.ScreenProbe,2013-04-13
09:12:12,1824858,"{
  ""TIMESTAMP"": 1365858732,
  ""PROBE"": ""edu.mit.media.funf.probe.builtin.ScreenProbe"",
  ""SCREEN_ON"": true
}"

```

The events in this probe were merged and sorted into a continuous timeline of app launch events. We assume that the timestamps on the tablets all refer to the same reference date within a given deployment and among deployments. The merged data set is an approximate time series log of all tablet activity (i.e. app launches, Launcher app pauses, and screen on and off) across the study. There were many duplicate data lines (i.e. same event, with identical timestamps, recorded multiple times), and the analysis simply removed these redundant data points. The analysis assumes that two subsequent LauncherApp events of the same app refers to the same launch event and removes the second log entry. This duplication happened only occasionally in the timeline, and the events were no more than a second or two apart. In addition, the analysis removed LauncherAppPaused probes where the value was “True” (i.e. the Launcher app paused, and a different app launched). Those events were redundant alongside the LauncherApp probe events. A sample from the scrubbed log is in Listing 5.4.

Listing 5.4: Merged event log from remote field data.

```

timestamp,name,device_id,event
2014-01-16 12:34:34,LauncherApp,129,com.dornbachs.zebra
2014-01-16 12:34:41,LauncherAppPaused,129,False
2014-01-16 12:34:44,LauncherApp,129,russh.toddler.game
2014-01-16 12:34:54,LauncherAppPaused,129,False
2014-01-16 12:35:05,LauncherApp,129,BMA_CO.Phonics_Lv1_Unit6
2014-01-16 12:35:42,LauncherAppPaused,129,False
2014-01-16 12:36:56,edu.mit.media.funf.probe.builtin.ScreenProbe
,129,False

```

Data was further partitioned by condition as mapped from the device id. Given the various start dates between the conditions, timestamp deltas from the initial date helped to normalize the deployment dates. These normalized timestamps indicated which week numbers were assigned to the data lines (ex. "Week 1", "Week 2", "Week 3", "Week 4").

Determine the duration involves scanning the event log per tablet. When there is a LauncherApp event, there is an assumption the child launched the app at that timestamp. In addition, there is an assumption that the next event in the log after an app launch will indicate the end of that particular app use. The LauncherApp might be unpaused (indicating the child returned to the home screen), or the tablet screen is turned off. As a result, the difference in timestamps between a LauncherApp event and the following event is an approximation of the duration of the app use, recorded in seconds.

There were a small number of extreme outliers in the event log. For example, app durations that seemed to last for dozens of hours. It was unlikely a child would use the same app for that long a period, so these were likely errors in data collection. For the approximately 400 out of 49754 data points where app duration was longer than 30 minutes, the duration rounded down to 30 minutes.

5.4 Results

5.4.1 Dwell Time

A one-way repeated measures ANOVA was conducted to determine whether there were statistically significant differences in the average duration of app use over the course of the first four weeks for each of the deployments.

5.4.2 Roanoke Full Day

For the Roanoke FULL DAY deployment (see Figure 12), as assessed by inspection of a boxplot, there was high outlier in Week 1 and two high outliers in Week 2. The analysis of the data with and without the outliers did not indicate a significant impact on the overall result, so they remained. Duration was normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$). Mauchly's test of sphericity indicated that the assumption of sphericity had not been violated, $\chi^2(2) = 8.069$, $p = .153$.

Duration was statistically significantly different at the different time points, $F(3, 45) = 2.954$, $p = .042$, partial $\eta^2 = .165$. Duration increased from 146.6115 ± 37.51433 seconds in week 1 to 177.5926 ± 58.07165 in week 2 to 184.7422

± 41.19269 in week 3 to 204.4089 ± 73.34511 . A planned contrast showed there was a statistically significant increase in duration from week 1 to week 4, a mean difference of 57.797, 95% CI[20.2, 95.39] seconds, $p = .005$, partial $\eta^2 = .417$. This result appears to support H1, that dwell time on individual apps increases as children have more exposure to the tablet.

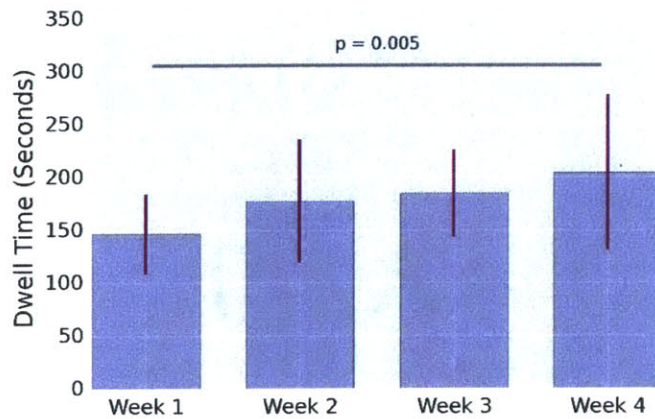


Figure 12: Mean and 95% confidence intervals for dwell time in Roanoke FULL DAY condition. Analysis by one-way repeated measures ANOVA $F(3, 45) = 2.954, p = .042$. Annotation indicates contrast significance.

5.4.3 Roanoke Half Day

For the Roanoke HALF DAY deployment (see Figure 13), as assessed by inspection of a boxplot, there were two high outliers in Week 3 and one high outlier in Week 4. There was not a significant impact on the overall result with or without the outliers, so they remained. Duration was normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$). Mauchly's test of sphericity indicated that the assumption of sphericity had not been violated, $\chi^2(2) = 8.467, p = .133$.

There were no statistically significant differences at the different time points, $F(3, 42) = .868, p = .465$, partial $\eta^2 = .058$. Duration increased from 161.0515 ± 36.58639 seconds in week 1 to 168.4995 ± 65.21650 in week 2, but decreased to 155.6739 ± 47.81246 in week 3 and further decreased to 144.1509 ± 43.43065 . This result does not seem to support H1. One speculation could be that as children have less time at the preschool to use tablets, they and the teachers prioritize other activities rather than spending extended focused time in apps.

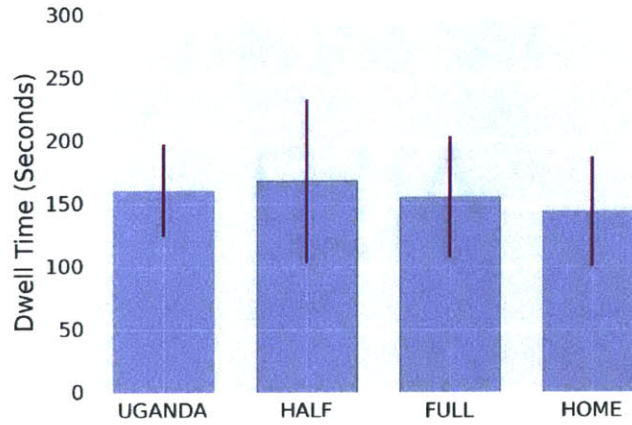


Figure 13: Mean and 95% confidence intervals for dwell time in Roanoke HALF DAY condition. Analysis by one-way repeated measures ANOVA $F(3, 45) = 1.294, p = .288$.

5.4.4 Roanoke Home

The Roanoke at-home condition (see Figure 14) only had usable data for the first three weeks.

As assessed by inspection of a boxplot, there was one high outlier in Week 3. Data analysis with and without the outlier did not demonstrate a significant impact on the overall result, so it remained. Duration was normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$). The assumption of sphericity was not met, as assessed by Mauchly's test of sphericity, $\chi^2(2) = 8.912, p = .012$. Greenhouse & Geisser (1959), was used to correct the one-way repeated measures ANOVA to compensate, with $\epsilon = .513$.

There were no statistically significant differences at the different time points, $F(1.026, 4.105) = .788, p = .430$, partial $\eta^2 = .163$. Duration increased from 205.9007 ± 77.32375 seconds in week 1 to 255.7503 ± 152.27185 in week 2 to 523.3840 ± 724.43682 in week 3. Since there was an increase in dwell time, this result seems to support H1 but does not hold significance. More telling is the dramatic drop in use altogether in tablets after week 3. Likely, children either simply stopped using the tablets when they had so much access at home or the data did not properly transfer at the end of the study; this is a potential deficit in study implementation.

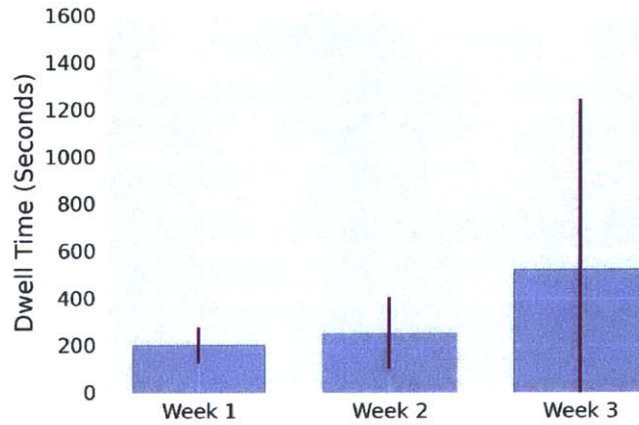


Figure 14: Mean and 95% confidence intervals for dwell time in Roanoke HOME condition. Analysis by one-way repeated measures ANOVA $F(1.026, 4.105) = .788, p = .430$.

5.4.4.1 Uganda

For the UGANDA deployment (see Figure 15), as assessed by inspection of a boxplot, there was one high outlier in Week 2. Analysis with and without this outlier demonstrated no significant impact on the overall result, so it remained. Duration was normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$). Mauchly's test of sphericity indicated that the assumption of sphericity had not been violated, $\chi^2(2) = 5.119, p = .402$.

Duration was statistically significantly different at the different time points, $F(3, 60) = 15.960, p < .001$, partial $\eta^2 = .444$. Duration increased from 85.1121 ± 20.78423 seconds in week 1 to 103.8184 ± 25.51005 in week 2 to 116.2030 ± 37.87527 in week 3 to 130.2672 ± 28.21125 . A planned contrast showed there was a statistically significant increase in duration from week 1 to week 4, a mean difference of 45.155, 95% CI[29.5, 60.76] seconds, $p < .001$, partial $\eta^2 = .645$. This finding supports H1 since the children dwelled on individual apps longer as the weeks progressed. Furthermore, this result and the results from the Roanoke FULL DAY and HOME conditions partially support H2. These differing deployment scenarios all seemed to have increases in dwell time over the course of the study, but not all increases were significant. The Roanoke HALF DAY condition seemed to show a decrease in dwell time, although it was a statistically insignificant difference. Nevertheless, since this contradicts H2.

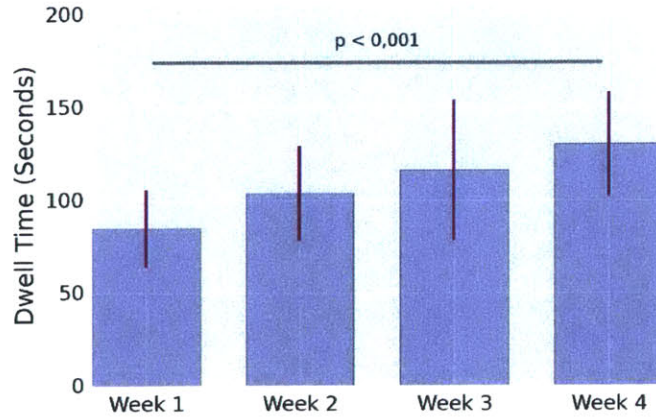


Figure 15: Mean and 95% confidence intervals for dwell time in UGANDA condition. Analysis by one-way repeated measures ANOVA $F(3, 60) = 15.960, p < .001$. Annotation indicates contrast significance.

5.4.5 Cross Deployment Dwell Time Analysis

A one-way Welch ANOVA was conducted to determine if the mean dwell time for apps on tablets was different among the deployments (see Figure 16). A boxplot revealed two high outliers in the Roanoke HOME condition for week 3. Closer inspection suggested removing the data from those tablets as they only seemed to represent a few very long app sessions over the course of the three weeks. An assumption is these were data collection errors or otherwise unusual usage patterns (ex. children stopped using the tablets after week two). Otherwise, Duration was normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$). However, the assumption of homogeneity of variances was violated, as assessed by Levene's test for equality of variances ($p < .005$).

Dwell time was statistically significant for different conditions, Welch's $F(3, 33.455) = 34.607, p < 0.0005$. Data is presented as mean \pm standard deviation. Mean app dwell time increased from the UGANDA ($n = 25, 99.3071 \pm 22.45547$), to Roanoke HALF DAY ($n = 17, 158.8160 \pm 31.12340$), to Roanoke FULL DAY ($n = 16, 166.7110 \pm 22.27814$) to Roanoke HOME ($n = 16, 201.2216 \pm 147.80516$) conditions, in that order. Games-Howell post hoc analysis revealed that the increase from UGANDA to Roanoke FULL DAY (67.404, 95% CI (48.029 to 86.778)) was statistically significant ($p < .0005$), as well as the increase from UGANDA to Roanoke HALF DAY (59.509, 95% CI (35.476 to

83.542), $p < .0006$). No other differences were significant. More ethnographic study would be necessary to identify differences between the conditions that might explain the differences in dwell time.

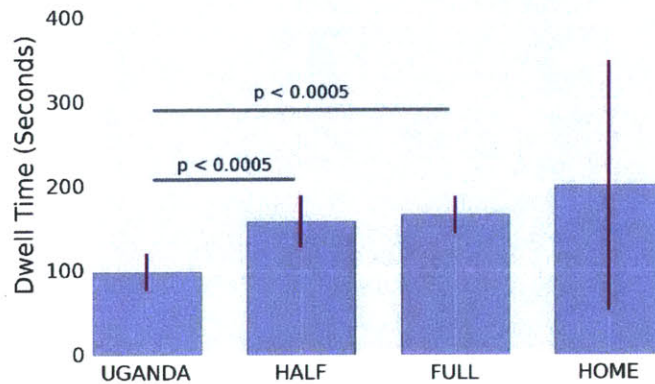


Figure 16: Mean and 95% confidence intervals for dwell time in UGANDA, Roanoke HALF DAY, Roanoke FULL DAY, and Roanoke HOME conditions. Analysis by one-way Welch ANOVA $F(3, 33.455) = 34.607, p < 0.0005$. Annotation indicates contrast significance.

5.4.6 Most Used Apps

An observational study of the most used apps in each condition yields reflections about the pattern of exploration. The total dwell time for each app was assessed as a percentage of all dwell time for all apps per condition per week. The analysis ranked the usage percentages in descending order to determine the top ten apps per condition per week. The visualizations in Figures 17, 18, 19, and 20 show the top ten apps used per week across tablets, their relative dwell times, and indicate how apps change in rank over the course of the weeks of deployment. Recall that the UGANDA condition did not contain video content, so no observations about videos could be made there, but otherwise some observations about interactive apps are possible.

5.4.6.1 Video vs. App Content

The tablets in the Roanoke WHOLE DAY, HALF DAY and HOME conditions all seemed to favor interactive apps in the top ten most used apps during

at least the first two weeks. In the second two weeks, the WHOLE DAY and HALF DAY conditions seemed to maintain interactive content in the top ten apps. However, in the Roanoke HOME condition, weeks 3 and 4 showed increasing numbers of video content appearing in the Top Ten. Indeed, in Week 4, eight out of the top ten were video content. This observation might indicate that in the home setting, without the supervision or structure introduced during daytime pre-school activities, children might revert to more passive “tv-watching” behaviors.

5.4.6.2 Most Used Videos

In all three Roanoke conditions, the video “Blue Goes to School 1”, a story about a dog’s first days at pre-school (Koyalee Chanda, Bruce Caines, Elizabeth Holder, Jonathan Judge, Nancy Keegan (Directors), Nickelodeon (Producer), 2003), appeared consistently in the top ten. In the two pre-school conditions, it was the top video. A speculative explanation could be that children watched this video as they empathized with the dog’s initial experiences in a school setting.

5.4.6.3 Most Used Interactive Apps

In all three Roanoke conditions and Uganda, the app “Monkey’s Preschool Lunchbox” was in the top three apps for at least three weeks and in the top 6 for all four weeks. This app especially encourages multiple, diverse playful interactions. It includes a vibrant monkey character and a variety of games to play. Children collect virtual stickers as they continue to use the app, encouraging repeat play, perhaps.

Additionally, ZebraPaint (Dornbach, 2003), Drawing Pad (Darren Murtha Design, 2003), and Kid Coloring (Luyen, 2003) all ranked highly among the conditions. These apps are all open-ended creativity apps, inspiring free sketching and creation.

Collectively, these observations seem somewhat to support H3 (i.e. the top apps will be similar across multiple deployment scenarios). Particular apps seem to have more engaging interaction design for children than other apps. However, the top ten lists are not identical across the deployments, so there is still room to understand what universalities exist about app engagement. In particular, an understanding of how children first discover apps and how sharing occurs among communities of tablet users may help explain convergence on the popular apps.

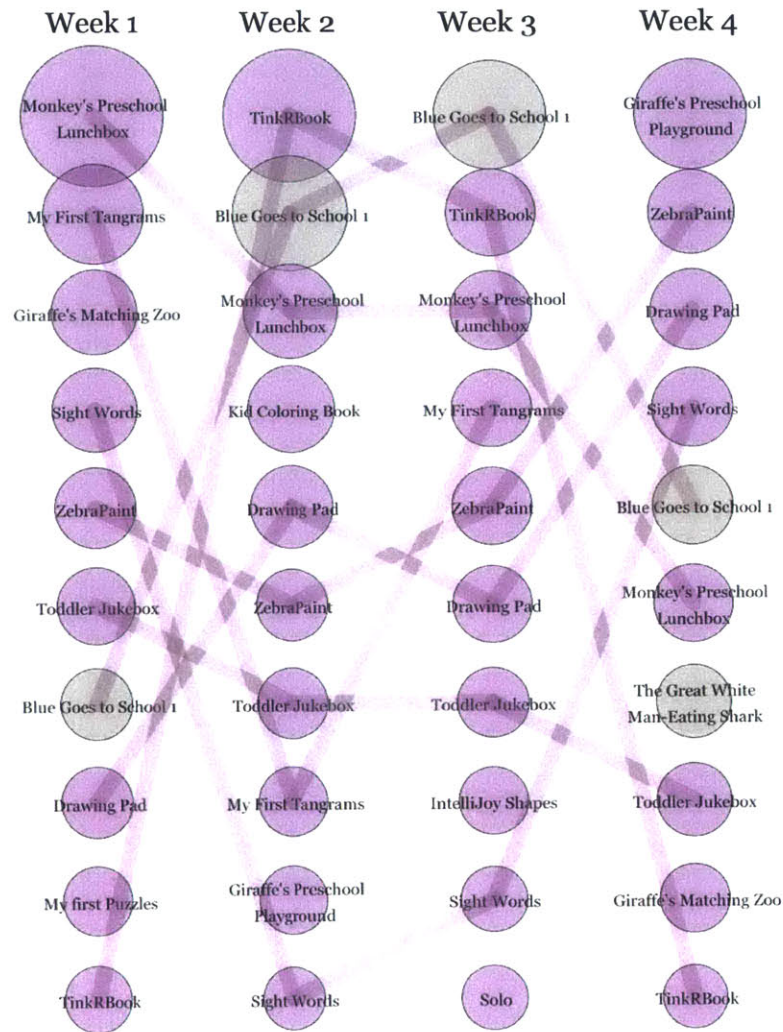


Figure 17: Top ten used apps in Roanoke WHOLE DAY deployment by week. Labels in circles indicate app identification number. Size of circles indicates usage percentage of that app compared to all others.

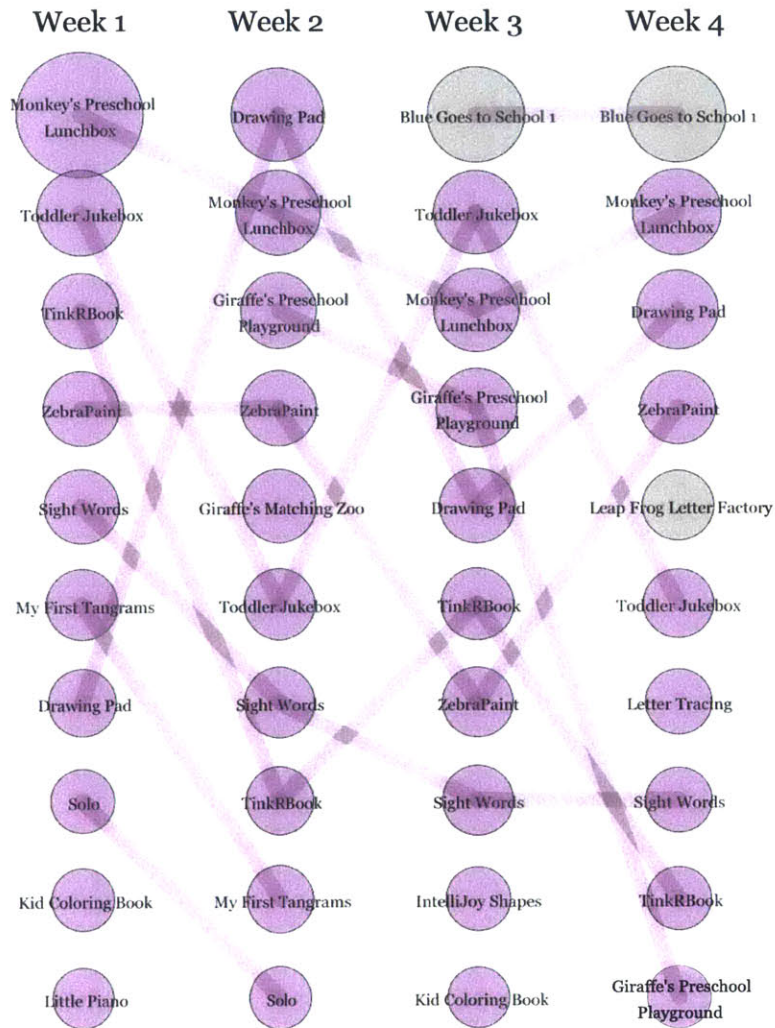


Figure 18: Top ten used apps in Roanoke HALF DAY deployment by week. Labels in circles indicate app identification number. Size of circles indicates usage percentage of that app compared to all others. Grey indicates video content, purple indicates interactive apps.

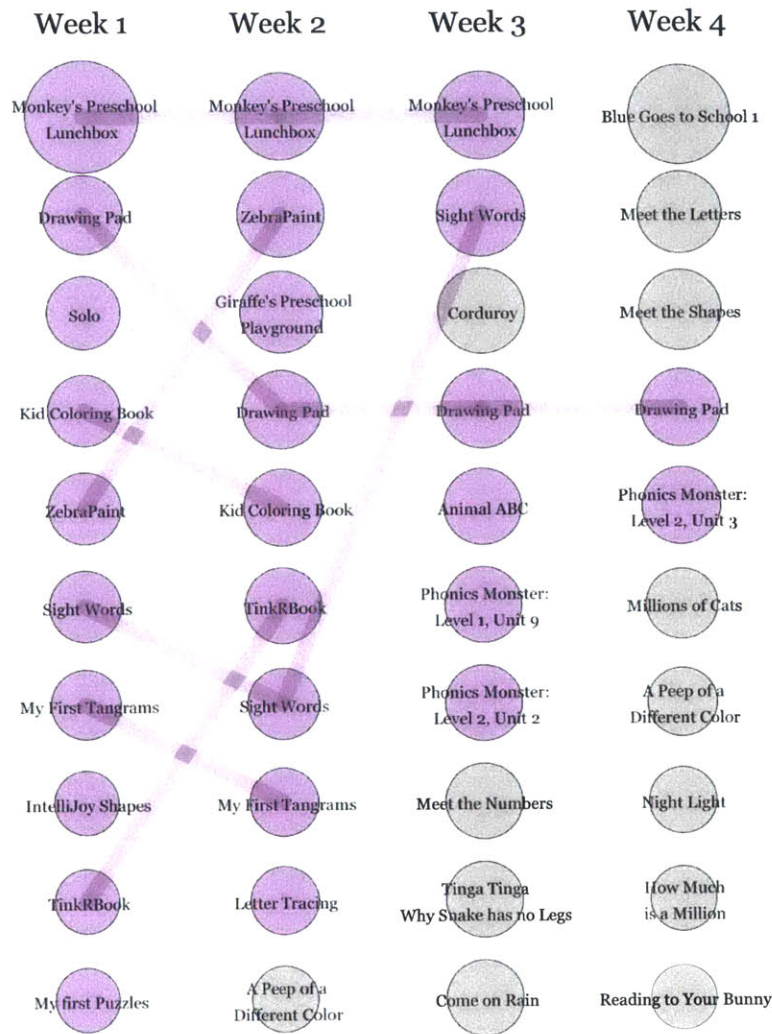


Figure 19: Top ten used apps in Roanoke HOME deployment by week. Labels in circles indicate app identification number. Size of circles indicates usage percentage of that app compared to all others. Grey indicates video content, purple indicates interactive apps.

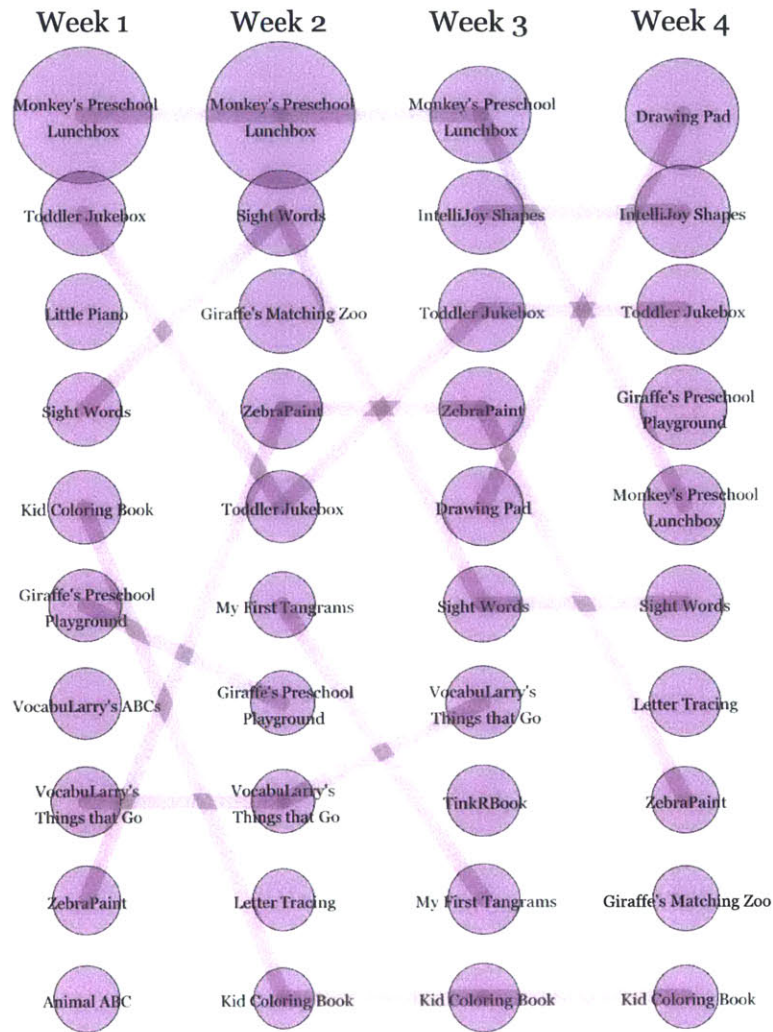


Figure 20: Top ten used apps in Uganda deployment by week. Labels in circles indicate app identification number. Size of circles indicates usage percentage of that app compared to all others. Grey indicates video content, purple indicates interactive apps.

5.5 Discussion

This chapter's case study examined app usage data across multiple deployment scenarios. It demonstrates a few ways to examine the data from tablets within and between deployments. For example, dwell time on apps seemed to increase over time. That is, as children use tablets over the course of a few weeks, it appears that children spend more time on any given app. The case study found the top ten apps per deployment and how the app rankings changed over the weeks. From this, the analysis provided observations about what types of apps were popular. An exploration of Dwell time between deployments revealed some differences in the overall time children spent in app session in the various site configurations.

However, it is important to put the speculative results obtained from purely observational examination of this data in perspective. Patterns may occur in the data, but without ground truth understanding of the context in a child's classroom or home, it can be difficult to infer meaning from the data. Furthermore, without literacy assessment, no claims are possible regarding if these apps have any impact on reading. Other studies using this data or future deployment data should include pre and post test measurements and controlled conditions to see if the apps have an impact. Since the interaction effects between the apps are complex, it may be useful to create constrained validations of individual apps. To determine apps' efficacy in isolation or carefully constructed content exposure sequences, researchers will need to understand the particulars of the apps in great detail. Future scenarios might include children having evolving access to subsets of the apps based on a timeline or gated by particular user interactions. If the tablet includes learning assessment apps, then those apps could be used to modulate which apps are available to children. The system could act in a mentoring capacity, helping to construct scaffolding for child-driven learning.

6 Case Study - GlobalLit Live

This thesis demonstrates the viability of the GlobalLit Live system as a useful platform for studies of children and tablet computers through an experiment at MIT using Tinkerbook and GlobalLit Live. The study examines if a virtual mentoring character, introduced in an interactive picture book, could encourage engagement and positive dialogic reading behavior between parents and children.

6.1 Introduction and Background

Parents can have a strong influence on their children's education and are an integral component of a multi-faceted approach to developing reading skills (see (Reese, Sparks, & Leyva, 2010) for a recent survey of parent-child interventions). For example, vocabulary, oral language complexity, and narrative skill development all seem to benefit when parents and their children read together (Bus, IJzendoorn, & Pellegrini, 1995). Co-reading experiences between parents and very young children (1-3 years old) point to more easily developed literacy skills in the long term, especially if this co-reading occurs frequently and consistently (Wells, 1985). Reading in a manner that captures a child's attention, and imagination can promote emergent literacy and language development while encouraging a lifelong love of reading (Duursma, Augustyn, & Zuckerman, 2008).

6.1.1 Dialogic Reading

Dialogic reading is a technique introduced by Whitehurst et al. (G J Whitehurst, Falco, & Lonigan, 1988, Grover J Whitehurst, Arnold, Epstein, Angell, & al (1994), C. J. Lonigan & Whitehurst (1998)) that uses conversation between adults and children about stories and picture books as a means to facilitate language development and emergent literacy. Parents are trained in techniques that encourage active participation by children in the reading experience through a processes of asking open-ended questions and connecting the story to the child's own interests and actual events in the child's daily life (Reese et al., 2010).

This scaffolding introduces specific prompting methods that parents might use with their children as part of the reading experience.

1. Completion prompts: Fill-in-the-blank questions [...]
2. Recall prompts: Questions that require the child to remember aspects of the book [...]
3. Open-ended prompts: Statements that encourage the child to respond to the book in his or her own words [...]
4. Wh-prompts: What, where, and why questions [...]
5. Distancing prompts: Questions that require the child to relate the content of the book to aspects of life outside the book. [...] (Zevenbergen, Whitehurst, Van Kleeck, Stahl, & Bauer, 2003)

Parents can learn how to engage in dialogic reading techniques through live training programs, where an instructor models and teaches best practices, or via offline methods like video recorded demonstrations and tutorials (Blom-Hoffman, O'Neil-Pirozzi, Volpe, Cutting, & Bissinger, 2007; J. Cutting, 2011). When parents receive coaching in dialogic reading, there is a strong impact on parent-child reading behaviors and parental attitudes to joint storybook reading (Pillinger & Wood, 2014). However, parents of at-risk children (i.e. those from lower-income families with higher rates of illiteracy) do not have as effective an impact on their children when attempting dialogic reading techniques (Mol, Bus, Jong, & Smeets, 2008). It is important to note that parents with less literacy or education were better able to acquire dialogic reading skills through direct training rather than video instruction (Huebner & Meltzoff, 2005).

6.1.2 Dialogic reading with Apps

An initial investigation showed parent-child dialogic reading and story comprehension were negatively affected by the interactive features in an electronic book (Parish-Morris, Mahajan, Hirsh-Pasek, Golinkoff, & Collins, 2013). However, the authors also suggested there might be ways to improve the design and implementation of these apps to improve their impact.

In addition, enhanced interactivity in electronic storybooks might negatively impact story recall but improve engagement and physical interaction (C. Chiong, Ree, Takeuchi, & Erickson, 2012).

Miller and Warschauer suggest that there are many unanswered questions about digital media and its impact on literacy development, particularly with

such new technology as mobile devices. They suggest focusing research on studying more fine-grained, specific features of interactive books and their affordances and efficacy in improving literacy development before declaring a verdict about their utility (E. B. Miller & Warschauer, 2013).

There are some cases where technology might be a necessity or provide enhanced learning tools (ex. in areas where parents are illiterate, themselves, or access to school is limited). So it is worth further exploration about how best to use these tools. Indeed, the dialogic reading technique has been shown to be possible using tablet-based storybooks even over long distances (Raffle et al., 2010) or when the dialogic prompts were delivered by a virtual character (Mori et al., 2011).

Technology might even facilitate new forms of learning through parent-child reading intervention. Interactive stories could provide affordances that are simply not available in physical books, such as those additional modalities categorized by Kucirkova: audio representation, visual representation, touch screen, interactivity, customization, personalization (Kucirkova, 2013). Tablet-based storybooks might, inherently, promote engagement and parent-child interaction if their design encourages exploration and discussion (Alonso, Chang, & Breazeal, 2011). TinkrBook provided interaction design techniques and mechanisms that encouraged interactive storytelling during shared storytelling sessions (Chang & Resnick, 2011). Whether these can be viable as generalized tools for enhancing literacy is an ongoing question.

6.2 Experiment

Dialogic reading is a powerful technique, but for parents of at-risk children, training in the methodology can be more involved and might not scale well (Mol et al., 2008). For this reason, an experiment was conducted to test whether a virtual character could demonstrate prompting techniques from dialogic reading in a storybook app. If this approach shows promise, then it could potentially be used as an intervention to encourage positive reading behaviors. This experiment is also an example case study for the GlobalLit Live system.

6.2.1 Hypotheses

This experiment proposes a design feature for a tablet-based, story reading experience between parents and children. Will a virtual character that models dialogic prompts increase engagement and promote positive dialogic reading? This experiment explores the following hypotheses.

H1 - Engagement: *Parents and children will have increased engagement with tablet reading experiences in the presence of a virtual character demonstrating dialogic reading techniques.*

Engagement in activities relates to motivation and persistence in learning activities (Haugland, 1999). Engagement increases using dialogic reading techniques. Tablets can, inherently, also promote engagement in reading activities (Couse & Chen, 2010). Dialogic techniques, role modeled by a virtual character in a storybook app on a tablet might also correlate with increased engagement.

H2 - Vocabulary: *Parents and children will generate more oral language of a more diverse nature in the presence of a virtual character demonstrating dialogic reading techniques.*

Dialogic reading can encourage more rich verbal interactions among family reading experiences (Brannon & Dauksas, 2012). The hypothesis predicts similar impacts on verbal interactions when a reading app demonstrates dialogic reading prompts compared to an app without these prompts.

H3 - Dialogic Reading: *Parents and children will exhibit conversation more consistent with dialogic reading techniques in the presence of a virtual character demonstrating dialogic reading techniques.*

Parish-Morris speculated that the inclusion of distancing prompts in e-books could encourage parent-child dialogic reading behaviors (Parish-Morris et al., 2013), so this experiment is an attempt to test that hypothesis and demonstrate this effect.

H4 - Crowd Sourced Prompts: *Dialogic reading prompting questions generated through crowd-sourcing algorithms can be as effective as questions crafted by a literacy expert when demonstrated by a virtual character.*

It may be impractical to craft precise dialogic prompts for every potential tablet reading experience; stories, cultural settings, and a child's background will vary wildly around the world. For this reason, this experiment also explores computationally generated dialogic prompts, in an effort to validate the potential for automated approaches.

6.2.2 Study Task & System

6.2.2.1 Tinkerbook App

This experiment used a tablet computer (Motorola XOOM) running an interactive storybook app based on TinkrBook (Chang & Resnick, 2011), a story about a day in the life of a baby duck. Readers of the book can page forward and back through the story pages or “scenes.” The book app contains 25 scenes, most of which have interactive elements. Sample interactions are illustrated in Figure 21 and also include:

- The words are spoken aloud when tapped.
- Words and graphics animated when user taps graphics.
- Iconimation animation hints appear when user taps certain words.
- Highly interactive scenes have enhanced play modes (ex. color mixing bath scene, feeding scene, singing scene).
- Some scenes auto-advance after a certain amount of time has passed, while other scenes require a user interaction before advancing is possible.

6.2.2.2 Prompting Character

The Tinkerbook can be optionally configured so a prompting character (a butterfly) (see Figure 22) appears in certain scenes in the story. All scenes in Tinkerbook with prompting scenes highlighted in Figure 23. When the reader gets to a scene containing the prompt, the butterfly animation appears on the page and moves across the top of the screen. If the reader taps on the butterfly character, the Tinkerbook optionally plays a spoken audio prompt.

To prevent voice inflection confounding the potential impact of the prompts, a computer speech synthesizer from Google generated the audio for these prompts (“Google Translate Web Application,” n.d.).

Listing 6.1: Audio Generation Script

```
#!/bin/bash
# translate.sh filename "text to convert to audio"
wget -q -U Mozilla -O "$1.mp3" \
    "http://translate.google.com/translate_tts?ie=UTF-8&tl=en&q=$2"
```

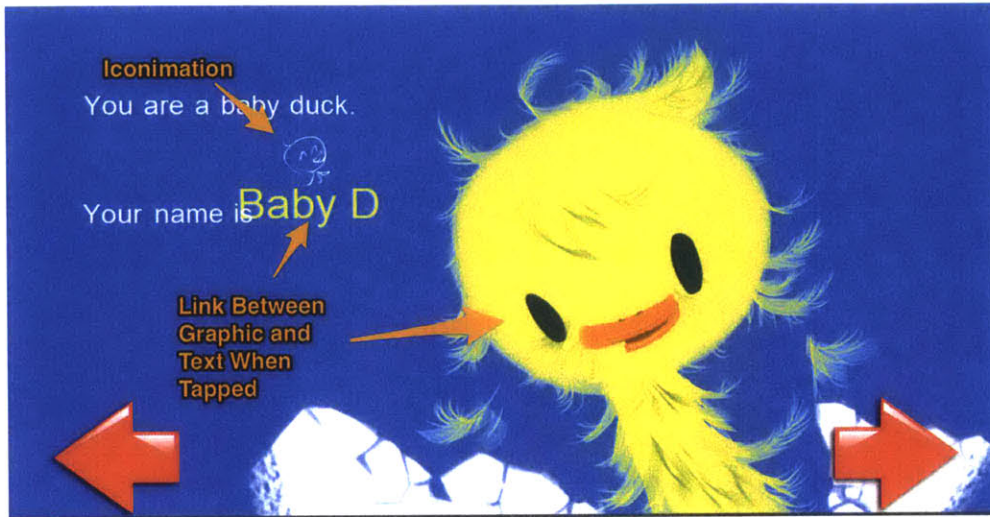


Figure 21: Sample Tinkerbook interactions include Iconimation hints and word-graphic highlighting.

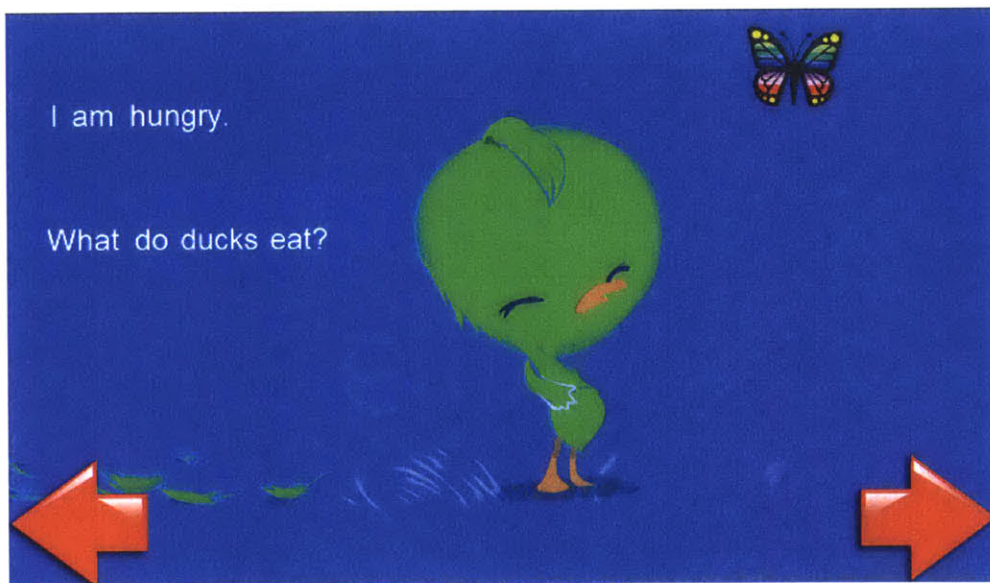


Figure 22: Tinkerbook with prompting character.

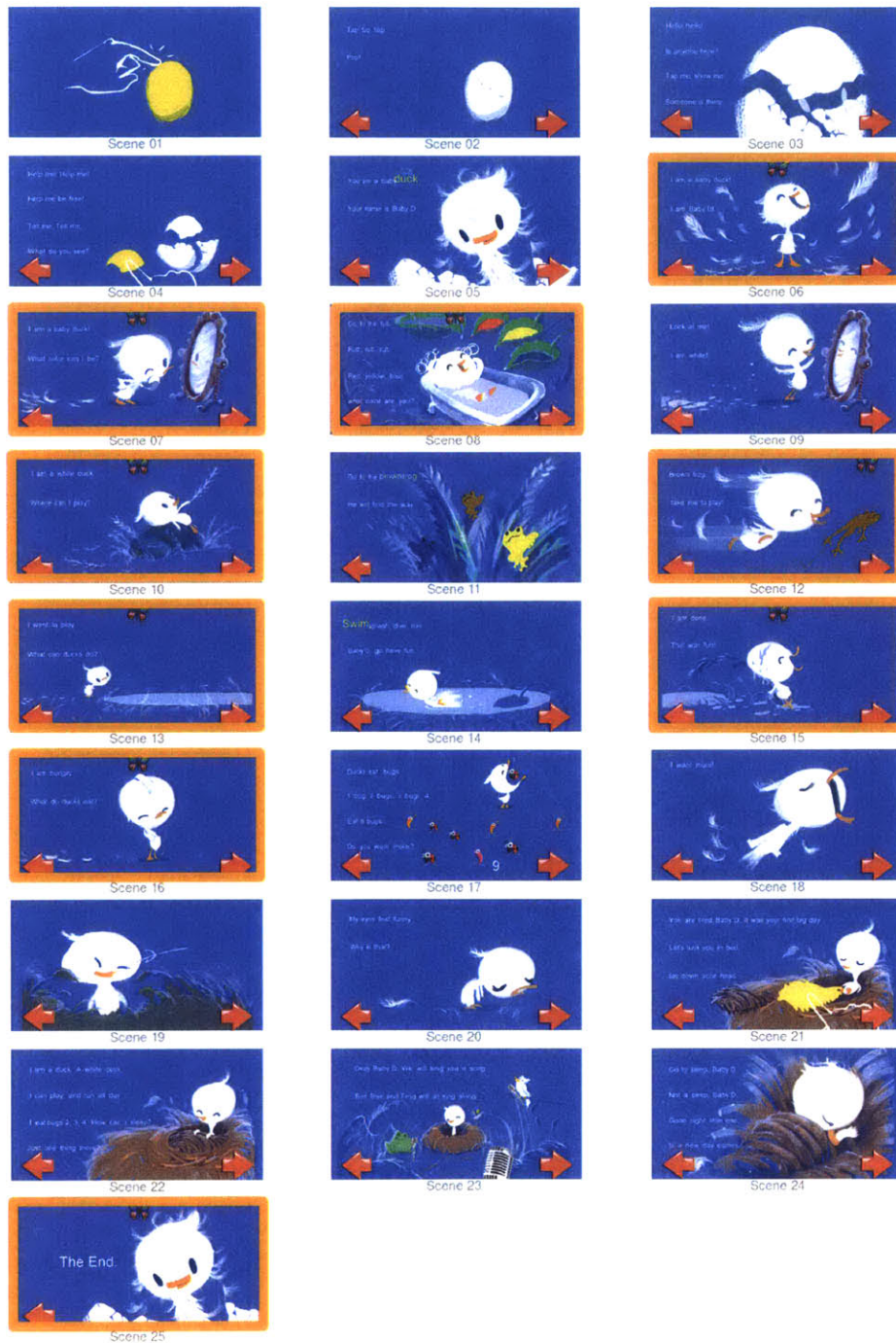


Figure 23: Tinkerbook with Prompting Scenes Highlighted.

6.2.2.3 Prompting Questions

The prompting questions, themselves, were generated from two different sources.

6.2.2.3.1 Expert Questions A set of questions was designed for each page of the Tinkerbook based on guidance from a literacy expert (N. Lasaux, personal communication, April 2013) and from an educator’s guide to literacy conversations with children (Barbara A Wasik & Iannone-Campbell, 2012). These led to manually authored prompts based on dialogic reading techniques. Each scene had its own limited and specific set of prompts from which the system randomly chose when the user tapped the prompting character. Prompts are listed in Table 5.

Table 5: Expert Prompts for Tinkerbook by Scene.










Scene	Questions
	<ul style="list-style-type: none">• Describe how Baby D is feeling right now.• What do you notice about Baby D’s Feathers?• What do you think Baby D will do next? ...
	<ul style="list-style-type: none">• What color can Baby D be?• What is Baby D thinking about?• What are you thinking right now?
	<ul style="list-style-type: none">• What color can Baby D be?• How do you make green?• What color can Baby D be next?• Why is Baby D in the bathtub?
	<ul style="list-style-type: none">• What will happen next?• What is Baby D thinking about?• What does this make you wonder about?

Table 5 – continued from previous page

Scene	Questions
	<ul style="list-style-type: none"> • Why do you think Baby D is following the frog? • What do you notice about the frog? • What are you thinking right now?
	<ul style="list-style-type: none"> • What will happen next? • What can ducks do? • What is Baby D thinking about?
	<ul style="list-style-type: none"> • What does this make you wonder about? • What do you notice about Baby D? • What do you think about the way Baby D plays?
	<ul style="list-style-type: none"> • What do you notice about Baby D's belly? • What is Baby D thinking about?
	<ul style="list-style-type: none"> • How is this story like other stories you have read? • Tell me the story in your own words. • How would you describe Baby D? • Is there something you didn't understand in the story?

6.2.2.3.2 Crowd-Sourced Questions Collaborators from Worcester Polytechnic Institute applied crowdsourcing and algorithmic techniques to model spoken dialog and topics sourced from recorded conversations between parents and children using Tinkerbook at the Boston Museum of Science in 2013. Their work generated a large set of new questions revolving around topics discussed during these interactions (Boteanu & Chernova, 2012). Questions were manually filtered for age-appropriateness (i.e. questions involving unacceptable adult topics were removed). When Tinkerbook is prompting from crowd-sourced questions, regardless of which scene a prompting character appeared, the prompt was chosen randomly from this entire set and presented to the readers. A list of these questions can be found in Appendix C).

6.2.3 Method

6.2.3.1 Participants

Parents with children aged 4-8 years old ($n = 88$, $M = 5.64$, $SD = 1.33$) were recruited in the Boston, Cambridge, Somerville and Arlington, Massachusetts areas. A combination of email announcements to various family lists and invitations to a volunteer subject list maintained by the Personal Robots Group generated participants. Parents who recommended subjects that successfully attended the study receive a referral bonus (gift certificate to online store). All subjects were informed of their rights and parents provided consent as mandated by the MIT Committee on the Use of Humans as Experimental Subjects (COUHES). Two subjects refused consent for video and audio recording, and their data was excluded and destroyed.

6.2.3.2 Experimental Manipulation

Subjects were randomly assigned to one of the three conditions during this study. The condition dictated the version of the story the child and parent would experience. In particular, the condition designated the appearance and behavior of the prompting character. The nature of the prompting made available to the subjects was the primary manipulation in this experiment.

- **Condition: NONE** - The first condition, as control, was the Baby D Tinkerbook story without any prompting from the mentoring character. If the parent or child tapped on this butterfly, nothing happens.
- **Condition: EXPERT** - In the second condition, on some scenes in the Tinkerbook, the story contained a mentoring character (a butterfly that

appeared at the top of the screen). When subjects tapped the butterfly, it would speak a prompt selected from a set derived from expert guidance and catered to the specific scene.

- **Condition: CROWD** - In the final condition, on some scenes in the Tinkerbook, the story contained a mentoring character (a butterfly that appeared at the top of the screen). When subjects tapped the butterfly, it would speak a random prompt selected from a large set, previously generated from the algorithmic, crowd-sourced method described by Boteanu and Chernova (Boteanu & Chernova, 2012).

In many cases, even if the butterfly character appears in the EXPERT or CROWD condition, the parent or child seemed not to notice or tap it to initiate the audio prompt. Sometimes the subjects referenced the butterfly in their conversation but did not tap on it. These subjects were reassigned to the NONE condition. Table 6 shows the number of subjects per condition.

Table 6: Number of Subjects per Condition.

NONE	43
EXPERT	22
CROWD	20
Total (n)	85

6.2.3.3 Procedure

Upon arrival to the lab at MIT, the parent and child were directed to the study room to sit in two chairs in front of a table (adult on the left, child on the right). Here, a video camera, microphone, and tablet were already present. The researcher started the recording using GlobalLit live; audio, video, and data about any interaction with the tablet would stream into a computer, also in the room.

The researcher would engage in a small talk with the child, for example, asking her name, age, and birth date. He would then show the child the tablet computer and ask if they “had seen something like this before... like a tablet or an iPad.” Regardless of the answer, the researcher would then explain that the child could touch the words and pictures they see on the screen. Then, the researcher would tell the parent and child that they would be using the tablet to “read a story together about a baby duck.” The two were instructed

to talk about what they saw on the screen and to treat this like a book they would read together at home.

The researcher asked parent and children if they have any questions. If none, the researcher would hand the child the tablet and instruct the pair to alert him when “they are done.” As the two played together, they would have no further interaction with the researcher, who would be observing the interaction through the GlobalLit Live dashboard. When the pair indicated the experience was over, the researcher indicated the end of the session in the data and the system cropped the recording and the submitted the video for annotation.

6.2.3.4 Measures

The tablet was instrumented using the GlobalLit Live system to collect interaction data and record the verbal utterances from the readers as described elsewhere in this thesis. The system captured every finger up, down, and move event. Additionally, specific interactions with words, graphics, and prompts were noted in the data probes.

In addition, the sessions were recorded via video and microphone and the streams were annotated, in real-time with information about the progress of the story. For example, the system superimposed on the video labels showing the timestamp, the current page, and an indicator when the user touches the screen. Figure 24 shows a sample annotated video frame. At the conclusion of a session, the system automatically packaged and uploaded the video and audio streams for transcription using a third-party API and service (“CastingWords,” 2013). Listing 6.2 shows a sample snippet from one transcript. These text transcriptions were analyzed per the measures described below. A summary of all measurements can be found in Table 9.

Data collected using the GlobalLit Live system was analyzed to obtain measurements for engagement, oral vocabulary, and dialogic reading behaviors. Measures that could be obtained requiring only minimal human oversight (ex. automatic text analysis) take precedence over measures that would require large amounts of researcher labor (ex. coding video). In particular, automatic measures might be better suited in the context of the Global Literacy Project, especially at large scales.

However, some measurements required coding the transcripts, manually, per the methodology suggested by (Parish-Morris et al., 2013). First, the transcripts were segmented by sentence utterances using the Natural Language Toolkit (Bird, 2006). This processing roughly designated phrase based utterances (phrases could include one word utterances or filler words like “uh”

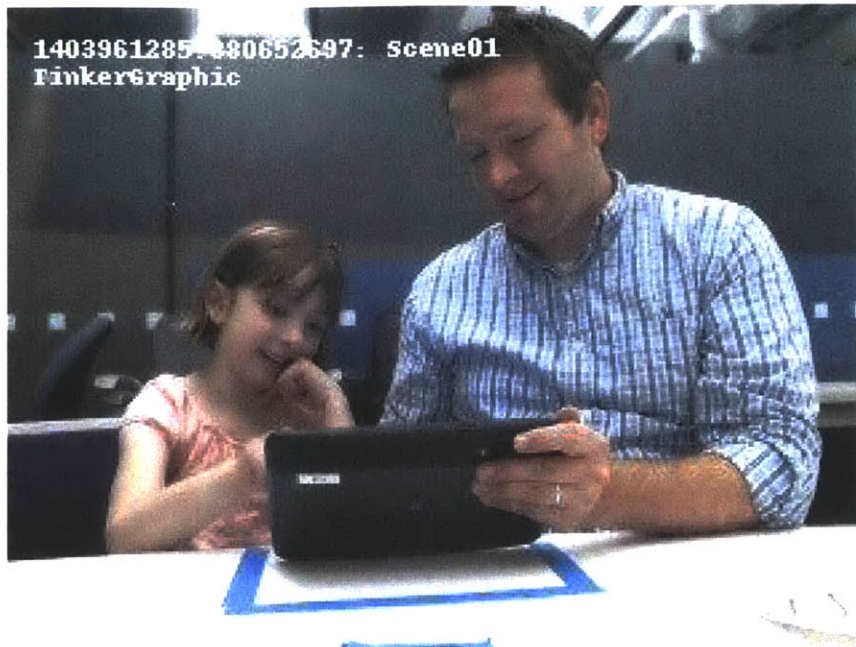


Figure 24: Annotated video frame from parent-child interaction.

Listing 6.2: Sample transcript from Tinkerbook session.

Child:	[0:22] Tap, tap.
Tablet:	[0:23] Tap, tap, tap, pop.
Parent:	[laughs] [0:25] Hello, hello. Is anyone here?
Child:	[0:30] Yeah.
Parent:	[0:30] Tap me, show me. Someone is there.
Parent:	[0:33] Help me, help me. Help me be free.
Tablet:	[0:48] Baby D.
Parent:	[laughs] [0:49]
Tablet:	[0:51] Baby D.
Parent:	[0:56] I'm a baby duck. I am Baby D.
Tablet:	[0:59] What do you notice about Baby D's feathers?
Child:	[1:04] Flying off?
Parent:	[1:06] What do you notice about the feathers?

and "hmmm"). These text-based utterances were then coded by a single researcher into one of the five categories as described in the table below.

Table 7: Categories for Coding.

Code	Description	Examples
Reading	The child or parent reading the text on the screen.	<ul style="list-style-type: none"> • "Tap me. Tap me. What do you see?" • "Go to the blue frog." • "He will find the way." ...
Non	Monosyllabic responses, non-verbal utterances, or indecipherable words	<ul style="list-style-type: none"> • "uhmm" • "yes" • "garrp." ...
Behavior	Utterances not related to the story. Could refer to interactive elements, interactions with the app, or commands to child.	<ul style="list-style-type: none"> • "Tap on the egg," • "Scootch closer," • "Do you want to read?" • "what happens if we touch the duck?" • "Go back a page." • "I want to touch the arrow." ...
Story	Comments about the story or about words in the story	<ul style="list-style-type: none"> • "Baby D looks hungry." • "It looks really tired." • "What does that letter sound like?" ...
Distancing	Prompts that extend the story beyond the here and the now. Includes prompts about what a child is thinking or guessing.	<ul style="list-style-type: none"> • "What do ducks eat?" • "This is like the other duck story we read." • "What does his voice make you think about?" • "What, what do you think is gonna happen next?" ...

6.2.3.4.1 Measures of Engagement Engagement with the app is an indication of how involved the parent and child are in the reading experience. We use surrogate measurements of engagement as other researchers have in different reading experiments (Couse & Chen, 2010).

- **Total time** in story is an indication of how long the subjects are willing to commit to the experience since the subjects self-selected when to end the reading session.
- **Engagement-related utterances** is the sum of story-related utterances, a manually coded measurement of how many utterances relate to the story in proportion to other utterances (Parish-Morris et al., 2013) and proportion of distancing prompts. Note that distancing prompts are a subcategory of story-related utterances, but they are coded separately to measure dialogic behaviors.
- **Behavior-related utterances** is a manually coded measurement of how many utterances not related to the story. These could refer to interactive elements, interactions with the app, or commands to the child. These utterances would relate to topics that are distracting or otherwise off-topic from the story, indicating a lack of engagement. Distancing prompts, even if not necessarily related directly to the story, are still considered to maintain engagement. (Parish-Morris et al., 2013).

6.2.3.4.2 Measures of Vocabulary The quantity and nature of the vocabulary words uttered were measured.

- **Full Utterances** is a count of all complete utterances made by the subjects, including phrases, sentences, or distinct single word utterances, normalized by session length.
- **Single Words** is a count of every word uttered by subjects, normalized by session length. This measure is different from full utterances as it is an attempt to differentiate between quantity of words spoken versus complexity of sentences.
- **Novel Words** are words not included in the text of the story uttered by parent and child, normalized by session length. These words could point to a general richness of vocabulary during dialog, including distancing language (i.e. conversation above and beyond what is happening in the story). Non-novel words are those words already included in the story (see Table 8).
- **Lexical Diversity** is a measurement of what percentage of words spoken that are different from all the other words spoken. The analysis

calculates this measure by counting the number of unique words spoken and dividing by the total number of words spoken. (Dale & Fenson, 1996)

Table 8: Words Included in Tinkerbook.

• a	• eight	• little	• take
• all	• eyes	• look	• tap
• along	• feel	• me	• tell
• am	• find	• more	• ten
• and	• first	• my	• that
• anyone	• five	• name	• the
• are	• four	• new	• there
• at	• free	• night	• thing
• baby	• frog	• nine	• three
• be	• fun	• not	• til
• bed	• funny	• okay	• tired
• bee	• go	• one	• to
• big	• good	• orange	• tub
• bird	• green	• peep	• tuck
• blue	• have	• play	• two
• brown	• he	• pop	• want
• bug	• head	• purple	• was
• bugs	• hello	• red	• way
• can	• help	• rub	• we
• color	• here	• run	• what
• comes	• how	• see	• where
• d	• hungry	• seven	• white
• day	• i	• show	• why
• dive	• I	• sing	• will
• do	• in	• six	• yellow
• done	• is	• sleep	• you
• down	• it	• someone	• your
• duck	• just	• song	
• ducks	• lay	• splash	
• eat	• let's	• swim	

6.2.3.4.3 Measurements of Dialogic Reading Behaviors Spontaneous examples of dialogic reading behavior were noted and measured.

- **Turn-taking Exchanges** is a measure of the number of conversational exchanges, normalized by session length, in the dialog between parent and child, indicating a two-way conversation rather than just the parent or child speaking.
- **Distancing Prompts** is a manually coded measurement of prompt utterances that relate the story to other experiences or examples in the child's life as a percentage of all utterances (Zevenbergen et al., 2003, Parish-Morris et al. (2013)).
- **Question utterances** is a count of the number of questions the parent asks the child. It is a measure of prompting behaviors (Blewitt, Rump, Shealy, & Cook, 2009), and it is estimated by counting the number of question marks that appeared in the transcripts, normalized by session length.

Table 9: Summary of Measurements.

Measure	Method	Norm.
H1 - Engagement		
Total Time	Seconds from the beginning of the interaction until subjects declare they are finished.	None
Engagement-Related Utterances	Manually Coded Utterances	%
Behavior-Related Utterances	Manually Coded Utterances	%
H2 - Vocabulary		
Full Utterances	Distinct utterances counted in text transcripts via lexical parsing.	Time
Single Words	Individual words counted in text transcripts.	Time
Novel Words	Unique words counted in text transcripts.	Time
Lexical Diversity	Unique words as a proportion of all words counted in transcripts.	None
H3 - Dialogic Reading		
Turn-Taking	Count number of transitions between speakers in text transcripts.	Time
Distancing Prompts	Manually Coded Utterances.	%
Question Utterances	Approximated by counting question marks in text transcripts.	Time

6.2.4 Results

6.2.4.1 Coding

Immediately after a session concluded, data analysis began. The Global-Lit Live software and cloud-based transcription service took about 22 hours to process each session (i.e. generating text transcript and analyzing the text) ($n = 85$, $M = 22.835$, $SEM = 1.108$). For the utterance categorization measurements, accomplished via manual coding, a single researcher processed 12110 utterances across 10.09 hours across the 85 trials. As shown in Figure 25, most of the utterances were coded as reading out loud from the story (31.4%). Story-related (21%) and behavior-related utterances (20.7%) were coding approximately the same amount. Dialogic prompts were most rare (7.2%), and non-utterances made up almost a fifth (19.7%) of the coded utterances. See Figure 26 for breakdown of coding per condition.

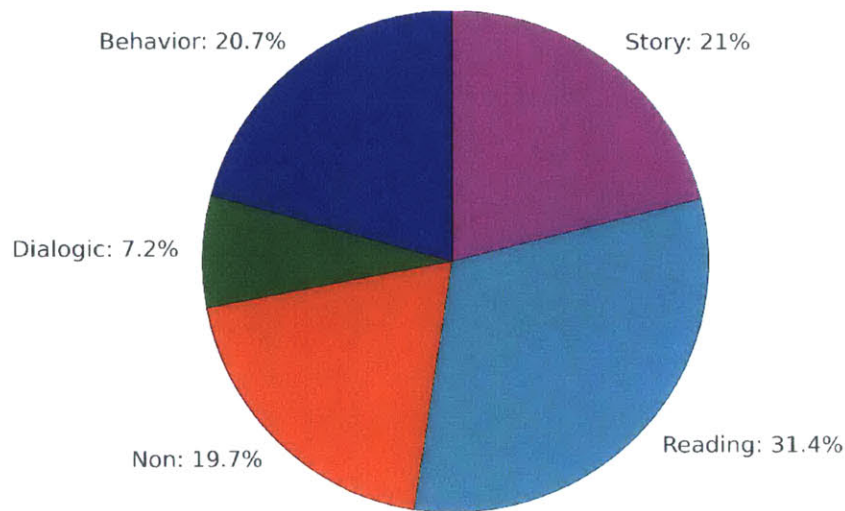
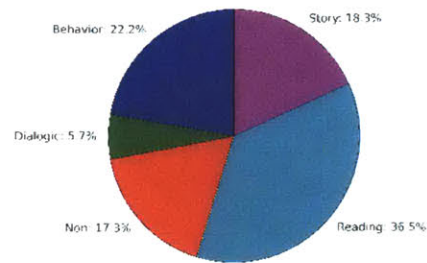
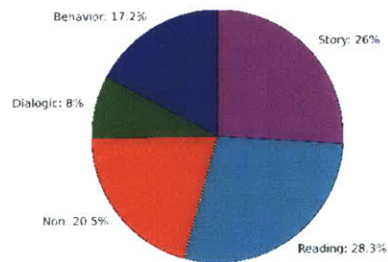


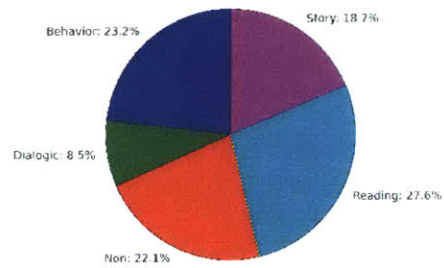
Figure 25: Percentage of coded utterances per category across all conditions.



(a) Coded Utterances in NONE



(b) Coded Utterances in EXPERT



(c) Coded Utterances in CROWD

Figure 26: Percentage of coded utterances per category by condition.

6.2.4.2 Engagement

The measures of engagement seem to support H1 (i.e. prompts would increase engagement). Data is presented as mean \pm standard deviation.

6.2.4.2.1 Total seconds A one-way Welch ANOVA was conducted to determine if the proportion of total seconds was different for each condition (see Figure 27). There were no outliers, as assessed by boxplot; data was not normally distributed for each group, as assessed by Shapiro-Wilk test, so it was log-transformed before analysis. There was not homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = 0.071$). The log-transformed total seconds increased from the CROWD ($M = 2.6531$, $SD = 0.20702$, 95% CI [2.5563, 2.75], *back-transformed* $M = 449.88$, *back-transformed* 95% CI [360.0, 562.34]) to the NONE ($M = 2.6573$, $SD = 0.13838$, 95% CI [2.6136, 2.701], *back-transformed* $M = 454.26$, *back-transformed* 95% CI [410.77, 502.34]), and EXPERT ($M = 2.7724$, $SD = 0.13859$, 95% CI [2.7109, 2.8338], *back-transformed* $M = 592.11$, *back-transformed* 95% CI [513.93, 682.02]) conditions, in that order. Between the different conditions, the transformed total seconds were statistically significantly different, Welch's $F(2,39.24) = 5.251$, $p = 0.01$, $\omega^2 = 0.0929$. Planned contrasts on the transformed data revealed that the mean difference from NONE to EXPERT ($\Delta M = 0.1151$, $t(43.015) = 3.143$, $p = 0.003$) was statistically significant, the mean difference from NONE to CROWD ($\Delta M = -0.0042$, $t(27.563) = -0.081$, $p = 0.936$) was not statistically significant, the mean difference from EXPERT to CROWD ($\Delta M = -0.1193$, $t(32.722) = -2.171$, $p = 0.037$) was statistically significant. This supports H1 since subjects spent more time with the app when a prompting character appeared expressing expert-generated questions, perhaps encouraging more time with the story. However, there wasn't a significant increase in time with the crowd-sourced questions, in opposition to H4.

6.2.4.2.2 Engagement-related utterances A one-way Welch ANOVA was conducted to determine if the proportion of engagement-related utterances was different for each condition (see Figure 28). There were no outliers, as assessed by boxplot; data was normally distributed for each group, as assessed by Shapiro-Wilk test ($p > .05$). There was not homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = 0.279$). Engagement-related utterances increased from the NONE ($M = 0.2014$, $SD = 0.11868$, 95% CI [0.1639, 0.2388]) to the CROWD ($M = 0.2155$, $SD = 0.10297$, 95% CI [0.1673, 0.2637]), and EXPERT ($M = 0.3112$, $SD = 0.08956$, 95% CI [0.2715, 0.3509]) conditions, in that order. Between the different conditions, Engagement-related

utterances were statistically significantly different, *Welch's* $F(2,45.748) = 9.456$, $p < 0.001$, $\omega^2 = 0.169$. Planned contrasts revealed that the mean difference from NONE to EXPERT ($\Delta M = 0.1098$, $t(54.033) = 4.126$, $p < 0.001$) was statistically significant, the mean difference from NONE to CROWD ($\Delta M = 0.0141$, $t(43.021) = 0.479$, $p = 0.634$) was not statistically significant, the mean difference from EXPERT to CROWD ($\Delta M = -0.0957$, $t(37.901) = -3.197$, $p = 0.003$) was statistically significant. This seems to support H1 since more engagement related utterances were measured, perhaps inspired by the story, when the expert prompts appeared. However, the crowd sourced prompts did not show a noticeable increase in engagement-related prompts.

6.2.4.2.3 Behavior-related utterances A one-way ANOVA was conducted to determine if the proportion of behavior-related utterances was different for each condition (see Figure 29). There were no outliers, as assessed by boxplot; data was normally distributed for each group, as assessed by Shapiro-Wilk test ($p > .05$). There was homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = 0.025$). Behavior-related utterances increased from the EXPERT ($M = 0.1725$, $SD = 0.06713$, 95% CI [0.1427, 0.2022]) to the CROWD ($M = 0.2025$, $SD = 0.09573$, 95% CI [0.1577, 0.2473]), and NONE ($M = 0.2477$, $SD = 0.11518$, 95% CI [0.2114, 0.2841]) conditions, in that order. Between the different conditions, Behavior-related utterances were statistically significantly different, $F(2,80) = 4.332$, $p = 0.016$, $\omega^2 = 0.0748$. Planned contrasts revealed that the mean difference from NONE to EXPERT ($\Delta M = -0.0752$, $t(80) = -2.849$, $p = 0.006$) was statistically significant, the mean difference from NONE to CROWD ($\Delta M = -0.0452$, $t(80) = -1.66$, $p = 0.101$) was not statistically significant, the mean difference from EXPERT to CROWD ($\Delta M = 0.03$, $t(80) = 0.972$, $p = 0.334$) was not statistically significant. This seems to support H1, as off-topic, behavior-related utterances decreased when prompts appeared. However, the increase was only significant for the expert prompts, showing lack of support for H4.

6.2.4.3 Vocabulary

The measures of vocabulary seem partially to support H2 (i.e. prompting increases the quantity and quality of vocabulary).

6.2.4.3.1 Full utterances A one-way Welch ANOVA was conducted to determine if the proportion of full utterances was different for each condition (see Figure 30). There were no outliers, as assessed by boxplot; data was normally distributed for each group, as assessed by Shapiro-Wilk test ($p >$

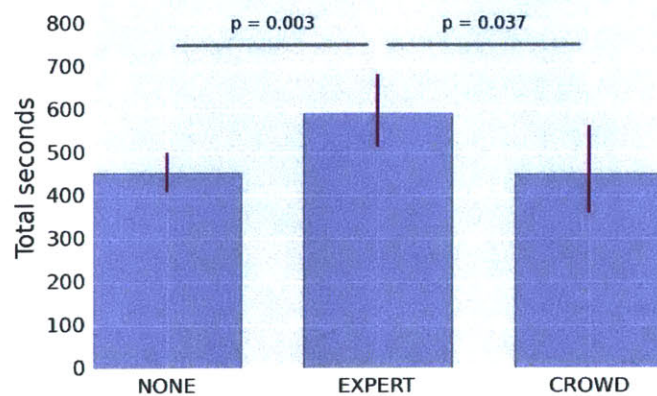


Figure 27: Mean and 95% confidence intervals for total seconds by condition. Analysis by one-way Welch Anova $F(2,39.24) = 5.251, p = 0.01$. Annotation indicates contrast significance. Analysis conducted on log-transformed data, then back-transformed for this graph.

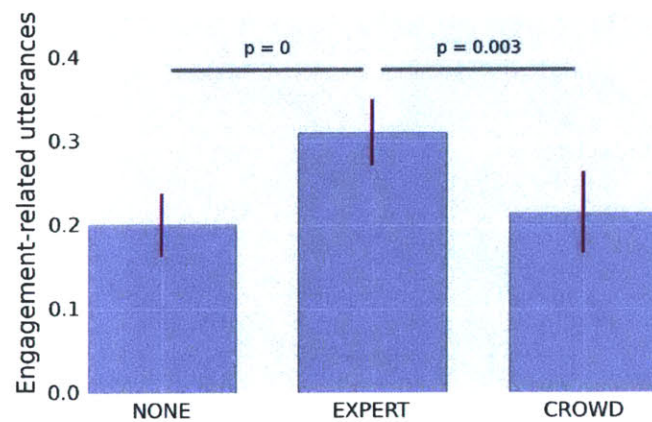


Figure 28: Mean and 95% confidence intervals for engagement-related utterances by condition. Analysis by one-way Welch Anova $F(2,45.748) = 9.456, p < 0.001$. Annotation indicates contrast significance.

.05). There was not homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = 1.792$). Full utterances increased from the NONE ($M = 0.2448$, $SD = 0.10962$, 95% CI [0.2102, 0.2794]) to the CROWD ($M = 0.3051$, $SD = 0.10109$, 95% CI [0.2578, 0.3524]), and EXPERT ($M = 0.3092$, $SD = 0.07705$, 95% CI [0.2751, 0.3434]) conditions, in that order. Between the different conditions, Full utterances were statistically significantly different, *Welch's* $F(2,45.365) = 4.169$, $p = 0.022$, $\omega^2 = 0.0709$. Planned contrasts revealed that the mean difference from NONE to EXPERT ($\Delta M = 0.0644$, $t(56.436) = 2.714$, $p = 0.009$) was statistically significant, the mean difference from NONE to CROWD ($\Delta M = 0.0603$, $t(40.688) = 2.124$, $p = 0.04$) was statistically significant, the mean difference from EXPERT to CROWD ($\Delta M = -0.0041$, $t(35.428) = -0.149$, $p = 0.882$) was not statistically significant. When prompts appeared, in either the expert or crowd-sourced conditions, the number of utterances increased, supporting H2 and H4.

6.2.4.3.2 Single words A one-way Welch ANOVA was conducted to determine if the proportion of single words was different for each condition (see Figure 31). There were no outliers, as assessed by boxplot; data was normally distributed for each group, as assessed by Shapiro-Wilk test ($p > .05$). There was not homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = 1.634$). Single words increased from the NONE ($M = 1.0882$, $SD = 0.4634$, 95% CI [0.9419, 1.2344]) to the CROWD ($M = 1.3125$, $SD = 0.4019$, 95% CI [1.1244, 1.5006]), and EXPERT ($M = 1.3174$, $SD = 0.31617$, 95% CI [1.1772, 1.4576]) conditions, in that order. Between the different conditions, Single words were statistically significantly different, *Welch's* $F(2,45.365) = 3.132$, $p = 0.022$, $\omega^2 = 0.0489$. Planned contrasts revealed that the mean difference from NONE to EXPERT ($\Delta M = 0.2292$, $t(57.326) = 2.318$, $p = 0.024$) was statistically significant, the mean difference from NONE to CROWD ($\Delta M = 0.2243$, $t(43.037) = 1.944$, $p = 0.058$) was not statistically significant, the mean difference from EXPERT to CROWD ($\Delta M = -0.0049$, $t(36.065) = -0.044$, $p = 0.965$) was not statistically significant. When the expert prompts appeared, the parents and children uttered more words, supporting H2. However, there was not support for H4, as the increase in words in the CROWD condition was not significant.

6.2.4.3.3 Novel words A one-way Welch ANOVA was conducted to determine if the proportion of novel words was different for each condition (see Figure 32). There were no outliers, as assessed by boxplot; data was normally distributed for each group, as assessed by Shapiro-Wilk test ($p > .05$).

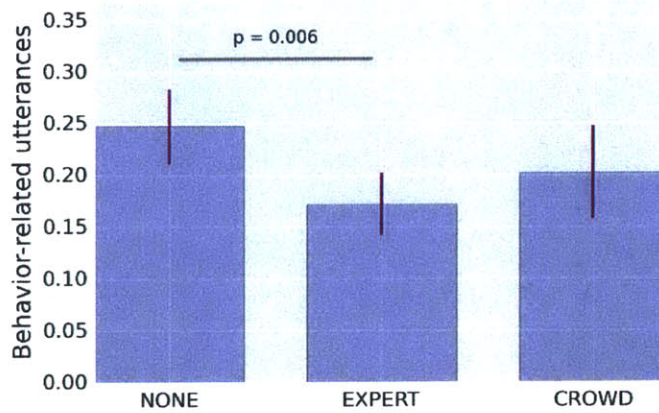


Figure 29: Mean and 95% confidence intervals for behavior-related utterances by condition. Analysis by one-way ANOVA $F(2,80) = 4.332$, $p = 0.016$. Annotation indicates contrast significance.

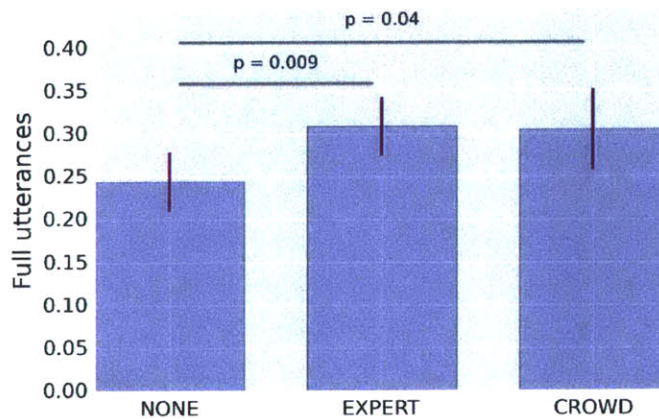


Figure 30: Mean and 95% confidence intervals for full utterances by condition. Analysis by one-way Welch Anova $F(2,45.365) = 4.169$, $p = 0.022$. Annotation indicates contrast significance.

There was not homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = 3.418$). Novel words increased from the NONE ($M = 0.3254$, $SD = 0.15379$, 95% CI [0.2768, 0.3739]) to the CROWD ($M = 0.4308$, $SD = 0.22971$, 95% CI [0.3233, 0.5383]), and EXPERT ($M = 0.4576$, $SD = 0.16021$, 95% CI [0.3866, 0.5287]) conditions, in that order. Between the different conditions, Novel words were statistically significantly different, *Welch's* $F(2,38.939) = 5.552$, $p = 0.008$, $\omega^2 = 0.0988$. Planned contrasts revealed that the mean difference from NONE to EXPERT ($\Delta M = 0.1322$, $t(41.565) = 3.168$, $p = 0.003$) was statistically significant, the mean difference from NONE to CROWD ($\Delta M = 0.1054$, $t(27.59) = 1.859$, $p = 0.074$) was not statistically significant, the mean difference from EXPERT to CROWD ($\Delta M = -0.0268$, $t(33.579) = -0.435$, $p = 0.666$) was not statistically significant. More words, not included in the story, were uttered when expert prompts appeared, supporting H2. While more novel words were uttered with crowd-sourced prompts, the increase was not significant, so we do not have support for H4.

6.2.4.3.4 Lexical diversity A one-way Welch ANOVA was conducted to determine if the proportion of lexical diversity was different for each condition (see Figure 33). There were no outliers, as assessed by boxplot; data was normally distributed for each group, as assessed by Shapiro-Wilk test ($p > .05$). There was not homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = 2.549$). Lexical diversity increased from the EXPERT ($M = 0.35275$, $SD = 0.052392$, 95% CI [0.32953, 0.37598]) to the CROWD ($M = 0.39671$, $SD = 0.10904$, 95% CI [0.34568, 0.44774]), and NONE ($M = 0.42413$, $SD = 0.11725$, 95% CI [0.38712, 0.46114]) conditions, in that order. Between the different conditions, Lexical diversity were statistically significantly different, *Welch's* $F(2,43.951) = 5.889$, $p = 0.005$, $\omega^2 = 0.105$. Planned contrasts revealed that the mean difference from NONE to EXPERT ($\Delta M = -0.071378192$, $t(59.59) = -3.328$, $p = 0.002$) was statistically significant, the mean difference from NONE to CROWD ($\Delta M = -0.027422396$, $t(40.378) = -0.899$, $p = 0.374$) was not statistically significant, the mean difference from EXPERT to CROWD ($\Delta M = 0.043955796$, $t(26.747) = 1.639$, $p = 0.113$) was not statistically significant. This result is in opposition to H2, as the language used in the interaction between parents and children was more diverse when prompts did not appear.

6.2.4.4 Dialogic Reading

6.2.4.4.1 Turn-taking exchanges A one-way Welch ANOVA was conducted to determine if the proportion of turn-taking exchanges was different for each

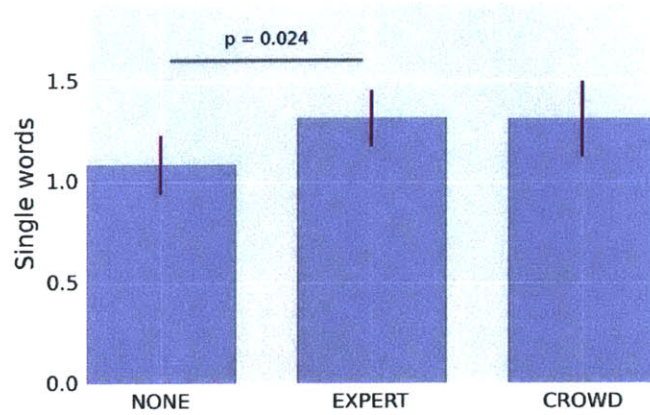


Figure 31: Mean and 95% confidence intervals for single words by condition. Analysis by one-way Welch Anova $F(2,45.365) = 3.132, p = 0.022$. Annotation indicates contrast significance.

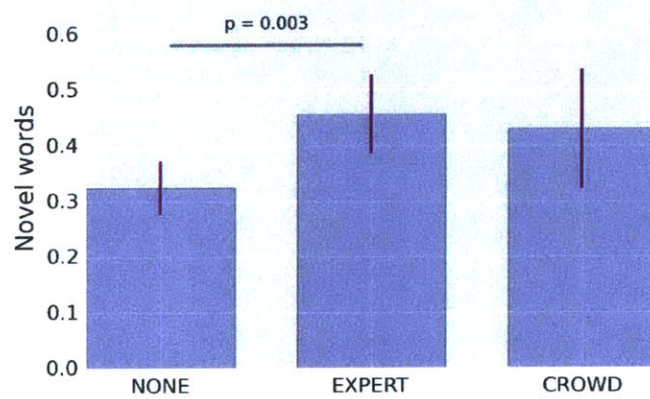


Figure 32: Mean and 95% confidence intervals for novel words by condition. Analysis by one-way Welch Anova $F(2,38.939) = 5.552, p = 0.008$. Annotation indicates contrast significance.

condition (see Figure 34). There were no outliers, as assessed by boxplot; data was not normally distributed for each group, as assessed by Shapiro-Wilk test, so it was square-root transformed before analysis. There was not homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = 0.696$). The square-root transformed turn-taking exchanges increased from the NONE ($M = 0.2554$, $SD = 0.09345$, 95% CI [0.2259, 0.2849], *back-transformed* $M = 0.065229$, *back-transformed* 95% CI [0.051031, 0.081168]) to the CROWD ($M = 0.3039$, $SD = 0.07544$, 95% CI [0.2686, 0.3392], *back-transformed* $M = 0.092355$, *back-transformed* 95% CI [0.072146, 0.11506]), and EXPERT ($M = 0.3572$, $SD = 0.07489$, 95% CI [0.324, 0.3904], *back-transformed* $M = 0.12759$, *back-transformed* 95% CI [0.10498, 0.15241]) conditions, in that order. Between the different conditions, the transformed turn-taking exchanges were statistically significantly different, *Welch's* $F(2,46.172) = 10.923$, $p < 0.001$, $\omega^2 = 0.193$. Planned contrasts on the transformed data revealed that the mean difference from NONE to EXPERT ($\Delta M = 0.1018$, $t(51.775) = 4.706$, $p < 0.001$) was statistically significant, the mean difference from NONE to CROWD ($\Delta M = 0.0485$, $t(45.878) = 2.174$, $p = 0.035$) was statistically significant, the mean difference from EXPERT to CROWD ($\Delta M = -0.0533$, $t(39.563) = -2.295$, $p = 0.027$) was statistically significant. Parents and children took turns when talking more times when a prompt appeared, from either the EXPERT or CROWD condition, in support of H3 and H4, indicating that prompts might be helpful for inducing dialogic reading techniques.

6.2.4.4.2 Question utterances A one-way Welch ANOVA was conducted to determine if the proportion of questions was different for each condition (see Figure 35). There were no outliers, as assessed by boxplot; data was not normally distributed for each group, as assessed by Shapiro-Wilk test, so it was log-transformed before analysis. There was not homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = 0.531$). The log-transformed questions increased from the NONE ($M = 0.0286$, $SD = 0.01428$, 95% CI [0.0241, 0.0331], *back-transformed* $M = 0.068071$, *back-transformed* 95% CI [0.057061, 0.079195]) to the CROWD ($M = 0.0392$, $SD = 0.01884$, 95% CI [0.0304, 0.048], *back-transformed* $M = 0.09446$, *back-transformed* 95% CI [0.072507, 0.11686]), and EXPERT ($M = 0.0397$, $SD = 0.0138$, 95% CI [0.0336, 0.0459], *back-transformed* $M = 0.095721$, *back-transformed* 95% CI [0.080438, 0.11148]) conditions, in that order. Between the different conditions, the transformed questions were statistically significantly different, *Welch's* $F(2,40.448) = 5.487$, $p = 0.008$, $\omega^2 = 0.0976$. Planned contrasts on the transformed data revealed that the mean difference from NONE to EXPERT ($\Delta M = 0.0111$, $t(44.373) = 3.02$,

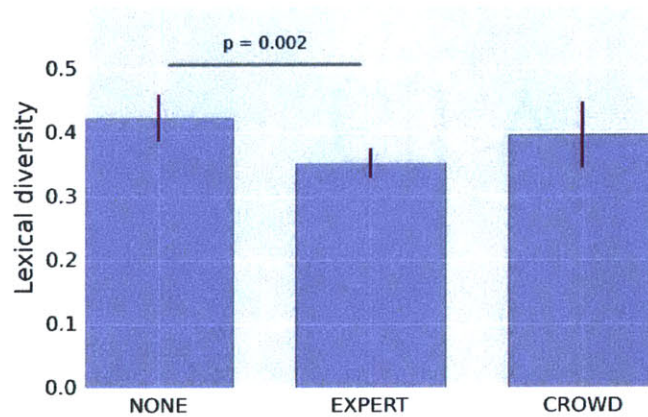


Figure 33: Mean and 95% confidence intervals for lexical diversity by condition. Analysis by one-way Welch Anova $F(2,43.951) = 5.889, p = 0.005$. Annotation indicates contrast significance.

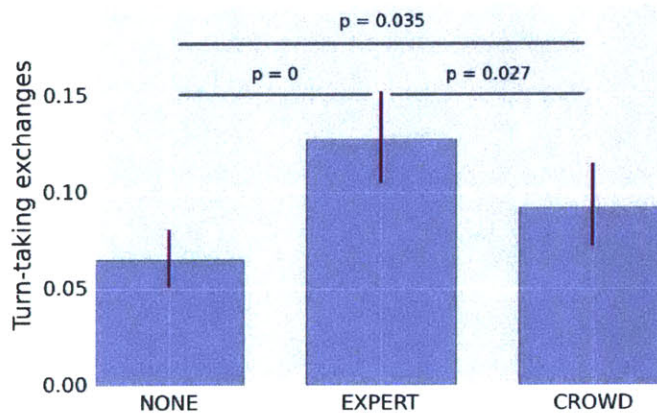


Figure 34: Mean and 95% confidence intervals for turn-taking exchanges by condition. Analysis by one-way Welch Anova $F(2,46.172) = 10.923, p < 0.001$. Annotation indicates contrast significance. Analysis conducted on square-root transformed data, then back-transformed for this graph.

$p = 0.004$) was statistically significant, the mean difference from NONE to CROWD ($\Delta M = 0.0106$, $t(30.016) = 2.222$, $p = 0.034$) was statistically significant, the mean difference from EXPERT to CROWD ($\Delta M = -0.0005$, $t(34.601) = -0.107$, $p = 0.915$) was not statistically significant. In support of H3 and H4, parents tended to ask more questions of their children when either the expert or crowd-sourced prompts appeared.

6.2.4.4.3 Distancing prompts A one-way Welch ANOVA was conducted to determine if the proportion of distancing prompts was different for each condition (see Figure 36). The data for dialogic utterances was not normally distributed and could not be corrected through reasonable transformations. The NONE condition had many high outliers (i.e. evidence against the hypothesis). The exclusion of these outliers did improve the normality of the data in the NONE condition, and it did not seem substantially to alter the overall result of the ANOVA. For this reason, since ANOVAs can be fairly robust with non-normal data (Maxwell & Delaney, 2004), I chose to proceed with the Welch ANOVA with the original data.

There was not homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = 0.054$). Distancing prompts increased from the NONE ($M = 0.0276$, $SD = 0.0277$, 95% CI [0.0189, 0.0364]) to the EXPERT ($M = 0.0434$, $SD = 0.02489$, 95% CI [0.0323, 0.0544]), and CROWD ($M = 0.0476$, $SD = 0.03884$, 95% CI [0.0294, 0.0658]) conditions, in that order. Between the different conditions, Distancing prompts were statistically significantly different, Welch's $F(2, 40.506) = 3.665$, $p = 0.034$, $\omega^2 = 0.0603$. Planned contrasts revealed that the mean difference from NONE to EXPERT ($\Delta M = 0.0158$, $t(47.232) = 2.299$, $p = 0.026$) was statistically significant, the mean difference from NONE to CROWD ($\Delta M = 0.02$, $t(28.755) = 2.059$, $p = 0.049$) was statistically significant, the mean difference from EXPERT to CROWD ($\Delta M = 0.0042$, $t(31.822) = 0.416$, $p = 0.68$) was not statistically significant. In support of H3, when the butterfly appeared, more distancing prompts were uttered, encouraging children to think about the story beyond the present reading experience, a critical dialogic reading technique. Furthermore, subjects in the CROWD condition tended to utter the most distancing prompts.

6.2.4.5 Crowd

As described above, H4 seemed to be partially supported as the CROWD condition yielded results that were consistent with measures supporting engagement, vocabulary, and dialogic reading behaviors more than the NONE

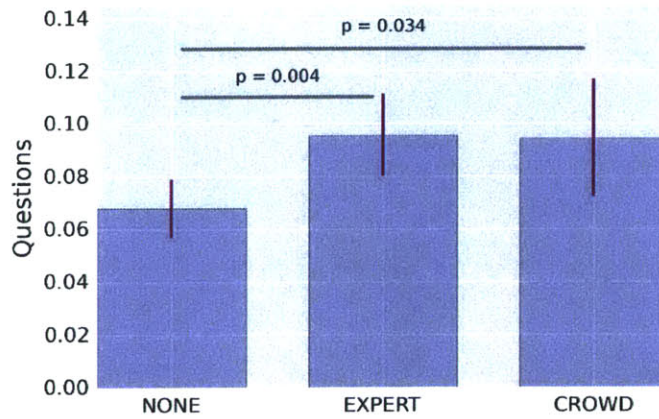


Figure 35: Mean and 95% confidence intervals for questions by condition. Analysis by one-way Welch Anova $F(2,40.448) = 5.487, p = 0.008$. Annotation indicates contrast significance. Analysis conducted on log-transformed data, then back-transformed for this graph.

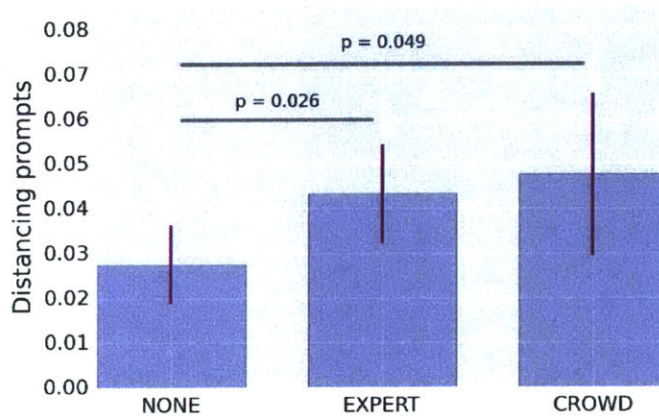


Figure 36: Mean and 95% confidence intervals for distancing prompts by condition. Analysis by one-way Welch Anova $F(2,40.506) = 3.665, p = 0.034$. Annotation indicates contrast significance.

condition in almost all cases. However, it did not have as much an effect as the EXPERT condition.

6.3 Discussion

Almost entirely across the board, the surrogate measures pointed strongly towards a significant positive correlation between the mentoring prompts and measures of engagement, vocabulary, and dialogic reading behaviors. The expert prompts seemed to relate to increased engagement more than the crowd prompts, but they both relate to increased engagement compared to the control.

The experiment shows support for H1, when the expert prompts appeared; parents and children spent more time reading the story together and talking about the story. When the prompts did not appear, the dyads tend to talk about behavior-related topics. More of their utterances were unrelated to the story and were more about the context around the interaction itself, indicating a disengagement with the story. However, H4 was not supported by the engagement measurements, as the crowdsourced prompts did not seem significantly to impact the engagement. Perhaps this could be explained by the non-sequitur nature of some of the prompts. For example, when a crowdsourced prompt appeared that did not obviously connect to the story, sometimes this would distract the subjects from the reading experience. As crowdsourcing algorithms improve, they might approach the content and form of expert generated prompts, and if they become indistinguishable from the expert prompts, we would expect they would provide similar improvements in engagement.

H2, the hypothesis related to increasing vocabulary, seemed to be partially supported by the experiment. Parents and children expressed more words and more full utterances in the sessions where the prompting character triggered. They also spoke more words that were not included in the story, suggesting a richer vocabulary pool. While the crowd-sourced prompts seemed to encourage more language, the only significant difference was found in the measure of full utterances, only partially supporting H4.

Their conversations were not lexically diversified with the prompts, however, as they tended to use more varied vocabulary in their dialog in the absence of prompts, suggesting a lack of support for H2. One possible explanation is that prompts tend to focus the parent and child's conversation towards the context of the prompt collection rather than completely undirected and wandering (albeit off-topic) discussion.

Most encouragingly, dialogic reading behaviors were observed to be more present with the prompting conditions, regardless of whether they were expert-generated or crowd-sourced. Parents and children took more turns in the conversation, indicating more dialog involving production and consumption of language rather than a single modality. Distancing prompts were not common among the sessions, but the prompts also seemed to encourage more of these, where the child is encouraged to think about the story beyond the here and the now. Parents asked more questions of their children when prompts appeared, encouraging more generation of language and, perhaps, critical thinking about the story.

These results point to the potential of a virtual character modeling dialogic reading behaviors as a method of delivering instruction in dialogic reading, when more traditional techniques are not available. The crowd-source prompts were most effective at encouraging distancing prompts, in particular. This result may be due to the sometimes off-topic nature of the questions, encouraging parents to attempt to make connections between the prompts and the story, themselves.

However, in all the other cases, when an expert helps to craft prompts, they are more effective than the crowdsourced prompts. This result seems probable, as sometimes the crowd prompts were rather off-topic. However, the crowd prompts were more effective than no prompts at all, suggesting that prompts can be a useful tool even if the questions are not perfect. As the algorithms improve to generate prompts that are closer to expert prompts, this bodes well for scalability.

The study did not control for reading levels (including not screening for atypical reading development). In addition, since there was such a diversity of age and styles of parent-child interaction, inter-subject comparisons might be problematic. There are opportunities to investigate this data further, within-subjects, to better understand the trajectory through a given story reading experience, giving us further insight into temporal dynamics of a reading session. For instance, we could measure novel vocabulary words spoken per second before the child triggers the first mentoring prompt to see if the mentoring character impacts the production of new spoken language in the same child. However, more care would be necessary to control for timing of triggering of prompts and ensuring the same prompts appear in the same way for each subject.

Another major weakness in the methodology revolves around the manual coding of transcripts. At the very least, multiple coders should be used to cross-validate results. While excluding video from the coding sessions was a deliberate choice, video could have been a useful contextual aid to reduce

misclassification of data. Nevertheless, each hypothesis was evaluated with a combination of automated and manual measures; take together, these measures seemed to corroborate the results. The data should be reexamined using multiple coders to validate the outcome.

While very encouraging for prompts in interactive storybooks as a potential literacy aid, connecting the results of this study to any gains in reading skills is tenuous as there were no literacy-related pre and post tests. In addition, for such short sessions, it is unlikely that there would be any long-lasting effects without repeated use. A future study might investigate how prompts can encourage parents and children to re-engage with the same story multiple times over a longer timeframe. Indeed, this is also consistent with dialogic reading techniques where parents are instructed to ask deeper questions with every re-reading of a story.

It may be possible for more deliberate modeling of the dialog between the parent and child. The system could track turn taking and conversation dynamics throughout the story reading, showing how the conversation evolves over the course of a session with and without prompting.

Data could be further analyzed for the dynamics of the reading session related to the story, itself. For example, it could show how long do the parent and child spend on any given page, and how their language relates to the specific elements of the story.

Data could be segregated by speaker (i.e. parent language versus child language) to understand better how the prompts affect the individuals rather than taking the aggregate language.

There is also an opportunity to innovate along the data collection methodology. For example, we could apply machine learning techniques to the coding of dialog to reintroduce automation in the data collection. Finally, we could analyze the story-driven interaction with more detail, seeing which graphics and words are touched on the screen and connecting them to words that are verbalized, for instance.

7 Conclusion

In this thesis, I presented a prototype platform for collecting, analyzing, and reacting to field data from children using educational apps on tablets.

I presented some design goals for the system, and throughout this work, I discussed how well I felt these goals were achieved with this first implementation. I pointed out where the system did not successfully meet these aims and how architecture decisions might be improved in future versions.

The thesis contained two use cases which provide examples and templates for how the GlobalLit system could enable other researchers to explore child-driven learning on tablets. One case focused on observations that could be derived from the data in remote locations (without making claims about the meaning of those analytics, but simply demonstrating a few types of questions that could be asked), and the second use case details how GlobalLit can facilitate a controlled experiment with tablets, children and parents that seemed to show how prompts in the tablet could lead to increased engagement.

The GlobalLit system has been a useful start with many successes, but as I reflect on the initial goals, it is useful to enumerate some specific recommendations for future improvements.

7.1 Goal: Geo-Shifted data is normalized across the globe.

This system has been used in multiple configurations. With every new deployment, modifications were made to the system based on lessons we learned from previous attempts. The degree to which we could normalize the data was greatly affected by the instability of the system. Data formats would change across deployments, for example, and the reliability of the system was suspect, leaving many gaps and noise in the incoming data. There were many unforeseen and difficult circumstances that made data normalization very difficult some were technical (like the clock resets in Ethiopia), but others were related to workarounds children discovered to circumvent our best efforts to direct their experiences. The reality is that the children will use the devices in very unexpected ways and the software, being a work-in-progress, often fell short on robustness. Deployment platforms should be well tested and bug checked ahead of deployments as they increase in distance from potential remedies.

7.2 Goal: Time-Shifted analysis allows us to understand life-cycle of tablet use and exploration models.

Most deployments lasted on the order of months, much longer than typical field studies but not long enough to reasonably answer questions about literacy development. Furthermore, children would get bored with the content available on the tablet, so it will become important to constantly refresh the available apps and games on these deployments to maintain children's interest.

7.3 Goal: Probes can be examined at a variety of scales (ex. Regions, Deployments, Tablet Groups (Conditions), Individuals)

Technically, much work needs to occur to redesign the architecture to scale to extremely large sizes of deployments (1000+ or 100000+ tablets, for example). From both a protocol and methodology perspective, providing support for that many users will be difficult. Furthermore, the technology should be buttressed to provide for more server and network infrastructure to filter the enormous amount of data that could be collected. This might involve providing standard network architectures of failover servers along with deploying hierarchical data caches in locales. Work towards providing true support for more devices without requiring highly customized builds will be mandatory.

GlobalLit Live provided some support for studying individuals, but those interactions were short (minutes) and in the unrealistic lab setting. Our aim is to support tablet data collection in the field, but it is nearly impossible to ensure that a given tablet is tied to a specific child. Indeed, we presumed that tablets were shared among siblings or classmates regularly. With careful experiment design using hierarchical statistics, the effects of sharing could be mitigated, but it would also be powerful to focus on individual progress using the devices. This would require inventing some mechanism to uniquely identify a child using a tablet.

In addition, future tablets should incorporate literacy assessment tools, like the matching app, more closely with the app experience. Arguably, this data is extremely important for answering questions about the efficacy of this approach when it comes to actual learning. User studies should be con-

ducted to discover how frequently to make assessments without disrupting the child's engagement with the tablet.

7.4 Contributions and Outlook

Despite some clear areas of necessary improvement, this system served an important first step in conducting impactful research on child-driven learning.

- The GlobalLit system currently runs on hundreds of tablets with hundreds of children in many communities on multiple continents. Children in Ethiopia, South Africa, India, Bangladesh, Uganda, and the United States have been using the system and the data they are contributing to the research is incredibly valuable. These active deployments are still generating new data to explore.
- GlobalLit is a unique system that focuses on data capture of the holistic and comprehensive behavior of children on tablets. Previous data collection systems were app-specific and closed; GlobalLit's design accounts for the sampling and exploration behavior that children exhibit as they use a variety of apps.
- GlobalLit is the first system of its kind to incorporate both school and home learning and in a variety of unusual circumstances. Researchers can study child-driven learning on mobile devices in an air-conditioned classroom in Georgia but also in an electricity-free hut in Ethiopia.
- The thesis describes design goals and presents wisdom about technical challenges for tablet-based educational deployments in the field.
- The this presents a case study showing how the GlobalLit system allows for the collection of data in the field for within and between deployment analysis.
- This thesis also contributed example studies generating evidence for specific beneficial app interactions (i.e., that virtual characters in apps can encourage engagement and dialogic reading behaviors).

As the system evolves and continues to deploy around the world, new lessons will emerge, improving our understanding of how children acquire literacy skills using these devices.

8 References

- Aharony, N., Pan, W., Ip, C., Khayal, I., & Pentland, A. (2011). Social fMRI: Investigating and shaping social mechanisms in the real world. *Pervasive and Mobile Computing*, 7(6), 643–659.
- Ali, S. M. (2013). Challenges and Benefits of Implementing Tablets in Classroom for e-Learning in a K-12 Education Environment Case Study of a School in United Arab Emirates. *Research Inventy: International Journal of Engineering and Science*, 3(4).
- Ally, M. (2014). Large Scale Deployment of Tablet Computers in K-12 Schools in Brazil. Retrieved from <http://auspace.athabascau.ca/handle/2149/3413>
- Alonso, J. B., Chang, A., & Breazeal, C. (2011). Values Impacting the Design of an Adaptive Educational Storybook. In *Interactive storytelling* (pp. 350–353). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Ananian, C. S., Ball, C. J., & Stone, M. (2012). Growing Up With Nell : A Narrative Interface for Literacy, 228–231.
- Barker, R. G. (1965). Explorations in ecological psychology. *American Psychologist*, 20(1), 1–14.
- Barker, R. G., & Wright, H. F. (1951). *One boy's day; a specimen record of behavior*. Harper.
- Barker, R. G., & Wright, H. F. (1954). *Midwest and its children: The psychological ecology of an American town*. New York, NY, US: Row, Peterson; Company.
- Bird, S. (2006). NLTK: the natural language toolkit. In *Proceedings of the cOLING/ACL on interactive presentation sessions* (pp. 69–72). Association for Computational Linguistics.
- Blewitt, P., Rump, K. M., Shealy, S. E., & Cook, S. a. (2009). Shared book reading: When and how questions affect young children's word learning. *Journal of Educational Psychology*, 101(2), 294–304.
- Blom-Hoffman, J., O'Neil-Pirozzi, T., Volpe, R., Cutting, J., & Bissinger, E. (2007). Instructing Parents to Use Dialogic Reading Strategies with Preschool Children. *Journal of Applied School Psychology*, 23(1), 117–131.
- Boteanu, A., & Chernova, S. (2012). Modeling Topics in User Dialog for Interactive Tablet Media. In *Eighth artificial intelligence and interactive digital*

entertainment conference.

- Brannon, D., & Dauksas, L. (2012). Studying the Effect Dialogic Reading Has on Family Members' Verbal Interactions during Shared Reading. *SRATE Journal*, 21(2), 9–20.
- Bus, A. G., IJzendoorn, M. H. van, & Pellegrini, A. D. (1995). Joint Book Reading Makes for Success in Learning to Read: A Meta-Analysis on Intergenerational Transmission of Literacy. *Review of Educational Research*, 65(1), 1–21.
- Carrot App. (2010). App protector pro [Software].
- CastingWords. (2013). [Web Service]. Retrieved from <http://castingwords.com/>
- Chang, A., & Resnick, M. (2011). *TinkRBooks : Tinkerable Story Elements for Emergent Literacy by TinkRBooks : Tinkerable Story Elements for Emergent Literacy* (PhD thesis). Massachusetts Institute of Technology.
- Chang, A., Nunez, D., Roberts, T., Sengeh, D., & Breazeal, C. (2013). Pre-pilot findings on developing a literacy tablet. In *IDC '13: Proceedings of the 12th international conference on interaction design and children* (pp. 471–474). New York, New York, USA: ACM Request Permissions.
- Chang, A., Tilahun, L., & Breazeal, C. (2014). Visualisations of Data from the Literacy Tablet Reading Project in Rural Ethiopia. In *Electronic visualisation and the arts 2014*. BCS, The Chartered Institute for IT.
- Chiong, C., & Shuler, C. (2010). *Learning: Is there an app for that? Investigations of young children's usage and learning with mobile devices and apps*. New York: The Joan Ganz Cooney Center at Sesame
- Chiong, C., Ree, J., Takeuchi, L., & Erickson, I. (2012). *Print Books vs. E-books: Comparing parent-child co-reading on print, basic, and enhanced e-book platforms*.
- chrony. (n.d.). [Software]. Retrieved from <http://chrony.tuxfamily.org/>
- Couse, L. J., & Chen, D. W. (2010). A Tablet Computer for Young Children? Exploring its Viability for Early Childhood Education. *Journal of Research on Technology in Education*, 43(1), 75–96.
- Cutting, J. (2011). *The use of video-based instruction, performance feedback and role play in teaching caregivers of preschool aged children to use dialogic reading strategies* (PhD thesis). Northeastern University.

- Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers*, 28(1), 125–127.
- Darren Murtha Design. (2003). Drawing pad [Software]. Retrieved from <https://play.google.com/store/apps/details?id=org.pogi.DrawingPad>
- Dornbach, P. (2003). Zebra paint coloring app [Software]. Retrieved from <https://play.google.com/store/apps/details?id=com.dornbachs.zebra>
- Duursma, E., Augustyn, M., & Zuckerman, B. (2008). Reading aloud to children: the evidence. *Archives of Disease in Childhood*, 93(7), 554–557.
- Fleer, M., & others. (2013). Digital positioning for inclusive practice in early childhood: The cultural practices surrounding digital tablets in family homes. *Computers in New Zealand Schools*, 25(1-3), 56–76.
- Fletcher, J. D., Tobias, S., & Wisher, R. A. (2007). Learning Anytime, Anywhere: Advanced Distributed Learning and the Changing Face of Education. *Educational Researcher*, 36(2), 96–102.
- Forget-Dubois, N., Dionne, G., Lemelin, J.-P., Pérusse, D., Tremblay, R. E., & Boivin, M. (2009). Early Child Language Mediates the Relation Between Home Environment and School Readiness. *Child Development*, 80(3), 736–749.
- funf | Open Sensing Framework. (n.d.). [Software]. Retrieved from <http://funf.org/>
- Gonzalez, M. (2014). App evaluation: UDL tool 1.0. In *Society for information technology & teacher education international conference* (Vol. 2014, pp. 1809–1811).
- Gonzalez, M., & Fryer, C. (n.d.). The effect of the iPad on the readiness skills of at-risk preschoolers. In *Society for information technology and teacher education international conference* (pp. 4380–4383).
- Goodwin, K., & Highfield, K. (2012). iTouch and iLearn: an examination of educational Apps. *Early Education and Technology for*
- Google Translate Web Application. (n.d.). [Web Service]. Retrieved from <http://translate.google.com/>
- Guernsey, L., Levine, M., Chiong, C., & Severns, M. (2012). *Pioneering literacy in the digital wild west: Empowering parents and educators*. Joan Ganz Cooney Center.

- Gutnick, A. L., Robb, M., Takeuchi, L., Kotler, J., Bernstein, L., & Levine, M. H. (2011). Always connected: The new digital media habits of young children. Joan Ganz Cooney Center at Sesame Workshop.
- Harfield, A., Jormanainen, I., Rungrattanaubol, J., & Pattaranit, R. (2013). An Open Monitoring Environment for primary school children engaged in tablet-based learning. *The 2013 10th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, 195–199.
- Hart, B., & Risley, T. R. (2003). The early catastrophe: The 30 million word gap by age 3. *American Educator*.
- Haugland, S. W. (1999). What Role Should Technology Play in Young Children's Learning? Part 1. *Young Children*, 54(6), 26–31.
- Hisrich, K., & Blanchard, J. (2009). Digital Media and Emergent Literacy. *Computers in the Schools*, 26(4), 240–255.
- Huebner, C. E., & Meltzoff, A. N. (2005). Intervention to change parentchild reading style: A comparison of instructional methods. *Journal of Applied Developmental Psychology*, 26(3), 296–313.
- Ifenthaler, D., & Schweinbenz, V. (2013). The acceptance of Tablet-PCs in classroom instruction: The teachers perspectives. *Computers in Human Behavior*, 29(3), 525–534.
- Kinyanjui, M. (2014). 'Kids on the Tab': Kenya National Library Service's tablet computers project for slum school children.
- Koyalee Chanda, Bruce Caines, Elizabeth Holder, Jonathan Judge, Nancy Keegan (Directors), Nickelodeon (Producer). (2003). *Blue's Clues - Blue Takes You to School* [DVD].
- Kucirkova, N. (2013). Children's interactions with iPad books: research chapters still to be written. *Developmental Psychology*, 4.
- La Paro, K. M., & Pianta, R. C. (2000). Predicting Children's Competence in the Early School Years: A Meta-Analytic Review. *Review of Educational Research*, 70(4), 443–484.
- Lonigan, C. J., & Whitehurst, G. J. (1998). Relative efficacy of parent and teacher involvement in a shared-reading intervention for preschool children from low-income backgrounds. *Early Childhood Research Quarterly*, 13(2), 263–290.

- Luyen. (2003). Kid coloring [Software]. Retrieved from <https://play.google.com/store/apps/details?id=com.vanluyen.KidColoringNoAds>
- Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data*. New York, NY: Psychology Press.
- McManis, L. D. (2012). Achieving School Readiness & Success Using Tablets and Content. Hatch Innovations.
- Miller, E. B., & Warschauer, M. (2013). Young children and e-reading: research to date and questions for the future. *Learning, Media and Technology*, 39(3), 283–305.
- Mintz, J., & Aagaard, M. (2012). The application of persuasive technology to educational settings. *Educational Technology Research and Development*, 60(3), 483–499.
- Mol, S. E., Bus, A. G., Jong, M. T. de, & Smeets, D. J. H. (2008). Added Value of Dialogic ParentChild Book Readings: A Meta-Analysis. *Early Education and Development*, 19(1), 7–26.
- Mori, K., Ballagas, R., Reville, G., Raffle, H., Horii, H., & Spasojevic, M. (2011). Interactive rich reading: enhanced book reading experience with a conversational agent. In *Proceedings of the 19th ACM international conference on multimedia - mM '11* (pp. 825–826). New York, New York, USA: ACM Press.
- Neumann, M. M., & Neumann, D. L. (2014). Touch Screen Tablets and Emergent Literacy. *Early Childhood Education Journal*, 42(4), 231–239.
- Parish-Morris, J., Mahajan, N., Hirsh-Pasek, K., Golinkoff, R. M., & Collins, M. F. (2013). Once Upon a Time: ParentChild Dialogue and Storybook Reading in the Electronic Era. *Mind, Brain, and Education*, 7(3), 200–211.
- Pillinger, C., & Wood, C. (2014). Pilot study evaluating the impact of dialogic reading and shared reading at transition to primary school: early literacy skills and parental attitudes. *Literacy*, 48(3), 155–163.
- Plowman, L., Mcpake, J., & Stephen, C. (2008). Just picking it up? Young children learning with technology at home. *Cambridge Journal of Education*, 38(3), 303–319.
- Priyankara, K. W. T. G. T., Mahawaththa, D. C., Nawinna, D. P., Jayasundara, J. M. A., Tharuka, K. D. N., & Rajapaksha, S. K. (2013). *Android based e-Learning solution for early childhood education in Sri Lanka*. IEEE.

- Raffle, H., Ballagas, R., Revelle, G., Horii, H., Follmer, S., Go, J., ... Spasojevic, M. (2010). Family story play: reading with young children (and elmo) over a distance. In *CHI '10: Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1583–1592). New York, New York, USA: ACM Request Permissions.
- Reese, E., Sparks, A., & Leyva, D. (2010). A Review of parent interventions for preschool childrens language and emergent literacy. *Journal of Early Childhood Literacy*, 10(1), 97–117.
- Repetti, R., Wang, S.-w., & Saxbe, D. (2009). Bringing It All Back Home: How Outside Stressors Shape Families' Everyday Lives. *Current Directions in Psychological Science*, 18(2), 106–111.
- Roy, D., Patel, R., Decamp, P., Kubat, R., Fleischman, M., Roy, B., ... Gorniak, P. (2006). The Human Speechome Project Stepping into the Shoes of Children.
- Ruttkay, Z., Bényei, J., & Sárközi, Z. (2014). Evaluation of Interactive Children Book Design. In *Methodologies and intelligent systems for technology enhanced learning* (pp. 109–117). Cham: Springer International Publishing.
- Sabar, A. (2014). *The Outsider: The Life and Times of Roger Barker*. Kindle Single.
- Shuler, C., Levine, Z., & Ree, J. (2012). iLearn II: An analysis of the education category of Apple's app store. *New York: The Joan Ganz*
- Takeuchi, L. (2011). Families matter: Designing media for a digital age. New York: The Joan Ganz Cooney Center at Sesame Workshop.
- Tsai, P.-s., & You, M. (2013). Usability analysis of children's iPad electronic picture books. In *ACHI 2013, the sixth international conference on advances in computer-human interactions* (pp. 269–273).
- Twining, P., Cook, D., Ralston, J., & Selwood, I. (2005). Tablet PCs in schools: Case study report.
- Unity3d. (n.d.). [Software]. Retrieved from <http://unity3d.com/>
- Verenikina, I., & Kervin, L. (2011). iPads, digital play and pre-schoolers. *He Kupu*.
- Viriyapong, R., & Harfield, A. (2013). Facing the challenges of the One-Tablet-Per-Child policy in Thai primary school education. *Education*, 4(9).
- Walling, D. R. (2014). Tablet Take-Home Strategies. In *Designing learning for*

- tablet classrooms* (pp. 113–118). Cham: Springer International Publishing.
- Wasik, B. (2008). When Fewer Is More: Small Groups in Early Childhood Classrooms. *Early Childhood Education Journal*, 35(6), 515–521.
- Wasik, B. A., & Iannone-Campbell, C. (2012). Developing Vocabulary Through Purposeful, Strategic Conversations. *The Reading Teacher*, 66(4), 321–332.
- Wasik, B. A., Bond, M. A., & Hindman, A. (2006). The effects of a language and literacy intervention on Head Start children and teachers. *Journal of Educational*.
- Wells, G. (1985). *Language development in the pre-school years*. CUP Archive.
- Whitehurst, G. J., Arnold, D. S., Epstein, J. N., Angell, A. L., & al, et. (1994). A picture book reading intervention in day care and home for children from low-income families. *Developmental Psychology*, 30(5), 679–689.
- Whitehurst, G. J., Falco, F. L., & Lonigan, C. J. (1988). Accelerating language development through picture book reading. *Developmental Psychology*, 24(4), 552–559.
- Wolf, M., & Gottwald, S. (2013). *Android App Survey (Unpublished)*.
- Wolf, M., Gottwald, S., Galyean, T., Morris, R., & Breazeal, C. (2014). The Reading Brain, Global Literacy, and the Eradication of Poverty. In *Bread and brain, education and poverty* (pp. 1–22). Vatican City: Pontifical Academy of Sciences.
- Wolf, M., Miller, L., & Donnelly, K. (2000). Retrieval, automaticity, vocabulary elaboration, orthography (RAVE-O): a comprehensive, fluency-based reading intervention program. *Journal of Learning Disabilities*, 33(4), 375–386.
- Zevenbergen, A. A., Whitehurst, G. J., Van Kleeck, A., Stahl, S. A., & Bauer, E. B. (2003). Dialogic reading: A shared picture book reading intervention for preschoolers. In *On reading books to children: Parents and teachers* (pp. 177–200).

Appendix A Sample Probe Data

edu.media.mit.prg.matchingapp

```
name,timestamp,funf_file_id,value,device_id, id,data_uuid
edu.media.mit.prg.matchingapp,2013-09-06 18:04:52,1715370,"{"
  player_choice":["brown","Image","brown","WordAndImage","
  correct"]}],244,10638266,db292d42-b005-4520-bed3-5adde6b8c8b4_39
edu.media.mit.prg.matchingapp,2013-09-06 18:04:53,1715370,"{"
  player_choice":["yellow","WordAndImage","yellow","Image
  ","correct"]}],244,10638267,db292d42-b005-4520-bed3-5
  adde6b8c8b4_40
edu.media.mit.prg.matchingapp,2013-09-06 18:04:54,1715370,"{"
  all_items":["blue","red","red","brown","white","white
  ","black","brown"]}],244,10638268,db292d42-b005-4520-bed3-5
  adde6b8c8b4_41
```

edu.mit.media.funf.bgcollector.MainPipeline

```
name,timestamp,funf_file_id,value,device_id, id,data_uuid
edu.mit.media.funf.bgcollector.MainPipeline,2013-04-24
15:33:14,1831025,"{
  ""dataUploadUrl"": ""http://worldliteracy.media.mit.edu/gsu/
  data/?pwd=ekjfsad"",
  ""configUpdateUrl"": ""http://worldliteracy.media.mit.edu/gsu/
  config/index.html"",
  ""dataRequests"": {
    ""edu.mit.media.funf.probe.builtin.RunningApplicationsProbe"":
      [
        {
          ""PERIOD"": 60
        }
      ],
    ""edu.mit.media.funf.probe.builtin.BatteryProbe"": [
      {}
    ],
    ""edu.mit.media.funf.probe.builtin.HardwareInfoProbe"": [
      {
```

```

        ""PERIOD"": 86400
    }
],
    ""edu.mit.media.funf.probe.builtin.ScreenProbe"": [
        {}
    ]
},
    ""version"": 1,
    ""name"": ""gsu v2""
}","37,13081865,5fe27015-fa7f-4be8-b9a8-b3f0d0fb4b70_1

```

edu.mit.media.funf.probe.builtin.ActivityProbe

```

name,timestamp,funf_file_id,value,device_id, id,data_uuid
edu.mit.media.funf.probe.builtin.ActivityProbe,2012-09-10
    21:19:57,6,"{
    ""HIGH_ACTIVITY_INTERVALS"": 0,
    ""PROBE"": ""edu.mit.media.funf.probe.builtin.ActivityProbe"",
    ""TOTAL_INTERVALS"": 15,
    ""TIMESTAMP"": 1347326397,
    ""LOW_ACTIVITY_INTERVALS"": 0
}"
,1,11,4b5d94ee-184f-42b9-8059-a2a65233c0a2_1

```

edu.mit.media.funf.probe.builtin.BatteryProbe

```

name,timestamp,funf_file_id,value,device_id, id,data_uuid
edu.mit.media.funf.probe.builtin.BatteryProbe,2013-04-13
    08:19:48,1824849,"{
    ""icon-small"": 17302707,
    ""scale"": 100,
    ""present"": true,
    ""technology"": ""Li-ion"",
    ""level"": 83,
    ""voltage"": 8081,
    ""status"": 4,
    ""PROBE"": ""edu.mit.media.funf.probe.builtin.BatteryProbe"",
    ""invalid_charger"": 0,
    ""plugged"": 0,
    ""TIMESTAMP"": 1365855588,
    ""health"": 2,
    ""temperature"": 221
}"

```

```
}",48,12807207,49fb0953-cdd1-4472-a1bf-442d441ff1a3_56
```

edu.mit.media.funf.probe.builtin.HardwareInfoProbe

```
name,timestamp,funf_file_id,value,device_id, id,data_uuid
edu.mit.media.funf.probe.builtin.HardwareInfoProbe,2013-04-13
  14:11:14,1824979,"{
    ""PROBE"": ""edu.mit.media.funf.probe.builtin.HardwareInfoProbe"",
    ""WIFI_MAC"": ""3C:43:8E:ED:25:16"",
    ""MODEL"": ""Xoom"",
    ""TIMESTAMP"": 1365876674,
    ""BRAND"": ""motorola"",
    ""ANDROID_ID"": ""ccb2e1a21d9f2fe""
  }",38,12812927,630637c9-2c4b-421b-ac71-e617e150c4b7_16
```

edu.mit.media.funf.probe.builtin.LightSensorProbe

```
name,timestamp,funf_file_id,value,device_id, id,data_uuid
edu.mit.media.funf.probe.builtin.LightSensorProbe,2012-09-10
  22:28:26,4,"{
    ""SENSOR"": {
      ""RESOLUTION"": 0.05,
      ""POWER"": 0.003,
      ""NAME"": ""Ambient Light sensor"",
      ""VERSION"": 1,
      ""TYPE"": 5,
      ""MAXIMUM_RANGE"": 208076.8,
      ""VENDOR"": ""Maxim""
    },
    ""PROBE"": ""edu.mit.media.funf.probe.builtin.LightSensorProbe"",
    ""ACCURACY"": [
      0
    ],
    ""TIMESTAMP"": 1347330506,
    ""LUX"": [
      0.0
    ],
    ""EVENT_TIMESTAMP"": [
      39831863873000
    ]
  }",1,6,442292d5-0897-488d-91bd-26bfb65927a1_3
```

edu.mit.media.funf.probe.builtin.RunningApplicationsProbe

```
name,timestamp,funf_file_id,value,device_id, id,data_uuid
edu.mit.media.funf.probe.builtin.RunningApplicationsProbe,2013-04-13
    08:52:51,1824849,"{
    ""TIMESTAMP"": 1365857571,
    ""PROBE"": ""edu.mit.media.funf.probe.builtin.
        RunningApplicationsProbe"",
    ""RUNNING_TASKS"": [
        {
            ""baseActivity"": {
                ""mClass"": ""com.android.launcher2.Launcher"",
                ""mPackage"": ""com.android.launcher""
            },
            ""topActivity"": {
```

edu.mit.media.funf.probe.builtin.ScreenProbe_appended

```
name,timestamp,funf_file_id,value,screen_on,funf_timestamp,data_uuid,
    id,device_id
edu.mit.media.funf.probe.builtin.ScreenProbe,2013-04-13
    09:12:12,1824858,"{
    ""TIMESTAMP"": 1365858732,
    ""PROBE"": ""edu.mit.media.funf.probe.builtin.ScreenProbe"",
    ""SCREEN_ON"": true
    }",True,1365858732,9d3c680e-9ac1-464a-9191-ef6edf6eff47_44,12807687,1
```

edu.mit.media.funf.probe.builtin.ScreenProbe

```
name,timestamp,funf_file_id,value,device_id,id,data_uuid
edu.mit.media.funf.probe.builtin.ScreenProbe,2013-04-13
    09:12:12,1824858,"{
    ""TIMESTAMP"": 1365858732,
    ""PROBE"": ""edu.mit.media.funf.probe.builtin.ScreenProbe"",
    ""SCREEN_ON"": true
    }",1,12807687,9d3c680e-9ac1-464a-9191-ef6edf6eff47_44
```

FileMoverService

```

name,timestamp,funf_file_id,value,device_id,id,data_uuid
FileMoverService,2013-04-13 08:12:57,1824850,"{"type":"log", "msg
":"moved files automatically"}",1,12807223,9825d5e8-7598-45f6-
bb97-6977ae3403c5_1
FileMoverService,2013-04-13 08:06:29,1824852,"{"type":"log", "msg
":"moved files automatically"}",50,12807367,48ddc53a-de37-4dba-
a880-dbc8be179a8e_1
FileMoverService,2013-04-13 08:11:20,1824854,"{"type":"log", "msg
":"moved files automatically"}",15,12807490,7da05893-0120-461e-
a1ae-1d36628a1219_1

```

funfFileMover

```

name,timestamp,funf_file_id,value,device_id,id,data_uuid
funfFileMover,2012-09-10 14:23:30,3,moved files automatically,1,3,2
b00c036-0d21-49b2-b334-6d3213b13aee_1
funfFileMover,2012-09-10 14:48:56,8,moved files automatically,1,13,0
eeb882f-c544-4156-890e-e3eb16b5abf6_1
funfFileMover,2012-09-11 03:51:58,9,moved files automatically
,1,14,9420e583-d91e-46d1-b213-d095c1a78aec_1

```

GpsDateFixService

```

name,timestamp,funf_file_id,value,device_id,id,data_uuid
GpsDateFixService,1970-01-01 19:00:22,1827017,"{"type": "
InvalidDateDetected", "date": 86422}",1,12903977,5bbd335b-18f7
-4960-9327-0449cac90b57_297
GpsDateFixService,1970-01-01 19:01:52,1827017,"{"type": "
GpsDateFixOld", "date": 86512}",1,12903979,5bbd335b-18f7
-4960-9327-0449cac90b57_299
GpsDateFixService,1970-01-01 19:01:52,1827017,"{"type": "
GpsDateFixNew", "date": 1366989543}",1,12903980,5bbd335b-18f7
-4960-9327-0449cac90b57_300

```

LauncherApp

```

name,timestamp,funf_file_id,value,device_id,id,data_uuid
LauncherApp,2013-04-13 00:32:01,1825078,EZRA_JACK_KEATS5.m4v.mp4
,8,12816434,f5c10c68-f973-4ed6-bc48-5226cbbf3599_5
LauncherApp,2013-04-12 23:37:58,1825090,CHICKA_CHICKA_BOOM_BOOM5.mp4
,8,12817050,0e37c7cc-dcbf-4621-a00d-daae8c7522dd_3

```

LauncherApp,2013-04-13 11:19:36,1825165,MAXRUBY_SUMMER1.mp4
,8,12821349,6121fe8e-8b1c-4f1a-80df-aea35f1bd45c_38

LauncherAppPaused_appended

name,timestamp,funf_file_id,value,paused,data_uuid,id,device_id
LauncherAppPaused,2014-07-22 07:22:55,1876656,"{"paused": true}",
True,c926a820-a6c6-4e52-a6c3-87133e94cd4a_5,14239263,1
LauncherAppPaused,2014-01-25 19:45:40,1858636,"{"paused": true}",
True,30d814e6-b09a-47ce-98b1-7312dfd5c68a_3,14223776,1
LauncherAppPaused,2014-07-22 07:51:16,1876684,"{"paused": true}",
True,40e5d829-88d8-44c2-87f3-b5f0dff09318_14,14239304,1

LauncherAppPaused

name,timestamp,funf_file_id,value,device_id,id,data_uuid
LauncherAppPaused,2014-07-22 07:22:55,1876656,"{"paused": true
}",1,14239263,c926a820-a6c6-4e52-a6c3-87133e94cd4a_5
LauncherAppPaused,2014-01-25 19:45:40,1858636,"{"paused": true
}",1,14223776,30d814e6-b09a-47ce-98b1-7312dfd5c68a_3
LauncherAppPaused,2014-07-22 07:51:16,1876684,"{"paused": true
}",1,14239304,40e5d829-88d8-44c2-87f3-b5f0dff09318_14

tinkerbook

name,timestamp,funf_file_id,value,device_id,id,data_uuid
tinkerbook,2013-04-03 19:38:05,1825435,"{"type":"","SoundPlayed":"","
url":"","app:/audio/title/My_BabyD_by_Angela_Cynthia_and_Fardad.mp3
"}",50,12833093,50faea79-edaa-454a-9a8b-934aaf8633c8_81
tinkerbook,2013-04-03 19:38:05,1825435,"{"type":"","FingerUp":"","x
":1249.45,"y":380.7}",50,12833094,50faea79-edaa-454a-9a8b-934
aaf8633c8_82
tinkerbook,2013-04-03 19:38:05,1825435,"{"type":"","FingerDown":"","x
":347.6,"y":730.85}",50,12833095,50faea79-edaa-454a-9a8b-934
aaf8633c8_83

Appendix B Apps on Tablets

ID	Type	Name	Description
1	video	The Three-Legged Cat	video content (passive)
2	video	Little Tim and the Brave Sea Captain	video content (passive)
3	app	VocabuLarry's ABCs	letters and vocabulary
4	video	Caps for Sale	video content (passive)
5	video	The Little Red Lighthouse and the Great Grey Bridge	video content (passive)
6	video	Blue Goes to School 1	video content (passive)
7	app	Phonics Monster: Level 1, Unit 10	phonics curriculum
8	app	Phonics Monster: Level 1, Unit 2	phonics curriculum
9	app	Phonics Monster: Level 1, Unit 3	phonics curriculum
10	app	Phonics Monster: Level 1, Unit 4	phonics curriculum
11	app	Phonics Monster: Level 1, Unit 5	phonics curriculum
12	app	Phonics Monster: Level 1, Unit 6	phonics curriculum
13	app	Phonics Monster: Level 1, Unit 7	phonics curriculum
14	app	Phonics Monster: Level 1, Unit 8	phonics curriculum
15	app	Phonics Monster: Level 1, Unit 9	phonics curriculum
16	app	Phonics Monster: Level 2, Unit 10	phonics curriculum
17	app	Phonics Monster: Level 2, Unit 2	phonics curriculum
18	app	Phonics Monster: Level 2, Unit 3	phonics curriculum
19	app	Phonics Monster: Level 2, Unit 4	phonics curriculum
20	app	Phonics Monster: Level 2, Unit 5	phonics curriculum
21	app	Phonics Monster: Level 2, Unit 7	phonics curriculum
22	app	Phonics Monster: Level 2, Unit 8	phonics curriculum
23	app	Phonics Monster: Level 2, Unit 9	phonics curriculum
24	app	Phonics Monster: Level 3, Unit 10	phonics curriculum
25	app	Phonics Monster: Level 3, Unit 4	phonics curriculum
26	app	Phonics Monster: Level 3, Unit 5	phonics curriculum
27	app	Phonics Monster: Level 3, Unit 7	phonics curriculum
28	video	How Much is a Million	video content (passive)
29	video	Blueberries for Sal	video content (passive)
30	video	A Story A Story	video content (passive)

31	video	Millions of Cats	video content (passive)
32	video	Hondo and Fabian	video content (passive)
33	video	Mouse Around	video content (passive)
34	video	Hot Hippo	video content (passive)
35	video	Waiting for Wings	video content (passive)
36	app	My first Puzzles	puzzle games for motor skills
37	app	My First Tangrams	shapes and matching
38	app	Animal ABC	game about animals and the alphabet
39	app	VocabuLarry Things that Go	vocabulary
40	app	Solo	virtual music instrument
41	app	ZebraPaint	art and coloring
42	app	Beginning Blends	beginning blends word sounds
43	app	Sight Words	vocabulary and sight words game
44	app	Letter Tracing	letter writing and strokes
45	app	Monkey's Preschool Lunchbox	colors, letters, counting, differences, and matching games
46	app	Toddler Jukebox	plays selection of songs
47	app	Giraffe's Matching Zoo	mind and memory development game designed for children
48	app	Giraffe's Preschool Playground	ABCs and numbers, seek-and-find skills, and animal names
49	app	Kid Coloring	drawing and coloring book
50	app	Zoodles: Jack and the Beanstalk	interactive book
51	app	Zoodles: Little Red Riding Hood	interactive book
52	app	Zoodles: The Country Mouse and the City Mouse	interactive book
53	app	Zoodles: The Elves and the Shoemaker	interactive book
54	app	Zoodles: The Emperor's New Clothes	interactive book
55	app	Zoodles: The Three Little Pigs	interactive book
56	app	Zoodles: The Tortoise and the Hare	interactive book
57	app	Zoodles: The Velveteen Rabbit	interactive book
58	video	Corduroy	video content (passive)
59	video	Smile for Auntie	video content (passive)
60	video	The Happy Owls	video content (passive)

61	video	Goose	video content (passive)
62	video	Here Comes the Cat	video content (passive)
63	video	Frog Goes to Dinner	video content (passive)
64	app	TinkRBook	interactive book
65	video	Pet Show!	video content (passive)
66	video	A Letter to Amy	video content (passive)
67	video	The Trip	video content (passive)
68	video	Zin! Zin! Zin! A Violin!	video content (passive)
69	video	Elizabeth's Doll	video content (passive)
70	video	Time of Wonder	video content (passive)
71	video	The Foolish Frog	video content (passive)
72	video	Dot the Fire Dog	video content (passive)
73	video	I Want a Dog	video content (passive)
74	video	Angus Lost	video content (passive)
75	video	Danny and the Dinosaur	video content (passive)
76	video	Dinosaur Bones	video content (passive)
77	video	T is For Terrible	video content (passive)
78	video	I Love My ABCs - Animals	video content (passive)
79	video	Ban the Doughnut	video content (passive)
80	video	Noisy Nora	video content (passive)
81	video	The Little Red Hen	video content (passive)
82	video	Reading to Your Bunny	video content (passive)
83	video	Space Case	video content (passive)
84	video	Over in the Meadow	video content (passive)
85	video	Giraffes Can't Dance	video content (passive)
86	video	The Chinese Violin	video content (passive)
87	video	Changes Changes	video content (passive)
88	video	Dooby Dooby Moo	video content (passive)
89	video	Leap Frog Letter Factory	video content (passive)
90	video	LeapFrog Phonic Farm	video content (passive)
91	video	Little Einsteins	video content (passive)
92	video	Angus and the Ducks	video content (passive)
93	video	Cat and Canary	video content (passive)
94	video	Wings: A Tale of Two Chickens	video content (passive)
95	video	Meet the Letters	video content (passive)

96	video	Meet the Numbers	video content (passive)
97	video	Meet the Shapes	video content (passive)
98	app	Drawing Pad	creative drawing and art app
99	video	Night Light	video content (passive)
100	video	A Peep of a Different Color	video content (passive)
101	video	The Fish Museum	video content (passive)
102	video	Sounds Like..	video content (passive)
103	video	Peep Plants a Seed	video content (passive)
104	video	The Whatchamacallit	video content (passive)
105	video	The Great White Man-Eating Shark	video content (passive)
106	video	What's Under My Bed?	video content (passive)
107	app	Little Piano	virtual music instrument
108	video	Rapunzel	video content (passive)
109	video	The Three Billy Goats Gruff	video content (passive)
110	video	Lon Po PO (A Red-Riding Hood Story from China)	video content (passive)
111	app	Kids Alphabet	letters and soundboard
112	video	The Amazing Bone	video content (passive)
113	video	Let's Give Kitty a Bath	video content (passive)
114	video	Shrinking Violet	video content (passive)
115	video	Tinga Tinga 2-5	video content (passive)
116	video	Tinga Tinga 2-6	video content (passive)
117	video	Tinga Tinga 2-7	video content (passive)
118	video	Tinga Tinga 3-1	video content (passive)
119	video	Tinga Tinga 3-2	video content (passive)
120	video	Tinga Tinga 3-3	video content (passive)
121	video	Tinga Tinga 3-4	video content (passive)
122	video	Tinga Tinga 3-5	video content (passive)
123	video	Tinga Tinga 3-6	video content (passive)
124	video	Tinga Tinga 3-7	video content (passive)
125	video	Tinga Tinga - Why Hen Pecks the Ground	video content (passive)
126	video	Tinga Tinga - Why Snake has no Legs	video content (passive)
127	video	Bugs! Bugs! Bugs!	video content (passive)
128	video	Keeping House	video content (passive)
129	video	Chicken soup with rice	video content (passive)

130	video	One was Johnny	video content (passive)
131	video	Alligators All Around	video content (passive)
132	video	He's got the whole world in his hands (Song)	video content (passive)
133	video	Come on Rain	video content (passive)
134	video	Owl Moon	video content (passive)
135	video	Patrick	video content (passive)
136	app	IntelliJoy Shapes	jigsaw puzzles and shapes

Appendix C Crowd-Sourced Questions

- Why does someone think?
- Why do humans share?
- Why is a stick made of wood?
- Why is a book made of paper?
- What other things are like bugs and insects?
- Can a universe be huge?
- Why does a whale have a fin?
- Why does a finger have a knuckle?
- Why does a head have an ear?
- Why does a body have a head?
- Why does a computer have a keyboard?
- Why does a computer have a monitor?
- Why does a face have a mouth?
- Why does a hammer have a handle?
- Why does a bird have a wing?
- A book is a text. What other texts do you know?
- What other things are like hours and minutes?
- What other things are like hours and years?
- What other things are like minutes and years?
- Why does a shirt have a button?
- What other things are like centuries and minutes?
- What can something be besides purple and red?
- Why does a story have a chapter?
- Why does a bed have a pillow?
- What other things are like brothers and sisters?
- Why does a year have a month?
- Why does a body have an arm?
- Why does a finger have a nail?
- Why does a mouth have a tooth?
- Why does an car have a hood?
- Why does a day have a night?
- Can a chocolate be white?
- Why does someone hide?
- Why do kids explore?
- Why do people study?
- Why does a basket have a handle?
- Why is a chair made of wood?
- Can a computer be useful?
- Why do dogs run?
- Why does someone yawn?
- Why do balls roll?
- Why do kids play?
- What are nine and six?
- A hop is a jump. What other jumps do you know?
- Why does a finger have a fingerprint?
- Why does a school have a classroom?
- What can something be besides purple and yellow?
- Why does a house have a wall?
- Can a night be scary?
- Why do scientists experiment?
- Why does a body have a heart?
- A watch is a clock. What other clocks do you know?
- Why does a book have a page?
- Why does a head have an eye?
- Can a birthday be special?
- Why does a phone have a bell?
- What are nine and three?
- Why does a couch have a cushion?
- Why does an arm have an elbow?
- Why does a door have a handle?
- Why does a text have a chapter?
- Can a person be happy?
- Why does a finger have a fingernail?
- Why does a hand have a finger?
- Why do frogs jump?
- Why does someone burp?
- Why does an alphabet have a letter?
- A kindergarten is a school. What other schools do you know?
- Why does a phone have a keypad?
- Why do whales swim?
- Can a child be funny?
- What are one and two?
- Why does a couch have an upholstery?
- Why do dogs sit?
- Why do humans eat?
- What can something be besides purple and yellow?
- Why does an institution have a department?
- Why do people sleep?
- Why is a bread made of flour?
- Can a person be sad?
- Can a universe be strange?
- A bug is a insect. What other insects do you know?
- Why does a shirt have a sleeve?
- Why do teachers educate?
- Why do humans want?

- What can something be besides pink and red?
- Why do children skip?
- What can something be besides pink and purple?
- Why does someone teach?
- What are nine and ten?
- What are nine and seven?
- Why do people agree?
- Why do balls pop?
- What are six and ten?
- Can a birthday be happy?
- Why do dogs jump?
- A heart is a muscle. What other muscles do you know?
- Why do trees fall?
- Why do humans talk?
- Why do frogs swim?
- Why does a book have a chapter?
- Can a health be important?
- Why do friends talk?
- Can a chocolate be yummy?
- Can a person be funny?
- Can a chocolate be sweet?
- Can a person be special?
- What are eight and six?
- What can something be besides pink and purple?
- Why do humans run?
- What are eight and nine?
- What can something be besides red and yellow?
- What can something be besides pink and red?
- What can something be besides purple and red?
- Why does a hand have a palm?
- Why does someone type?
- Can a chocolate be brown?
- Why is a pillow made of feathers?
- Can a life be good?
- Can a person be kind?
- Can a sea be blue?
- What other things are like centuries and hours?
- Can a food be yummy?
- Why does someone count?
- Can a person be greedy?
- Why does someone jump?
- Why do actors pretend?
- What other things are like centuries and years?
- Can a friend be important?
- Why does a tooth have a gum?
- Can a disaster be bad?
- Can a person be different?
- Why do paints dry?
- Can a child be small?
- Why do people think?
- What are six and three?
- Why does a tree have a leaf?
- Why do dogs want?
- Can a car be useful?
- Can an art be colorful?
- Why do people jump?
- Why does a chapter have a paragraph?
- Why do dogs bite?
- Why do people laugh?
- Why does a bathroom have a bathtub?
- Why does a laptop have a screen?
- Why does a plant have a leaf?
- What can something be besides red and yellow?

Appendix D Colophon

I authored this document using the Markdown markup language by John Gruber and \LaTeX . My writing tools were Sublime Text and Rúben Cabaço's Byword. For data exploration and visualization, I relied on iPython Notebook and various libraries provided by SciPy along with IBM's SPSS. I used Marked 2 by Brett Terpstra for work-in-progress rendering and Pandoc for final document conversion. *Papers*³ helped me manage my references and bibliography. I used a self-hosted git repository to manage changes and versions of the document. Spotify provided my working soundtrack, with a heavy rotation of focus-inspiring music from artists like The American Dollar, M83, Junip, Halou, Helios, Mogwai, The Echelon Effect, and Hammock.

The body text is set in Palatino, headlines are set in Avenir Black and Avenir Book, and source code text is set in Inconsolata.

Icons used to create the Heterogeneous Connectivity figure include Antenna by Diego Naive, Database by Stefan Parnarov, Exclamation Point by MarkieAnn Packer, Router by Andrey Mo, RSS by P.J. Onori, School by Chris Cole, SD Card by Thomas Le Bas, Sneaker by Simon Child, Tablet by Pham Thi Dieu Linh, Target by Karsten Barnett, Thermometer by Jaime Carrion, Watch by Valerio Poltrini, Wireless by Thomas Uebe, and others in the Public Domain, all from The Noun Project.