

© Copyright 2019
Hyewon Suh

Supporting Long-Term Health Monitoring by
Reducing User Burden and Enhancing User Benefit

Hyewon Suh

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2019

Reading Committee:

Julie A. Kientz, Chair

Jennifer Turns

Sean A. Munson

Program Authorized to Offer Degree:
Human Centered Design and Engineering

University of Washington

Abstract

Supporting Long-Term Health Monitoring by
Reducing User Burden and Enhancing User Benefit

Hyewon Suh

Chair of the Supervisory Committee:

Julie A. Kientz

Human Centered Design and Engineering

The prevalence of smartphones and advances in low cost monitoring sensors have dramatically increased self-tracking in recent years. Among many other areas of self-tracking, health and wellness is one area that has received particular attention. However, along with this enthusiasm and growth, there exists a skepticism regarding these technologies and their ability to inspire and sustain individuals' engagement with their everyday health and wellness in the long term. Many people fail to adopt, or they abandon health technologies after very little use, even if they are seeing benefits of use. Depending on the domain, lapsing with or abandoning health-tracking may not result in serious consequences, but it can become critical when it comes to child development tracking, leaving many missed opportunities. Thus, I decided to investigate the domain of child development monitoring, as a first step, because of its importance as a public health goal.

In this dissertation I examine the feasible ways technology can reach families for regular child development tracking; how technology may impose burdens and benefits on its users; and how we can reduce user burden and enhance user benefits of child development tracking technology in support of long-term tracking. I report first on the design, development, and evaluation of two such systems, “@BabySteps” and “Baby Steps Text”, in which I investigated challenges and opportunities in child development tracking and explored how everyday technologies parents already use can be used to create an interactive public health system. Through these studies I was able to understand the feasibility and acceptability of social media, web portal and text messaging for child development tracking, as well as burdens and benefits parents experienced in their use. This triggered me to explore burdens and benefits associated with technology use in general and to develop the User Burden Scale and User Benefit Scale. Based on this background work, I examined the effects of specific design features in reducing burdens and increasing benefits in child development tracking system and evaluated the Baby Steps system by deploying it with 139 families over 20 months.

Across these studies, I found that reducing user burden and enhancing user benefit in tracking technology can engage, inform, educate, and empower parents and make monitoring more accurate and meaningful. . I hope that this work will shed new light on how health tracking technologies can be designed to support awareness and long-term engagement in their users.

DEDICATION

To my parents Clare Lim & Simon Suh,
for their never-ending support, encouragement and love

TABLE OF CONTENTS

List of Figures	vi
List of Tables	vii
Chapter 1. Introduction.....	14
1.1 Background and Motivation	14
1.1.1 Child Development Tracking.....	15
1.1.2 Keeping Motivation for Tracking	17
1.2 Thesis Statements and Research Questions	18
1.3 Thesis Overview	19
1.4 Contributions.....	21
1.4.1 Artifact Contributions	21
1.4.2 Methodological & Theoretical Contributions	21
1.4.3 Empirical Research Contributions.....	22
Chapter 2. Related Work	23
2.1 Technology and Healthcare	24
2.1.1 Use of SMS-based mHealth in Healthcare	24
2.1.2 Benefits of interactive SMS-based mHealth.....	26
2.2 Technologies for the Health of Young Children	27
2.3 User Burden and Benefit in Technology	28
2.3.1 Theoretical Framework	28
2.3.2 Evaluation of Computing Systems.....	29
2.3.3 Measurements for Different Types of Usability Issues and User Burden.....	30
2.3.4 Measurements for Different Types of User Benefits and Values.....	32

Chapter 3. @BabySteps.....	34
3.1 @BabySteps System Design.....	35
3.1.1 Functionality and Usage	35
3.1.2 Twitter System Implementation	37
3.1.3 Companion Website	38
3.2 Deployment Study Design	39
3.2.1 Study Procedure.....	40
3.2.2 Participants	41
3.3 Results	44
3.3.1 Overall Tweeting Activity	44
3.3.2 Using Twitter Syntax.....	46
3.3.3 Tweeting in Public	47
3.3.4 Social Interactions	48
3.3.5 Twitter and Website Integration.....	49
3.3.6 Overall Feedback	50
3.4 Discussion	51
3.5 Chapter 3 Summary	54
Chapter 4. Baby Steps Text.....	55
4.1 Preliminary Work & System Design.....	56
4.1.1 Design Methods	56
4.2 System Design & Implementation	61
4.2.1 System Design	61
4.2.2 System Implementation	64

4.3	Methods.....	65
4.3.1	Participants	65
4.3.2	Study Procedure.....	67
4.4	Results	67
4.4.1	Baby Steps Text Usage	67
4.4.2	Use of Baby Steps Text Syntax	69
4.4.3	Participants' Experience with Baby Steps Text.....	69
4.4.4	Continued Usage After Study Completion.....	70
4.5	Discussion	70
4.5.1	Recommended Changes to Baby Steps Text Design	70
4.5.2	Limitations	72
4.6	Chapter 4 Summary	72
	Chapter 5. Understanding User Burdens in Technology.....	74
5.1	Background.....	75
5.2	Initial Definition of User Burden.....	76
5.3	Developing User Burden Scale	77
5.3.1	Refining Initial User Burden via Interviews	77
5.3.2	Identification & Development of Scales	78
5.3.3	Principal Component Analysis	80
5.4	Final User Burden Constructs & Scale.....	82
5.4.1	Definitions of User Burden Constructs.....	82
5.4.2	Final Scale and Question Set	83
5.5	Validating User Burden Scale	85

5.5.1	Methods.....	85
5.5.2	Results.....	86
5.6	Discussion.....	88
5.6.1	UBS Scoring and Analysis and Guidelines for Use	89
5.6.2	Limitations	91
5.7	CHAPTER 5 summary	92
Chapter 6. Understanding User Benefitss in Technology.....		93
6.1	Initial Definition of User Benefit.....	94
6.2	Developing User Benefit Scale.....	95
6.2.1	Refining User Benefit via Interviews	95
6.2.2	Identification & Development of Scales	97
6.2.3	Principal Component Analysis	98
6.3	Final User Benefit Constructs & Scale	100
6.3.1	Definitions of User Benefit Constructs	100
6.3.2	Final Scale and Question Set	101
6.4	Validating User Benefit Scale.....	101
6.4.1	Methods.....	101
6.4.2	Results.....	104
6.5	Discussion.....	106
6.5.1	How to Administer the User Benefit Scale.....	106
6.5.2	Scoring Guidelines.....	107
6.5.3	Limitations	109
6.6	Chapter 6 summary	110

Chapter 7. Baby Steps Longitudinal Study	111
7.1 System Design	112
7.1.1 Web Portal.....	114
7.1.2 Text Messaging.....	115
7.2 Method	116
7.2.1 Participants	116
7.2.2 Study Procedure.....	117
7.3 Results	119
7.3.1 Overall Activity – Between Group Comparison	119
7.3.2 Overall Activity – Experimental Group (Phase 1 vs Phase 2)	124
7.3.3 Features Exclusive to Experimental Group.....	126
7.3.4 Findings from Surveys and Interviews	129
7.4 Discussion	132
7.4.1 Recommended Changes to Baby Steps.....	132
7.5 Chapter 7 Summary	135
Chapter 8. Discussion.....	137
8.1 Lessons learned	138
8.1.1 Active Health Monitoring and Patient Empowerment.....	138
8.1.2 Burdens and Benefits in Health Monitoring.....	139
8.2 Contributions.....	140
8.3 Future Work	141
8.4 Conclusion	142
References	144

Appendix A.	162
Appendix B.	172
Appendix C.....	173

LIST OF FIGURES

Figure 1. Value-based Adoption Model	29
Figure 2. @BabySetps: Exemplary milestone question tweet.....	36
Figure 3. @BabySetps: Exemplary response from users	36
Figure 4. @BabySetps: Simplified diagram of how @BabySteps works	38
Figure 5. @BabySetps: Baby Steps website design.....	39
Figure 6. @BabySetps: Study procedure.....	40
Figure 7. Baby Steps Text: User-centered design process.....	57
Figure 8. Baby Steps Text: Example persona for a parent of a child with special needs.	59
Figure 9. Baby Steps Text: Storyboard of Baby Steps Text used at the Latina health fair (in Spanish)	59
Figure 10. Baby Steps Text: Example script for "Wizard" in Wizard-of-Oz prototype	60
Figure 11. Baby Steps Text: Example of Baby Steps Text syntax usage.....	62
Figure 12. Baby Steps Text: Study Procedure	67
Figure 13. User Burden Scale: Identification and development process of UBS	Error! Bookmark not defined.
Figure 14. User Burden Scale: Validation process of UBS	85
Figure 15. User Burden Scale: UBS score distribution (n=750)	90
Figure 16. User Benefit Scale: Identification and development of UBeS.....	98
Figure 17. User Benefit Scale: Validation process of UBeS	103
Figure 18. User Benefit Scale: UBeS score distribution	108
Figure 19. Longitudinal Study: Baby Steps Feature differences between groups.....	116
Figure 20. Longitudinal Study: Participant details	119
Figure 21. Longitudinal Study: Timeline of Baby Steps longitudinal study by group	119
Figure 22. Longitudinal Study: Milestone response pattern for each group	121
Figure 23. Longitudinal Study: Usage of web portal and text messaging by participants over time	125

LIST OF TABLES

Table 1. Research Questions.....	19
Table 2. UBS-related literature based on the six constructs of user burden.....	32
Table 3. @BabySteps: Participant details	43
Table 4. @BabySteps: Overall usage.....	45
Table 5. Baby Steps Text: Participant details	66
Table 6. Baby Steps Text: Overall text activity	68
Table 7. User Burden Scale: Sample interview prompts and coded phrases for each initial category of user burden	79
Table 8. User Burden Scale: Factor loadings and communalities based on a principal component analysis (PCA) with varimax rotation for 26 items.....	81
Table 9. User Burden Scale: User Burden Scale items and categories.....	84
Table 10. User Burden Scale: List of systems participants rated using UBS	86
Table 11. User Burden Scale: Subscale reliability using Cronbach's Alpha coefficient	87
Table 12. User Burden Scale: Comparing different computing systems based on UBS score using Independent-Samples t-Test.....	88
Table 13. User Burden Scale: Score guidelines for UBS	89
Table 14. User Benefit Scale: Sample interview prompts and coded phrases for each initial category of user benefit	96
Table 15. User Benefit Scale: Factor loadings and communalities based on principal components analysis (PCA) with varimax rotation on 36 items	99
Table 16. User Benefit Scale: User Benefit Scale categories and questions.....	102
Table 17. User Benefit Scale: List of popular systems evaluated with UBeS	103
Table 18. User Benefit Scale: Scale reliability using Cronbach's alpha coefficient.....	104
Table 19. User Benefit Scale: Comparing different computing systems based on UBeS score using Independent-Samples t-Test.....	106
Table 20. User Benefit Scale: Score guidelines for UBeS	108

Chapter 1. Introduction

1.1 BACKGROUND AND MOTIVATION

The prevalence of smartphones and advances in low cost monitoring sensors have dramatically increased self-tracking in recent years [149]. Among many other areas of self-tracking, health and wellness is one area that has received particular attention. According to one report [29], one in ten Americans over the age of 18 owns an activity tracker and the number grows every year. However, along with this enthusiasm and growth, there exists a skepticism regarding these technologies and their ability to inspire and sustain individuals' engagement with their everyday health and wellness in the long term [73]. Many

people still fail to adopt systems that could benefit them greatly or abandon systems after very little use, even if they are seeing a benefit. A recent study suggests that more than half of people who own activity trackers no longer use them, and a third of people who own trackers stop using them within only six months [29]. Researchers in the field have explored reasons why people lapse with or stop using health tracking tools and the common reasons were: difficulties in remembering to track [30, 61]; losing motivation or getting worn out [79]; discomfort with the information revealed [31]; and finding health tracking frustrating or time-consuming [4, 48], all falling under the umbrella that the burden these systems place on the user is too high.

Depending on the domain, lapsing with or abandoning self-tracking may not result in serious consequences, but it can become critical when it comes to child development, leaving many missed opportunities. Among many different domains of health monitoring, I decided to investigate the domain of child development monitoring as a first step because of its importance as a public health goal [24, 122].

1.1.1 Child Development Tracking

Approximately 1 in 6 children between the ages of 3 and 18 in the United States are diagnosed with developmental disorders, such as autism, ADHD, or anxiety disorders [26]. It is known that the earlier these disorders can be diagnosed, the sooner early interventions can start, resulting in better outcomes for these children [52]. Unfortunately, approximately half of children who eventually receive diagnoses of developmental disorders are not identified until they reach kindergarten in the United States [23], with lower-income populations being missed at a higher rate (e.g., [56]). Tracking and regular screening of child development is the key to early detection of many types of developmental disabilities.

The current approach for screening for developmental progress is to ask parents a series of questions about their children's abilities in areas such as gross motor coordination, communication, and problem solving. For example, "does your baby say two syllable sounds, such as 'da-da' or 'ga-ga'?" or "can your child walk upstairs while holding onto a railing?" A popular screener tool, the Ages and Stages Questionnaire (ASQ), consists of 22 questionnaires with 30 questions each across the child's first five years [36]. The screener can either be conducted via paper or through a website and can be administered

through Well Child Visits with the child's pediatrician, by childcare providers, or by parents self-monitoring their child's development. There are some limitations to using milestone questionnaires for screening, such as cultural competence, access by more diverse families and/or the potential for parent anxiety when tracking development. However, developmental questionnaires are currently the most effective method for screening according to the American Academy of Pediatrics [19].

The problem with paper-based or even web-based screener surveys is that it is often difficult to ensure that parents complete each survey, especially since there are more screener questionnaires than scheduled Well Child Visits during the first 5 years of a child's life. Given that Well Child Visits are months to up to a year apart depending on the child's age, it is not the timeliest solution to only conduct screening through pediatrician visits. In addition, when parents answer questions in a single sitting, such as in a pediatrician's waiting room, they may not have the opportunity to try each activity with their child if the question asks about an activity that does not often occur or needs time to try out, such as drawing shapes on a piece of paper or interacting with a mirror. Also, busy parents often just forget to keep track of milestones using both paper and web-based systems [103]. What is more concerning is parents without access to developmental screening often have difficulties in monitoring because they may not be aware of what to watch for. And many resource-constrained populations, such as low-income or immigrant families, lack the insurance to cover Well Child Visits, and frequent moves or language barriers make it difficult to access information about normal child development.

Tools that can help prompt parents to answer questions at more regular intervals can possibly result in more complete and accurate responses, better parent involvement, and parent awareness of their child's activities. Especially because child development takes place in daily life outside of traditional clinical settings, parents' and caregivers' reports for the early signs of developmental delay are the most reliable way to track them. Therefore, it is important to engage, inform, educate, and empower parents to track and assess their child's development in an easy way in order to make monitoring more accurate and meaningful. The human-computer interaction community has recently begun developing such software (e.g., [102]). However, there are limitations with standalone software applications in that they require the parent to remember to open the application to answer questions. Even with proactive

notifications or email reminders, parents may still not engage as easily and as often as they might otherwise. Because these tools are intended to be used over a long period (e.g., up to 5 years), they require methods to maintain parent engagement.

1.1.2 Keeping Motivation for Tracking

For every child to be screened, leaving no child behind, it is also crucial to engage all caregivers, even the ones that are less motivated by the concept of health and wellness. Different individuals are motivated by different goals; in the case of health monitoring, goals may include maintaining one's own or another's health state or increasing wellness [43] but this may not be of interest for all. Therefore, it is necessary to maximize perceived benefits of use. One possible way to achieve this can be to frame and promote health monitoring besides just a way to gain deeper insight and discovery into their everyday health (e.g., [3, 173]). In case of child development monitoring, adding sentimental recording to make it more celebratory, positive, and a memory-making process can be one solution.

Also, a recent study [167] suggests that integrating a health intervention into social sites that people already visit can increase adherence. Although parents are often too busy to remember to complete baby books or other manual tracking systems [102], they are already using social networking systems to share information on their children [135]. However, more research is needed to understand what types of user burdens and benefits exist in health technologies and how to address them in the design process.

Thus, the need to design and develop a child development tracking system with thoughtful design features to reduce user burden and enhance user benefit becomes more apparent. With the growth of personal everyday technologies, the method of child development tracking can go beyond paper-based screening and blend into parents' lives, such as through mobile phones or social media, which parents already use and engaged with.

My dissertation aims to investigate the following: i) exploring the feasibility and acceptability of different technologies in child development monitoring; ii) understanding user burdens and benefits in technology use; and iii) evaluating how design might support long-term engagement. By learning about

the benefits and burdens of a specific health technology that requires regular monitoring over a long period, I anticipate it can possibly be generalized across other health domains.

1.2 THESIS STATEMENTS AND RESEARCH QUESTIONS

The claims of my dissertation can be summarized in the following thesis statements:

- T1** I propose that understanding how technology can be used and accepted in health tracking is necessary for identifying design opportunities, potential burdens, and potential benefits in supporting users' regular and continuous monitoring.
- T2** I propose that understanding how technology can impose burdens and/or provide benefits to its users is necessary to identify design opportunities and support its users.
- T3** I propose that adding specific technology design features that specifically aim to i) reduce user burdens and/or ii) enhance user benefits will increase users' awareness, engagement, and enjoyment in a health tracking application for the both the short-term and long-term.

To verify the thesis statements, I examined the following research questions (RQs) through a mixed method approach:

- RQ1** What are the feasible ways technology can reach families for regular child development tracking?
- RQ2** How does technology impose burdens on its users and how can we measure them?
- RQ3** How does technology provide benefits to its users and how can we measure them?
- RQ4** How can we design technology to support child development tracking in the long term?

I examined research questions **RQ1** to support **T1**, where I explored feasibility of technologies in child development tracking. To support **T2**, I investigated research questions **RQ2** and **RQ3**, in which I identify, develop and validate two new ways of assessing user burdens and benefits in technology. And finally, I investigated **RQ4** to address **T3**, in which I designed and developed an ecosystem of child development tracking intended to engage parents over a long-term period.

Table 1. Research Questions

	@BabySteps	Baby Steps Text	User Burden Scale	User Benefit Scale	Baby Steps Longitudinal study
RQ1. What are the feasible ways technology can reach families for regular child development tracking?	x	x			x
RQ2. How does technology impose burdens on its users and how can we measure them?	x	x	x		x
RQ3. How does technology provide benefits to its users and how can we measure them?	x	x		x	x
RQ4. How can we design technology to support child development tracking in the long term?					x

I answered these research questions through five different investigations (Table 1). To answer **RQ1**, I conducted two deployment studies. My goal was to study challenges and opportunities in child development tracking and explore how technology could be used to create an interactive public health system, supporting parents. I then designed and developed the @BabySteps and Baby Steps Text systems and deployed them to gauge feasibility and acceptability of social media and text messaging for child development tracking respectively. To address **RQ2** and **RQ3**, I identified burdens and benefits associated with technology use and developed the User Burden Scale and User Benefit Scale respectively. In addressing **RQ4**, I studied the effects of specific design features on reducing burdens and increasing benefits in a child development tracking system over 20 months.

1.3 THESIS OVERVIEW

This dissertation is divided into eight chapters.

In Chapter 2, *Related Work*, I summarize the prior work I leveraged in conducting investigations in this dissertation. I begin with the use of everyday technologies to promote healthcare and benefits of using technologies in health interventions. As the main body of my work is on child development tracking, I also summarize literature on technologies to support the health and wellbeing of young children and their families. I then describe the theoretical framework around user burden and benefit, as well a work that investigated evaluating computing systems and concepts relating to user burden and benefit.

In Chapter 3, *@BabySteps: How Social Media Can Be Used for Child Development Tracking*, I describe the design and implementation details for @BabySteps, a Twitter-based system, that allows parents to closely monitor their child's development by responding to milestone question prompts and checking their progress on a companion website. I also report on the results of a 3-week deployment study and design opportunities revealed from the study, which I build on in the following chapters.

In Chapter 4, *Baby Steps Text: How Text Messaging Can Be Used for Child Development Tracking*, I present feasibility and acceptability of text messaging in child development tracking by conducting a 4-week deployment study on Baby Steps Text, an automated, personalized, and self-contained two-way communication text messaging system. This work became a foundation of the longitudinal study I present in Chapter 7.

In Chapter 5, *Understanding User Burdens in Technology*, I seek to answer the questions 'what is user burden?' and 'how can we evaluate user burden systematically?' by designing and validating a new scale called the User Burden Scale (UBS). I report on the process of scale development and validation in detail.

In Chapter 6, *Understanding User Benefits in Technology*, I build on the work from Chapter 5 and develop a complementary model and measurement that focuses on positive aspects of computing systems, *user benefit*. I describe the development and validation process of the User Benefit Scale (UBeS), a new 25-item scale with seven subscales designed to measure the user benefit in a wide variety of computing systems.

In Chapter 7, Baby Steps longitudinal Study, I describe the design and deployment study of Baby Steps, a tool I created by building upon on the learnings from Chapters 3 and 4. I report on the quantitative and qualitative analysis of a 20-month longitudinal study with 139 families to understand the effectiveness of the system in supporting child development tracking in the long term.

In Chapter 8, I summarize the contributions of my thesis and opportunities for future work.

1.4 CONTRIBUTIONS

This dissertation presents three types of research contributions in the field of HCI: artifacts, methodological, theoretical, and empirical research contributions [74]. Specifically, my work has generated the following contributions.

1.4.1 Artifact Contributions

To investigate **T1**, I designed, developed, and evaluated three methods of child development tracking using social media (@BabySteps), text messaging (Baby Steps Text) and a website (Baby Steps Web Portal). @BabySteps makes a novel contribution because use of social media for public health had not been fully explored prior to this. Also, Baby Steps Text is a completely automated, personalized, and self-contained two-way communication system. It makes a unique contribution as it brings SMS-based mHealth interventions to a new domain (childhood developmental screening) and helps expand knowledge about how parent-generated health data can be collected in a simple but thoughtful way, thus expanding the reach of its benefits. Later, I combined Baby Steps Text and Web Portal to investigate **T3** and this ecosystem of tools works seamlessly together.

1.4.2 Methodological & Theoretical Contributions

Two of the investigations I presented in this dissertation make methodological and theoretical contributions. In an effort to address **T2**, I have identified the concept and constructs of user burden and benefit to lay the ground work for building new theory in this space. Also, by creating two quick and easy measurements for assessing user burdens and benefits in technology, I provide useful methods that can provide reliable, valid results to make a methodological contribution to human-computer interaction researchers.

1.4.3 Empirical Research Contributions

In each investigation in this dissertation, I used empirical methods to collect data. In addressing **T1** and **T3**, I used automated logging to see how parents use (or do not use) technology to track their child's development. I also used interviews and surveys to generate a deeper understanding of how parents felt about using technology to monitor child development, how their child development monitoring behavior has changed before and after the study, and concerns parents may have. From results of both quantitative and qualitative data, I created system design guidelines such as: i) Making screening activities more celebratory and less "medical" can motivate engagement, ii) Meeting parents where they are instead of requiring them to come to you can increase engagement. In addition to this, I have conducted a 20-month long deployment study to investigate **T3**. This study enabled me to see changes in adoption, tracking-adherence and health goal awareness over a long period of time. As our field evolves at a fast pace, it is rare to find long-term studies that run for this long. Thus, the approaches and strategies I used in this study can provide a guidance to researchers who seek to execute long-term studies. In addressing **T2**, I present the empirical finding of surveys on user burdens and user benefits in everyday technologies and I also created scoring guidelines based on the initial data sets.

Chapter 2. Related Work

In this chapter, I discuss background and work related to my dissertation. I begin with the use of everyday technologies to promote health care and its benefits. Because my work is particularly related to child development tracking, I also summarize literature on technologies to support the health and well-being of young children and their families. I then describe the theoretical framework on how user burden and benefit might impact the behavioral aspects of technology adoption, as technology adoption and continued usage is important in health tracking. Finally, I report on prior work that investigated evaluating computing systems and concepts relating to user burden and benefit. This section serves as an overview of related work in this area and will be expanded further for the dissertation.

2.1 TECHNOLOGY AND HEALTHCARE

In this section, I describe existing work that examines the ways technology might support health care using everyday technology that is widely adopted. Here, I review some of the domains of SMS-based mHealth and its benefits.

2.1.1 Use of SMS-based mHealth in Healthcare

The ubiquity and penetration of mobile phones presents an opportunity, particularly in under-resourced health eco-systems, via SMS (Short Message Service), a low-cost communication channel that is supported on even the most basic of mobile phones. Between 2000 and 2012, the number of mobile phones in use worldwide grew from fewer than 1 billion to around 6 billion [194]. In the United States, 91% of adults own a cell phone as of 2013 [130]. This widespread distribution of mobile phones across diverse populations may solve one of the most difficult problems in the public health domain – barriers to access. Thus, mHealth (mobile health) can make it possible to inform caregivers and let them collect and share relevant data at any time, allowing more rapid convergence on optimal treatment.

In recent years, public health outreach efforts have benefited greatly from the large-scale communication afforded by SMS messaging. With the rising penetration of mobile phones, public health campaigns are able to communicate with populations via SMS messaging, including traditionally considered hard-to-reach populations due to scale, geographic dispersal, limited technology use, or other concerns. Research has tackled different aspects of health care by integrating SMS with a variety of health interventions and existing literature suggests that SMS-based mHealth supports the following four health domains.

Disease Management & Treatment Adherence

The use of text messaging to reinforce treatment adherence is one of the most widely used applications of mHealth. The most common areas are disease management for chronic disease such as HIV / AIDS [25, 84, 119, 140, 180], diabetes [115, 165, 172, 191], and asthma [117, 157, 176, 188, 189]. Programs such as txt2stop [20] and a study by Rodgers [9] promote adherence for smoking cessation treatment. Other

programs have been developed to improve treatment adherence for domains such as weight loss [114], bulimia nervosa [160], mental illness [27, 42], and medication adherence [154, 177].

Health Information Delivery and Behavior Change

Recently, text messaging has been recognized as a relatively inexpensive tool for behavior change and the delivery of health information [62]. Text4Baby is one example of using text messaging for health promotion and behavior change that has been studied under various circumstances [41, 101, 159, 193]. Although also in a similar domain of text messaging for early childhood, Text4Baby focuses on one-way informational messages about general maternal and child health, rather than developmental screening. There is also an emerging body of evidence that supports the use of text messaging for other health behavior changes in the areas of vaccination for travelers [8], youth sexual health [34], weight management [11, 113], and encouraging physical activity [18, 88].

Patient-Provider Communication

mHealth is also effectively taking advantage of the mobile phone as a communication channel between patients and providers. Many studies have investigated the impact of short message reminders on increasing the likelihood of attending clinic appointments [116, 151, 181]. In addition to this, Mäkelä [110] conducted a pilot study of SMS communication between patients and providers in clinical psychiatry, and Kummervold [146] tested the use of SMS to communicate health information to visually impaired patients.

Early Detection & Screening

mHealth technology has been successfully used to promote screening for early detection of disease. Lee and his colleagues [64] developed and tested a 7-day mobile phone text message-based cervical cancer Screening intervention. Meanwhile, Khokhar [5] used SMS to encourage and remind women to do their monthly breast self-exam during their menstrual cycle. However, despite advantages of SMS in delivering information at the appropriate time, use of SMS for screening purpose has not been as well-explored as other areas. Baby Steps Text fits into this domain most closely and addresses a new area within it for children's developmental screening.

2.1.2 Benefits of interactive SMS-based mHealth

There are two ways of communicating using SMS-based mHealth depending on the communication type i) one-way communication and ii) two-way communication. One-way SMS communication is similar to mass media in that it “pushes” information to subscribers’ phones by using messages tailored to personal needs. A large body of SMS-based research has typically been designed this way. This includes simple reminder programs such as reminding of clinic appointments and other systems such as sending health promotion and dietary and disease management information. In one-way SMS systems, the subscriber is more of a passive recipient rather than an active user. To engage people more actively, some projects used two-way communication in their communication message design. For users, the interactive nature of two-way communication may require greater effort and generate greater interest [111]. While some projects support a fully natural way of communication by having people responding to questions and concerns sent by users [42, 154], other projects support fully automated communication by requiring users to follow a specific structure to enable automated parsing by the system [21, 49]. Recently, Perrier et. al. [186] deployed a hybrid two-way communication where the system sends out bulk messages to its users automatically, and a human behind the scenes takes care of user replies.

Several previous studies also have shown the benefits of individually tailored messages that are personalized for the user [64, 140, 165]. Thus, with Baby Steps Text, we tailored each message to be as specific to the child as possible. With automated personalization that takes into account each user’s specific context (e.g. age-appropriate questions with the child’s name), Baby Steps Text can convey a feeling of personal attention cost-effectively. Given that there are limited resources and time to engage in direct communication with individual users using human-generated messages, the Baby Steps Text model can be a good compromise to achieve a fully personal, two-way SMS communication. We developed a two-way, interactive text messaging system that sends child developmental screening questions to caregivers to increase their awareness and engagement. Baby Steps Text is unique in that all milestone questions are personalized to increase user attachment and entirely automated to run without behind-the-scenes human effort.

2.2 TECHNOLOGIES FOR THE HEALTH OF YOUNG CHILDREN

Designing interactive technologies to support the health and wellbeing of young children and their families has been a growing area of interest in the field of human-computer interaction. D'Alessandro & Dosa [40] describe how information technology can be empowering to parents and children, and thus there is great motivation to use it to support their health. For the past few decades, researchers have sought to design technologies in support of children's health and have primarily focused on general studies on the opportunities for design of technologies for children's health, technologies for use of parents and doctors, and technology to be used by children themselves.

Tang et al. [109] developed a mobile application called Estrellita for observing premature infants' daily living activities (such as feeding and diaper changes), prompting parents to do exercises at home, and reminders of upcoming appointments. Jeong et al. studied parents with the purpose of providing insights for the design of technology for tracking general health data about children [63]. Similarly, Fitbaby [53] helps parents of premature infants track their child's physical exercises and overall wellbeing. There are some commercial applications, such as Trixie Tracker [187] and LENA [124], which also allow for parents to track information about feeding and diapers and tracking language respectively. However, these systems focus on general health observations and are not intended for screening, as is the goal of Baby Steps.

When it comes to tracking and screening, there are researchers who have sought to design technologies to automatically assess a child's development. For example, researchers have used toys with sensors [38, 44, 185] to measure grasping actions of infants or computer vision to evaluate the position of an infant's leg and monitor the kicking patterns of infants [137]. Also, studies have used robots to evaluate children's sociability and detect autism spectrum disorders [100, 118] and there is a tool for monitoring emerging speech in young children called VisiBabble [60]. However, these automated child development assessments are often restricted by technical boundaries resulting in the accuracy of detection being fairly limited. In addition, because they tend to focus on one domain of child development, it is hard to capture the whole picture of child development, still leaving a questionnaire-based child development screening as a solid method.

In the domain of screening and tracking child development using interactive computing systems, Kientz et al. have done extensive research over time. She conducted a formative study on the opportunities for the design of developmental tracking systems [103] and then later conducted an in-depth evaluation of the opportunities for supporting parent-pediatrician interaction [104]. These two prior formative works have informed the original Baby Steps desktop software application, which helps parents track developmental progress at home. Recently, there was an attempt to tackle child development tracking with the help of Curated Crowd Intelligence (CCI). Baby CROINC [14] is an online early-childhood development tracker that relies on CCI to curate personalized inputs and connect with the crowd's aggregate data and to provide parents with objective and personalized feedback on their children's development. Also, Song et al. developed BebeCODE [168], a mobile system that encourages parents to independently answer all developmental questions for a given age and resolve disagreements through chatting, image/video sharing, or asking a third person.

The motivation and design of the Baby Steps ecosystem presented in this dissertation was inspired directly by the series of formative studies conducted by Kientz et al. I seek to build upon this previous literature by exploring a new delivery mechanism for developmental screening.

2.3 USER BURDEN AND BENEFIT IN TECHNOLOGY

In this section, I summarize the framework around how user burden and benefit might impact the behavioral aspects of technology adoption. I then report on prior work that investigated evaluating computing systems and concepts relating to user burden and benefit.

2.3.1 Theoretical Framework

Human behavior and decision-making theories such as the Technology Acceptance Model (TAM) and the Value-based Adoption Model (VAM) helped me understand the behavioral aspects of technology adoption. TAM, a popular framework developed by Davis [17], suggests that people's adoption of a new technology is dependent on their intention to use it, which is in turn dependent on their perception of the 1) technology's usefulness and 2) ease of use [16, 17]. To account for affective value that is missing from TAM, VAM brings to light the important role of affective value in consumer adoption of new

technology. According to VAM, consumers choose whether to adopt a technology or service by weighing the products' perceived cognitive and affective benefits such as usefulness or enjoyment against perceived sacrifices such as fees or technicality, including mental or physical effort (Figure 1). The VAM's perceived sacrifice and benefit framework is closely related to user burden and benefit as I define them. VAM argues that in order to raise perceived value and maximize people's intention to adopt them health monitoring technologies need to increase benefit while decreasing sacrifice (burden). My dissertation will be drawn from this and explore it further. For example, though VAM stops at predicting intention for adoption, I am interested to explore beyond adoption and how perceived user burden and benefit affects lapses or abandonment of health monitoring technologies. In addition, I aim to understand types of user burdens and benefits in technology use.

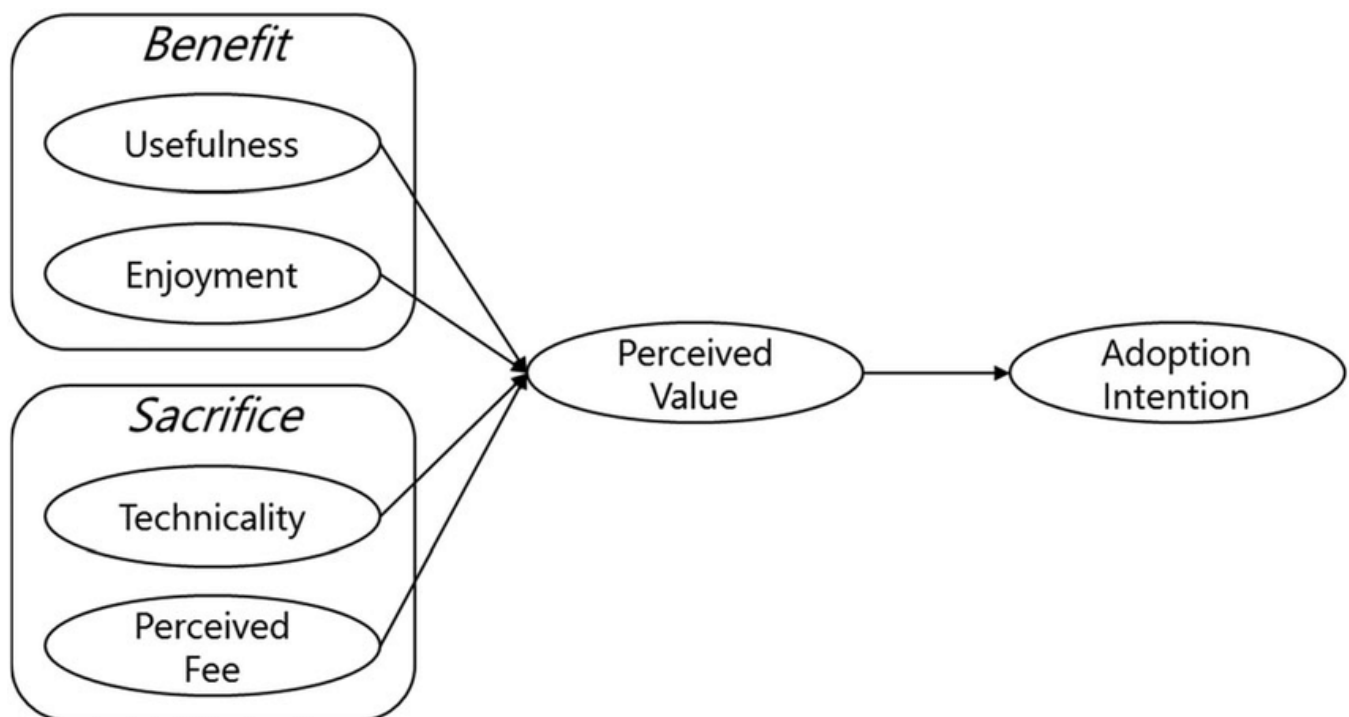


Figure 1. Value-based Adoption Model

2.3.2 Evaluation of Computing Systems

In the field of HCI, there have been many projects aimed at evaluating computing systems from the user's perspective. Usability is one of the most well-known and well-defined concepts in human-computer interaction (HCI) research. It was originally defined as the degree of efficiency and effectiveness of the

system [12, 17, 92, 93] and has been emphasized to be an important factor in making a successful system [6, 83, 97, 136]. In many early usability studies, the evaluators collected specific measures, such as ease of learning, efficiency of use, memorability, and speed and accuracy in performing a task, all of which are believed to reflect aspects of the usability of the system. Based on the collected measures, the evaluators make a conclusion that a certain system has good usability and is ready to be adopted by users or that one version of a system is better than others in terms of one measure or another. However, improving the system in terms of the objective measures does not always mean that the users are thoroughly satisfied with the system.

To overcome the limitations of objective measures, researchers in the field tried to extend the usability concept to emphasize the subjective aspects, including emotional and behavioral factors. Because user preference and satisfaction are not as easy to measure directly, researchers have developed a wide variety of evaluation instruments and user-completed scales. Some of these instruments include the Questionnaire for User Interaction Satisfaction (QUIS) [94], the Computer User Satisfaction Inventory [106], the NASA Task-Load Index (NASA-TLX) [164], the Software Usability Measurement Inventory (SUMI) [107], the System Usability Scale (SUS) [91], the Purdue Usability Testing Questionnaire (PUTQ) [58], and the IsoMetrics Usability Inventory [57]. However, no previous scales have attempted to characterize all aspects of user burden.

2.3.3 Measurements for Different Types of Usability Issues and User Burden

The study of user burden owes much to research into usability. In this section, I describe various instruments used to evaluate usability, which have been developed and had their reliability validated over time [78]. In light of other growing concerns over user experience beyond usability, I also include additional literature beyond usability instruments to provide a more holistic picture of user burden.

Many existing instruments use the concept of *cognitive* load as the primary lens to evaluate usability [47, 121]. In questionnaires such as the SUS [91] or NASA-TLX [164], subjects are asked to self-report on their experiences with the product in terms of how *mental*, *physical*, *emotional*, and *temporal*

factors contribute to their experience in using a particular system. For instance, emotion is measured in terms of when a user experiences frustration with a task they are asked to perform [89].

Growing awareness has led to the development of many tools to measure and evaluate usability in light of *accessibility* issues, as detailed by [161]. Guidelines like the ISO standards [87] and surveys like the Quebec User Evaluation of Satisfaction with Assistive Technology (QUEST) [127] provide designers and researchers methods for assessing user satisfaction from an accessibility perspective.

For the *financial* aspects of computing systems, quantitative models to estimate price sensitivity exist in the field of economics with relevance to HCI designers and software developers. Incorporating financial concerns, especially considering how it burdens a particular user, provides a more comprehensive assessment of user burden [155, 169]. In addition, the critical implications of *privacy* in relation to usability have given rise to various frameworks that can help researchers and designers make these issues more concrete, as demonstrated by [33, 82, 166].

The *social* lives of users are increasingly impacted by interactions with computing systems, and thus understanding social burdens is crucial. From the Internet to smartphones and social networking, these ongoing changes create new avenues for psychological researchers seeking to understand the influence these systems can have on the user's social relationships [132].

Through a review of the literature, we identified several characteristics or dimensions of usability related issues, which led to the initial user burden scale (Table 2). Although previous studies and theoretical frameworks provide dimensions regarding the subjective preferences for the different types of impacts that computing systems may have on the user, there has not yet been a systematic method for measuring user burdens. Accepting that one system has high user satisfaction is not directly translated to having no burden associated with using it. Thus, in Chapter 5, I attempt to define a new way of evaluating systems from a user burden point of view and develop a validated instrument that measures user burden.

Table 2. UBS-related literature based on the six constructs of user burden

Sub-scale	Related Literature
Difficulty of use	[58], [66], [87], [91], [94], [107], [127], [161]
Physical	[15], [71], [164]
Social & Time	[1], [35], [94], [98], [107], [132], [164], [182]
Mental & Emotional	[47], [57], [58], [89], [91], [94], [95], [107], [121], [164]
Privacy	[33], [55], [82], [126], [166]
Financial	[155], [169]

2.3.4 Measurements for Different Types of User Benefits and Values

The field of user experience (UX) has grown from concepts relating to usability [12] and ease-of-use, and thus there are numerous usability instruments assessing these aspects, including the Technology Acceptance Model (TAM) [51] and the System Usability Scale [91]. These metrics have been applied to various domains such as health systems [153], mobile banking [179], and eCommerce [131]. However, because the systems' usability and *ease-of-use* may vary across diverse demographics, more recent studies have examined ease-of-use of a technology for specific populations, such as older adults [184, 190] or people with disabilities [66, 161]. Although measures of *time* and *productivity* have long been a part of evaluating computing systems, it is still important to assess these metrics with more modern approaches. Gould et al. provide guidelines on how to make productivity-enhancing computer applications, and many usability measures [66, 99, 144] include questions around efficiency and effectiveness.

Because of the proliferation of new technologies, people have many options and discretion for adopting and abandoning various computing systems to meet their needs. Thus, it is essential to create computing systems that are more than *usable* and provide *enjoyable*, *engaging*, and *memorable* experiences. Previous research has advocated for the importance enhancing experience in the design of computing technologies [13, 145, 195], and Norman [39] has questioned the traditional view of usability and highlighted the importance of *aesthetic* aspects of user interfaces as well as users' emotional responses. Multiple researchers in the field have also investigated how aesthetics affects users' loyalty

[37], trust [196] of a technology, and overall satisfaction [80, 128]. A recent study by Cheng et al. [108] presented a user benefit survey that measures enjoyable, useful, usable, and *social* benefits of the system.

Looking at some of the other aspects of user benefit, previous research has explored how *financial* aspects of computing systems, such as the price to purchase and cost it takes to maintain, affects the user experience. Robert et al. [155] provided a model incorporating a price sensitivity measurement to software engineering, and usability metrics by Nielsen [76] include items related to the financial value of a system. There are also many technologies that support the *health and wellbeing* of their users, and they cover a variety of domains such as disease management and treatment adherence [20, 188, 191], health information delivery and behavior change [7, 193], patient-provider communication [28, 110], early detection and screening [65, 68], tracking activity, nutrition [2, 70], and more. Metrics related to health outcomes are common in the medical field, and a system that users value and benefit from can help achieve those outcomes.

Another significant way people benefit from technology is the value it gives in providing *information or awareness*. The reliability and credibility of content can greatly impact users' evaluation of a given technology. Across many domains, from online health information [163], to food-related hazards [170], to political information [183], researchers have investigated the reliability of information and its influence on users' perception and adoption. Likewise, from the Internet, to messenger apps, to social networking, computing systems are essentially reshaping and influencing the social lives of users and creating social value. Studies by Ducheneaut [141] and Ellison [142] show the quality of online social relationships and the benefits they can bring. There are also numerous online support groups about health conditions and social ties made online can have positive impacts on people's lives [133, 178], so it is an important consideration in the overall picture of user value.

While these previous studies and frameworks have incorporated aspects of user value, there still remains an opportunity to have a unified and systematic method for understanding and measuring user value across a broad spectrum and Chapter 6 will address that.

Chapter 3. @BabySteps: How Social Media Can Be Used for Child Development Tracking

As part of investigating **T1**, I examined how different everyday technologies can be used and accepted in health tracking. Due to its nature of easy accessibility and popularity in everyday life, Twitter became a key candidate to explore. In this chapter, I describe the design and implementation details for @BabySteps, a Twitter-based system that was designed to be a component of part of a larger suite of tools. @BabySteps allows parents to closely monitor their child's development by responding to milestone question prompts and checking their progress on a companion website. The design of @BabySteps was based on a series of formative studies ([63, 102, 103]), with additional input from a study on the use of social networking by mothers of young children [135]. After designing and developing the @BabySteps

system, I ran a 3-week deployment study with 14 parent participants. This deployment study provided me with empirical data that is relevant to the broader questions of **RQ1** (what are the feasible ways technologies can reach families for regular child development tracking?) **RQ2** (how does technology impose burdens on its users and how can we measure them?) and **RQ3** (how does technology provide benefits to its users and how can we measure them?) It allowed me to evaluate part of **T1** by revealing design opportunities, potential burdens and potential benefits of social media in child development tracking. I worked together with a team of 3 collaborators on this investigation.

3.1 @BABYSTEPS SYSTEM DESIGN

3.1.1 Functionality and Usage

To use the @BabySteps system, users first register in the system with their name, email address, and Twitter handle. For each of their children, they provide a name, gender, and date of birth. If a user is uncomfortable with their child's real first name being used in public communications, they are able to provide a pseudonym that can be used instead. After registration, a user follows the main Baby Steps Twitter account (@BabyStepsUW), which is used to send out announcements and study information to all participants and is manually operated by the research team. They also follow a Twitter account associated with their child's birth month, and if applicable, grant permission for it to follow them back. So, for example, a user with a child born in August 2017 would follow @BabyStepsAug17.

At regular intervals, this latter account sends out tweets asking age-appropriate milestone questions about the child's development. Each of these tweets contains a unique closing hashtag that the system refers to as a milestone ID (see Figure 2 for an example).

Users respond to these questions either publicly by posting to their timeline or privately via direct message to the account that originally posted the question. In either case, the user's response must contain a minimum of two specific elements: 1) the milestone ID hashtag from the prompt they are responding to and 2) their response to the question in the form of a hashtag – either #yes, #sometimes, or #no / #notyet (these two are handled identically). For users who have multiple children registered in the system, they must additionally include the relevant child's first name or pseudonym as a hashtag.

Beyond these required elements, users are free to include any other free text or additional hashtags in their tweet as well. This allows them to naturally include their response in an unassuming tweet that makes sense to their own followers, should they choose to do so. If all required pieces are included, such as in Figure 3, their response to the milestone will be logged in the database.



Figure 2. Exemplary milestone question tweet

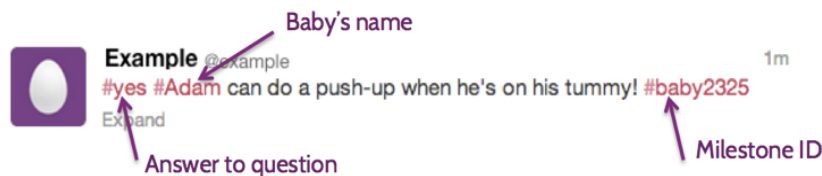


Figure 3. Exemplary response from users

Some milestones have figures that go along with them, such as showing a close up of a child doing a pincer grasp. I uploaded figures to *pic.twitter.com* and included a link in the prompt for relevant milestones. In addition to specific responses to the prompts, users are also able to create a free-form “memory” about their child by including the hashtag #babymemory in a tweet (e.g., #Billy visited the zoo today! #babymemory). When doing so, the remainder of the text in the tweet will be included as a memory on their child’s Baby Steps timeline on the companion website. Also, because there are only 30 milestones in a 2-3 months period, the system tweeted parenting tips and activities on a regular basis, such as advice about vaccinations and ideas for activities that encourage development. Tips and activities used the #tips and #activity hashtags respectively.

3.1.2 Twitter System Implementation

The Twitter system was developed in Python and consists of a series of scripts which run on the server that also houses the website. MySQLdb, a Python MySQL connector module, is used to interface with a centralized database shared across different pieces of the Baby Steps suite. Twitter interaction is achieved by making calls to Twitter's REST API v1.1. @BabySteps uses a collection of automated Twitter accounts, each of which have authorized full read/write access to the registered Python application. These accounts are each named with a given month and year combination, e.g. @BabyStepsAug11, @BabyStepsNov12. Generally speaking, the @BabySteps system can be broken into two general pieces: the "push" component and the "pull" component. Their basic operation is described as follows and overall system flow is mapped out in Figure 4.

Push Component

Three times per day, during normal daytime hours in our local time zone, this component of the system is executed via a cron job. It iterates through each of the authorized month-specific accounts sequentially. For each account, the system identifies the integer ID of the last milestone question that was tweeted out. This ID is then incremented, and the next milestone in the sequence is checked in the database. Each milestone has an age range for when it is appropriate to be asked. The age implied by the month and year in the Twitter account's handle is then compared against this range. If the new milestone is age-appropriate, it is sent out as a new tweet, with the hashtag #baby[ID] appended to the end, where [ID] is the milestone's integer ID. If all relevant milestones for the age bracket have been exhausted, the system sends out a random age-appropriate activity or tip.

Pull Component

This script executes every half-hour around the clock and also iterates through the accounts sequentially. It begins by identifying the unique Twitter-assigned ID of the last incoming tweet that was received. All public tweets as well as all direct messages from the users followed by and following this account created since this last check ID are then collected for processing. From the raw text of each individual tweet, the set of hashtags are extracted. To identify milestone responses, @BabySteps system looks for tweets whose hashtag sets contain the minimum set of required elements outlined previously. If all required

components are present, the user's response to the milestone prompt is recorded in the database. Alternately, if the tweet contains the hashtag #babymemory, the tag is stripped from the raw text, and the remainder is inserted into the user's log of baby memories. In either case, the entry is time stamped according to when the user tweet was originally authored.

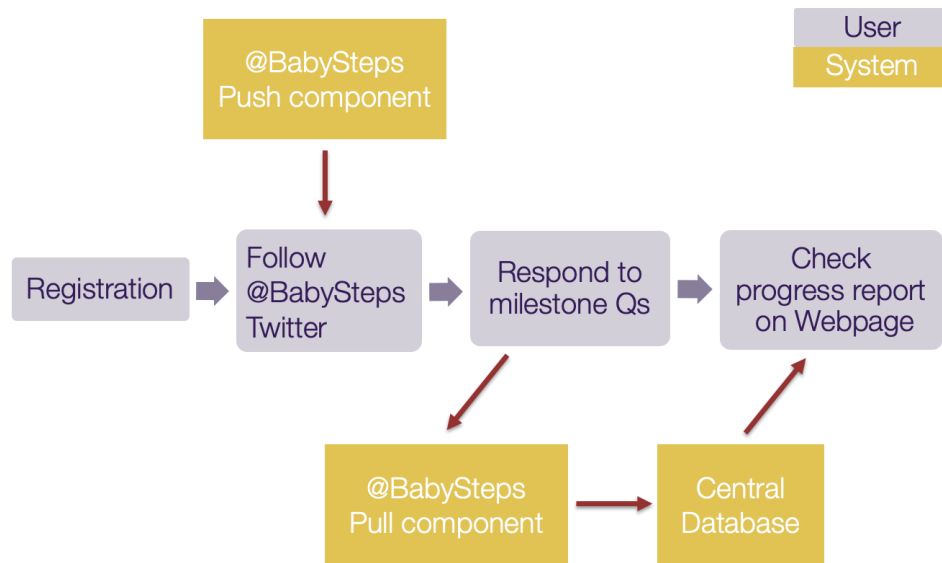


Figure 4. Simplified diagram of how @BabySteps works

3.1.3 Companion Website

In addition to the Twitter-based system, Baby Steps also had a companion website (Figure 5) where participants can log on to edit milestone responses, view a timeline of their child's progress and memories, and see an assessment of their child's progress based on their responses (e.g., whether they are on track developmentally or if they should get in touch with a doctor for additional assessment). The website also had a frequently asked questions page with information on how to use the system and a resources page that had contact information for public health resources to encourage parents to get in touch with their doctor if there was any concern about their child's development. The eventual goal of the Baby Steps work is to have our state's public health department conduct follow-up phone calls with parents based on their response to milestone prompts, but this system is not yet in place. Thus, the website served as an intermediary for parents to understand their child's progress.

3.2 DEPLOYMENT STUDY DESIGN

To test the feasibility of the @BabySteps system design, understand the usability of the hashtag syntax, and to gauge the acceptability of using Twitter for sharing a child’s milestone progress, I conducted a real-world deployment study with parents of young children. I conducted a qualitative, exploratory study, as Klasnja et al. have argued is the best first step for evaluating novel health systems [148] along with some quantitative data logging. This section includes the details of the study design and the participants.

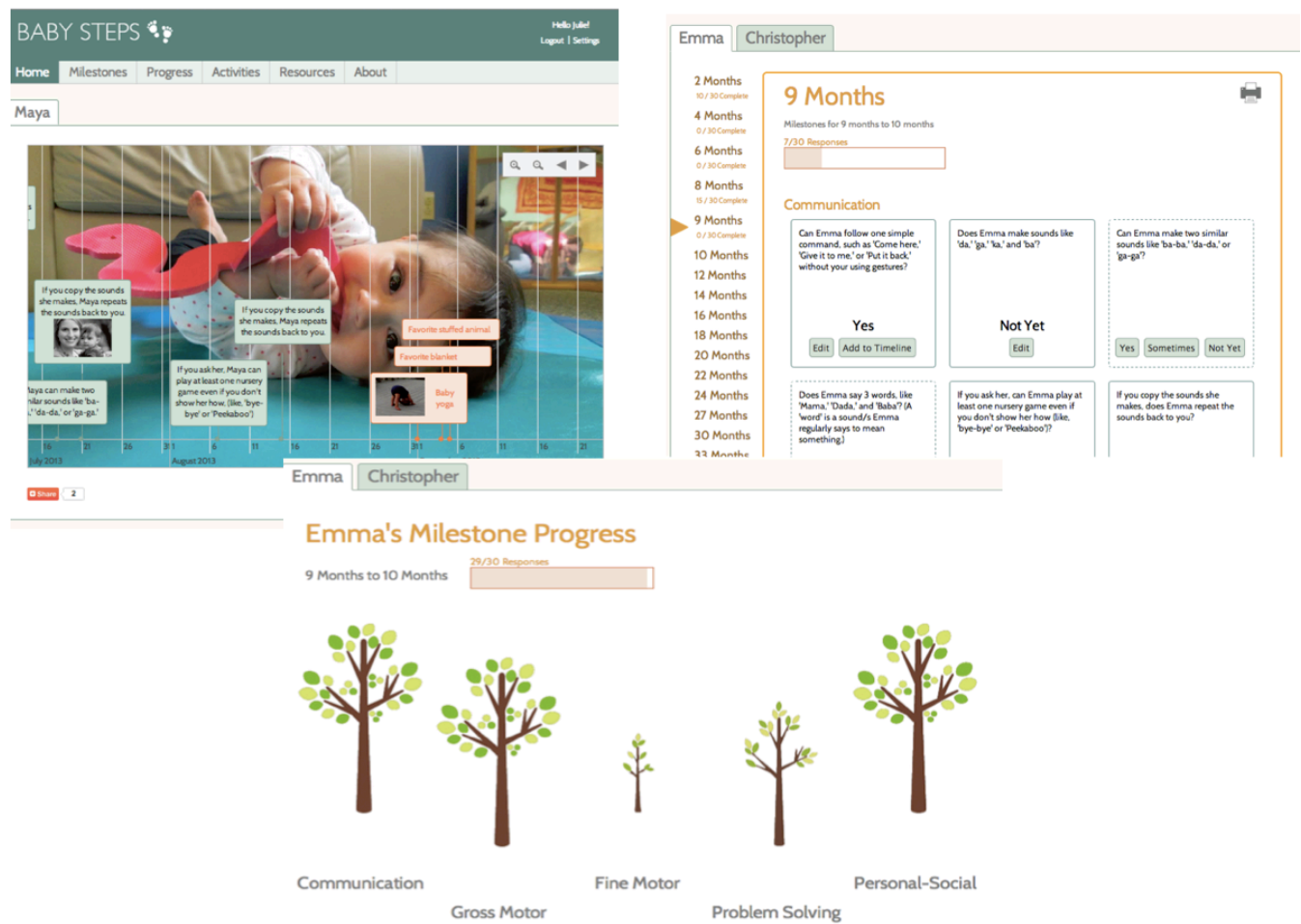


Figure 5. Baby Steps website design

3.2.1 Study Procedure

Before deploying @BabySteps, I sent out a pre-study survey asking participants about their demographics, family, Twitter usage, and parenting practices. With this information, I registered participants in @BabySteps system.

Stage 1: Two Weeks of @BabySteps Usage on Twitter

Because the @BabySteps system required use of special syntax in tweet responses, I sent out Twitter instructions to participants before Baby Steps monthly accounts started tweeting out. On the instruction page, there was information about how to use the three hashtags and information about how to follow the Baby Steps main account and monthly accounts. Also, because I acknowledge that privacy can be of concern to parents, I encouraged participants to check their account's privacy settings. The @BabySteps system also did not limit the form of response – it could be a

reply to tweets by @BabySteps, a new tweet, a re-tweet, or a direct message. That way, participants had the ultimate say in what they wanted to share publicly vs. privately. Then, for two weeks, each monthly account tweeted out developmental milestones, a set of functional skills or tasks that most children can do at a certain age range. Based on their observations, participants responded 'Yes', 'Sometimes', or 'Not Yet / No' accordingly. Occasionally, both the main account and monthly accounts sent out reminders for syntax and study-related announcements, as inspired by Tweak the Tweet [112].

Stage 2: One Week of @BabySteps plus Website Usage

After participants used @BabySteps and responded for two weeks, I made the Baby Steps companion website available for parents to view their child's progress based on milestones they answered via Twitter. The goal of this study was not to test the website itself, although participants were welcome to provide feedback on the system's design. For this study, the script only parsed the tweets with correct syntax. Due to this, not all of the participants' Twitter responses were successfully recorded on the website. However, because the progress report may not accurately show child developmental progress until all 30 milestones are answered for a certain age range, I notified participants about this fact and

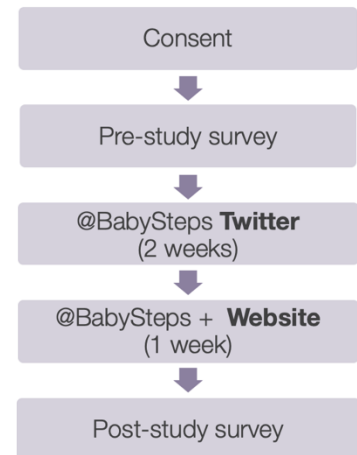


Figure 6. Study procedure

encouraged them to answer any milestones that they had missed via the Website before reviewing the progress report. To reduce any unnecessary parental stress, I explained in recruitment materials that child development is variable, with a wide range of what is typical as children grow in different areas at different rates. Participants had approximately one week to use the companion website before we deployed a post-study survey, though they were free to use the website indefinitely after the study completed. I did not open up the site initially since I wanted to study how parents used Twitter for responding to milestone prompts before they had a chance to answer questions via the website. The post-study survey included questions about the frequency of tweets, their ease of learning the hashtag system, any concerns about sharing child developmental information publicly, and open-ended questions about how to improve the design of @BabySteps. Figure 6 lays out overall study procedure.

3.2.2 Participants

As the main body of study occurred on Twitter, we recruited participants via Twitter. The research team tweeted a short study description and a link to a screener survey to their personal accounts and asked for people to re-tweet it. Some followers re-tweeted or forwarded it to a relevant person, with the majority of participants coming from a re-tweet by a person with nearly 20,000 followers (who also enrolled in the study as P5). The screener survey asked if they were parents of a baby between 2 months and 2 years, the number of children in their household, their experience with Twitter, and optional demographic information. I received 72 responses to the screener survey. Of those that responded to the screener, I sent invitations to approximately 40 participants to recruit a large, diverse set. Of the invitations sent, a total of 19 parents ended up consenting, completing the pre-study, and enrolling in the study. During the course of the study, one decided to withdraw because of concerns over sharing tweets about her child publicly on Twitter, and two people never sent tweets to @BabySteps and thus dropped out. Two participants replied via Twitter to milestone posts but did not respond to the post-study survey (P1 and P7). Overall, I was able to collect @BabySteps usage data from 14 participants and 12 responses from the post-study survey.

The participants were a mix of demographics from all over the United States, though most were younger as they had young children. Since we recruited via Twitter, most of participants had good

experience with it. Six participants had multiple Twitter accounts and nine participants had used Twitter for more than 3 years. All participants currently recorded information about their child in varying ways, and many often brought notes to doctors. However, lack of time and forgetting to record data were the most common difficulties they faced in keeping records regularly, which made them ideal candidates for @BabySteps. Two participants (P2 and P9) used @BabySteps with their two children, with P9 having fraternal twins. All participants consented via the web and received a \$20 Amazon.com gift card as a token of appreciation. See Table 3 for participant details.

Table 3. Participant details

ID	Age	Gender	# of children	Years Using Twitter	# of Twitter accounts	# Followed	% Personal acquaintance	# of Followers	% Personal acquaintance	Protected tweets	Twitter Frequency	Baby's Initial Age(s) in months	Baby's Gender(s)
P1	25 - 30	F	1	> 3 years	1	329	30-50%	226	< 10%	No	Several per day	21	F
P2	25 - 30	F	2	> 3 years	1	772	< 10%	329	< 10%	No	Several per day	4, 24	F, F
P3	25 - 30	M	1	> 3 years	3+	402	50-70%	602	30-50%	No	Several per day	12	M
P4	25 - 30	F	1	> 3 years	2	5	> 70%	5	> 70%	Yes	Once per 2 weeks	13	F
P5	25 - 30	F	1	> 3 years	1	8,169	< 10%	19,736	< 10%	No	Several per day	13	M
P6	31 - 40	F	2	1 - 3 years	2	175	< 10%	235	< 10%	No	Several per day	8	F
P7	31 - 40	F	1	> 3 years	2	577	< 10%	1598	< 10%	No	Several per day	22	M
P8	31 - 40	F	2	1 - 3 years	1	277	< 10%	214	< 10%	No	Several per day	16	F
P9	31 - 40	F	5	> 3 years	1	385	30-50%	269	30-50%	No	Several per day	5, 5	F, M
P10	31 - 40	F	1	1 - 3 years	1	150	< 10%	80	< 10%	No	Several per day	16	M
P11	31 - 40	F	2	1 - 3 years	1	305	10-30%	68	> 70%	Yes	Once per day	13	M
P12	31 - 40	F	1	> 3 years	1	795	< 10%	364	< 10%	Yes	Several per day	10	F
P13	31 - 40	F	2	> 3 years	3+	~900	10-30%	~4,000	10-30%	No	Several per day	22	M
P14	31 - 40	F	2	1 - 3 years	2	~100	10-30%	50	50-70%	No	Several per day	20	F

3.3 RESULTS

I here report findings from the @BabySteps deployment study on how participants used the @BabySteps system, learned pre-defined syntax, reacted to privacy issues, interacted with one another, used the website complementary to Twitter, and their overall reaction to the system.

3.3.1 Overall Tweeting Activity

For this study, Baby Steps set up twelve accounts including the main Baby Steps account. Over the three weeks, each account associated with a baby's birth month tweeted out age-appropriate milestone questions as well as useful parenting tips (#tip) and activities (#activity) that parents could try with their children. A total of 670 tweets were sent (5 from the main account) averaging approximately 60.5 tweets among eleven monthly accounts (average 2.88 tweets per day). The composition of those 60.5 tweets was 42.1 milestone questions, 8.1 tips, 8.4 activities, and 1.9 study-related announcements (e.g. reminders to use pre-defined hashtags, functionalities, etc.). Depending upon the age range, there could be slight difference in the number of milestone questions, tips, and activities. In response to @BabySteps' tweets, 14 participants made 319 milestone responses (22.3 per participant, see Table 4).

I also wanted to know what participants thought about the frequency and timing of tweets made by Baby Steps. In this study, an average of 2.88 tweets were sent out per day at random times, rather than tweeting at set intervals, to be more noticeable by participants checking Twitter at any time of the day. Eight participants thought this frequency was just right, while four thought it was too often. However, when it came to tweeting each tweet at different times, I received mixed opinions.

P6: "(...) whenever I checked twitter I usually would see at least one tweet from baby steps, which I really liked."

P12: [in response to how to improve system] "Sending out the tweets at designated times every day (or even repeating them) so I would know to expect them."

The @BabySteps system was designed to tweet parenting tips and activities after all milestone questions for an age group had been exhausted. One participant expressed that she would prefer if the system sends out them in parallel.

Table 4. Overall usage by the 14 participants in the study. The sparklines to the right are meant to give an overview of activity over 3 weeks. P5 started the study later than others

Participant ID	Total Tweets	Average Tweets per Day	Tweet Frequency by Time
P1	6	0.3	
P2	9	0.4	
P3	31	1.3	
P4	28	1.2	
P5	22	1.0	
P6	10	0.4	
P7	25	1.1	
P8	14	0.6	
P9	26	1.1	
P10	47	2.0	
P11	26	1.1	
P12	23	1.0	
P13	12	0.5	
P14	33	1.4	
Average	22.3	1	

P5: "The ideal frequency would be a balance of the milestones with the tips. Some days were just tips.

Others were just milestones. It seemed unbalanced, hard to predict, irregular."

From the pre-study survey, I found out that all participants had their own ways of recording child development and were currently keeping records of it. The most common difficulties they faced were lack of time and forgetting to record data. As Twitter has a "Favorite" feature, some participants used it to save some of @BabySteps' tweets for later.

P5: "Because I checked the Twitter when [my baby] was asleep or not with me, and it was a milestone activity I hadn't tried yet with him. I wanted to wait until I could try it, and sometimes favorited the milestone so I'd remember to."

It was more common to favorite #tip or #activities rather than milestone questions, but overall favoriting was low.

3.3.2 Using Twitter Syntax

For the @BabySteps system to correctly recognize and store user responses, the @BabySteps system required three hashtags (in any order) to be included in each milestone response - 1) answer to the milestone question, 2) baby's name or pseudonym, and 3) milestone ID. Hashtags are very common on Twitter, however, they are generally created by users organically. I anticipated it might be new for participants to be required to include specific hashtags for a certain purpose and thus they would need some training. I sent out instructions (via a website link) and occasional reminder tweets about the syntax. As a result, among 319 milestone responses, 198 (62%) were correctly formatted and 121 (38%) were missing some parts.

There were a few cases (8 tweets) where the response had different answers (e.g. maybe, rarely don't know, not sure, etc.) instead of the given answer options (yes/sometimes/ no/notyet) but the majority of errors came from missing the # sign before answers to the question or missing the milestone ID. For example:

- @BabyStepsAug12 #baby2464 yes. Can turn pages himself [missing # before 'yes']
- @BabyStepsDec12 #yes so cute! [missing milestone ID]

Because of reminder tweets, most of the participants learned the correct syntax over time, but some participants expressed difficulties following the rules.

P2: "It [Syntax] was good but a little complicated - would have been nice to not have to do a hashtag for all responses, as that doesn't seem like a typical way to use hashtags. As in, you would never normally do a hashtag in front of 'yes' or 'no'. I had to keep going back to the email with instructions to remember what format you wanted the responses in."

The most preferred method of responding to milestone questions was via the Twitter mobile phone or tablet application (54% of all responses), which made it more difficult to follow the syntax.

Approximately 28% of the posts were via the Twitter website, and another 11% were either via a special app like TweetDeck or TweetBot for iOS and an additional 7% were of unknown origin (due to being direct messages). Four participants reported they used only mobile phone applications, whereas there was only one participant who solely used websites.

P11: "It was difficult to remember to use the correct syntax, period, but probably a bit harder on the phone."

P9: "Using it from mobile, I had to remember the correct hashtag (ex. #baby1234) because it was not visible when I was typing. This would not be the case with Twitter on the web, but I most often use my phone to tweet."

There was one interesting case of a participant who was responding to milestone tweets for her two 5-month old twins. She responded to the milestone prompt for both children within the same tweet (@BabyStepsMar13 #yes #baby2325 #babyname1 #babyname2). The @BabySteps system could not handle this syntax, so it was not parsed correctly. She eventually split up the tweets into separate ones for each.

3.3.3 Tweeting in Public

Twitter, unless the user explicitly chooses to go private, is placed in the public domain and whatever the user tweets will be broadcast to the world at large.

Privacy

As the main body of study occurred on Twitter, privacy issues naturally came up to the surface. Some participants expressed that they were completely fine with sharing child developmental progress on Twitter.

P13: "I don't mind publicly responding. I already do a lot of public responses, and tweets about my children. Response did not affect my preference."

P12: "My account is private, but I don't mind responding publicly. My child is meeting milestones as expected. Even if she weren't I wouldn't mind responding publicly."

However, not everyone agreed on that. In fact, one of the participants withdrew from the study because she did not feel comfortable with sending tweets with her baby's developmental information, even though her timeline was protected. She expressed that although her account was protected and private, she was concerned about tweeting negative answers to milestone questions. Also, P5, who responded via Direct Message (DM), expressed a similar privacy concern before consenting. After the study, she added her reasons for responding via DM rather than public tweets and the frustration she had gone through because of her choice.

P5: "I prefer to respond privately, because a public message is a) incomprehensible to anyone viewing my public timeline and b) runs the risk of giving peering strangers more information about my baby than I care to give."

P10: "I didn't use DM b/c I thought it was too complicated on my phone. (...) I responded publicly b/c it was easier on my phone. But I didn't put a lot of identifying info. Just yes, no, etc."

Too Much Baby Talk?

In addition to and/or aside from privacy and security concerns, some parents expressed their awareness of posting too many baby stories on Twitter, which was a concern raised in Morris's study of mothers and social networking use [135]. There was also a fear of overshare, which might lead to unfollowing [58]. Therefore, DM option might work better for parents with these concerns.

P4: "I feel like most people don't want to see me tweeting about my child's development -- I try not to inundate people with pictures on fb (just Instagram), so I would feel the same on twitter."

P6: "I did not know DM was an option, but I would have preferred it. I did not want to tweet publicly multiple times in a row about my baby's development. It seemed a little spammy."

3.3.4 Social Interactions

Because the study was relatively short-term, and the number of participants was small, opportunities for social interaction were limited. However, as the nature of Twitter is a social networking service, I was also interested in finding out whether participants ever interacted with each other. Among the eleven monthly accounts created for this study's purpose, three had multiple participants—two with two

participants and one with three participants. I received several testimonies about interaction, from one following the other, to a brief conversation between the two.

P5: "I saw who else was following the @babystepsjul12 and followed those who seemed interesting."

P10: "I conversed with a woman on twitter who was also following the same BabySteps birth month account. We talked about enjoying the study."

In addition, P14 showed her desire to interact with parents with a baby born in the same month. Future studies will include more participants to study this social dynamic at a larger scale.

P14: "I think I might have been the only person in my group, so it wasn't as interactive as I would have hoped."

3.3.5 Twitter and Website Integration

After responding to milestone questions via Twitter for two weeks, participants were introduced to the Baby Steps website. They could view their child's developmental progress based on their responses via Twitter and were able to answer additional milestone questions, add photos, share a timeline of their child's progress, get useful resource information, etc. Although testing the Baby Steps website was not the main goal of this study, because users had a week to explore, we received quite a large amount of feedback about how well it served as a companion of @BabySteps and how well they complemented each other.

P6: "I love it! So well structured. And I love that I can use it to put in milestones that I missed on twitter."

Despite it not being a popular feature on @BabySteps, keeping sentimental records of a child with #babysteps and checking it on the website received positive feedback.

P12: "I liked that I could record #babymemory bits to check back on later."

However, because the system parsed and recorded only the tweets with correct syntax for this study, not all of participant responses were recorded on the website. And that obviously caused frustration for some participants.

P4: *"It didn't actually pull in the responses that I had posted to twitter -- it only recorded three of the responses, and I had responded to all the questions"*

P11: *"None of my responses are there because I messed up the syntax."*

3.3.6 Overall Feedback

In the post-study survey, I asked participants about their general perceptions and suggestions for improvements or new features. Most participants had a positive reaction to @BabySteps and the website. Nine out twelve participants completed the survey and said that they wanted to continue to use both @BabySteps and the website after the end of the study. They also said that they would like to recommend @BabySteps to friends. For the most part, many appreciated that the system identified and informed about tasks and developments that baby should be doing.

P4: *"It was really nice to learn about how her development is potentially matching up with where she should be."*

P8: *"It encouraged me to try some things with my daughter (like letting her feed herself with a spoon) which I hadn't thought to do yet."*

Also, some participants appreciated how @BabySteps made them more aware of their child's development and grateful for their child meeting development progress on time.

P6: *"I loved that it reminded me about the less obvious milestones she was achieving."*

P3: *"I think some of the development that happened we just took for granted... so getting to keep track of it showed that our kid was developing at a pace."*

Ratings in a 5-point Likert scale (5 being the positive answer) revealed that participants thought that @BabySteps can be helpful to both first-time parents ($M = 4.25$, $SD = 0.87$) and parents with multiple children ($M = 4.17$, $SD = 1.03$).

P9: *"I can see how the tips and activities could be useful, especially for a 1st time mom."*

P13: *"Used it for my second child and enjoyed re-learning about various milestones I had experience with my older child >3 years ago."*

It turned out that there was a pediatrician among the participants. She revealed this in the post-study survey. Her motivation for participating was because she recalled that remembering the developmental milestones was difficult during her training. She gave great comments about the @BabySteps system.

P12: "Full disclosure - I'm a pediatrician... You have created a fantastic tool for tracking and for guidance for parents on how their child is doing developmentally. I would love to be able to use this in my own practice. (...) This is an objective way to track milestones. I love it and I hope I get to continue using it! I've been using a blog that is open to only family to track milestones and post pictures - this is even better for milestone tracking. Thank you so very much!"

3.4 DISCUSSION

Overall, I believe the deployment study showed that the @BabySteps system was successful in achieving its initial goals. In the end, all participants were able to learn the syntax and successfully respond to milestone messages. For most participants, the system kept them engaged over the course of the study, and I received positive feedback on the overall use and idea behind the system.

Despite the initial success of the @BabySteps system design in its first trial of it, I believe there are a number of actions needed to improve the usability and overall acceptability of @BabySteps. For example, the original design had forced each participant to include 2-3 three different hashtags within their responses to milestone prompts. This included the response (#yes, #sometimes, or #no/#notyet), the baby milestone ID (e.g., #baby2343), and if there was more than one child registered in the system, the child's name or pseudonym (e.g., #sally or #daughter). However, because participants sometimes forgot the hashtags in the response, it would be easy to modify the system to allow for answers to not require the hashtags. In addition, I could allow participants to omit the baby milestone hashtag if they reply directly to the message, since the Twitter API allows the system to identify which tweets are in reply to a direct message. This would alleviate some of the difficulty of having to copy the hashtag, especially when using a mobile device. Finally, the @BabySteps system could simplify the baby code slightly. The system used 4-digit milestone IDs initially, but because there are only about 660 unique milestones across the entire 5 years of a child's life, the system could use numbers between #baby1 and

#baby660 to shorten what the participant has to remember. These changes may prevent more user error and allow for flexibility, though it may still be necessary to be strict in the event of multiple children (like the case with P9, who tried to combine a tweet for her two twins).

In addition to simplifying the hashtag system, I believe there are possible future improvements to the system itself that could help participants in the beginning to learn the syntax and to also account for human error, typos, etc. The Tweak the Tweet application [112] had similar issues with having participants tweet using the right hashtags as well, which is why the @BabySteps system had reminder tweets about the syntax. I believe adding to current script the capability of sending automated replies in the event that the user sends an improperly formatted tweet would help with correcting errors sooner. These should be via direct message (DM). For example, if the user tweeted “@BabyStepsNov12 #yes, my daughter can kick a ball,” the @BabySteps system could send a DM back that says “Thanks! However, we cannot process your request due to: missing child name hashtag (#sally), missing milestone id hashtag (#baby234).”

In addition, although I did not encounter this in the short duration of the study, it is possible that a person could learn about @BabySteps via their followers and wanted to start using the system, which I hope will happen to help broaden the reach of @BabySteps system’s design. However, if they are not registered in the system, their replies cannot be processed. Thus, the system could send an automated message if anyone outside of the system sends a message inviting them to register on the Baby Steps companion website.

Using Twitter for a public health application is also not without its challenges, which would be useful for others interested in working in this space. First, the 140-character limit and the need to allow for additional characters for replies was somewhat challenging. The milestone content was inspired by the Ages and Stages Questionnaire [36]. However, due to the 140-character limit, I could not use the exact wording of the milestone questions, which in some cases were as long as 300-400 characters. Because the ASQ is a validated tool, this likely would break validation and may not be as reliable as the original. There is a similar issue with other public health campaigns that have used SMS messaging and have run

into 160-character limits, and thus there are public health organizations that are working on shortened, validated messages for this format. This could be one solution to this issue but may take some time and cooperation from the original sources of the content. Second, because Twitter remains in the public domain, privacy can be a critical issue. I cannot expect all users to have a strong level of media literacy. Therefore, it is important for me as a researcher and designer, not only to inform users about the risk in advance, but also to give them options that they can choose to control privacy. My future work is exploring alternatives to Twitter to track milestones, so for those who are not comfortable with using this method, so that they have several other options.

There is also a question on how frequent postings should be to avoid overload, missing milestones, and keeping the parent engaged without forgetting the syntax. There is currently not much guidance on how often to send messages, and I believe it is likely dependent on the health domain (e.g., it is possible parents may be willing to receive more milestone messages than they would reminders about quitting smoking). Although most of the participants in the study believed that the three milestones per day frequency was just right, there were a few who believed it was too many, or they would like to see milestones balanced more with the tips and activity suggestions. I believe additional studies are necessary to determine this correct balance.

Finally, one last issue is the need for having a companion website. It would be ideal to have the system reside entirely on Twitter, but for developmental milestone tracking, it is not enough to just collect responses from parents. I also needed to prompt them to communicate that information with their doctor and follow up as necessary. Although the eventual use case for Baby Steps is to have our state's public health system monitor parent responses in the database and directly follow up with participants based on the results, this system is not yet in place. Thus, the @BabySteps system had to have a website where participants could log in, see all of their responses, print out the results to bring with them to the doctor, and find local resources on who to contact if they were concerned about their child's development. There may be other health applications where this may not be needed, since the tweet itself could direct them to the appropriate resources.

3.5 CHAPTER 3 SUMMARY

I presented the design and implementation details of @BabySteps, a system for using Twitter to create an interactive public health system for recording children's developmental milestone progress. I conducted a 3-week deployment study to understand the feasibility of such a system. The @BabySteps system was successful in teaching participants to use the proper syntax to reply to tweets, it engaged participants over the course of the study, and received a number of positive comments about the system's design.

This work provides empirical data to evaluate T1 by examining the questions of how everyday technologies can be used to support child development monitoring. This research has two primary contributions: 1) the design and development of a novel system for using Twitter as an interactive public health platform and 2) the results of the deployment study demonstrating its feasibility. The findings from this deployment study have implications for designing such systems and can inform others of lessons learned in using Twitter for this purpose.

This study also exposed several opportunities for future work. This system and study have just scratched the surface for what is possible with using Twitter as an interactive public health platform that can shift the flow of information from many to a single source. I documented design opportunities revealed in this study and built on these findings directly by applying them in the design of next iterations of the Baby Steps ecosystem. In the following chapters, I explore other platforms to understand how parents prefer to respond to and share milestone information when compared to just a website or other mediums, such as text messaging. I also explore whether a technological approach to child development monitoring can maintain engagement over the long-term and improve self-efficacy and patient activation.

Chapter 4. Baby Steps Text: How Text Messaging Can Be Used for Child Development Tracking

One of the primary goals for the overall Baby Steps project was to identify ways to reach underserved families for developmental screening. Although previous studies have shown that mobile technologies may be an effective method in overcoming the healthcare gap because it provides new opportunities to reach underserved populations [22, 134], it was not well investigated in the domain of child development. Thus, to explore **T1** and identify design opportunities for mHealth (mobile health) in child development monitoring, I investigated the feasibility and acceptability of text messaging in child development tracking. To investigate, I first conducted a user-centered design approach that formed the basis for the final design of Baby Steps Text. After designing and developing the Baby Steps Text system,

I conducted a 4-week deployment study with 14 parent participants, who were recruited from a local pediatric clinic that serves a high number of underserved, low-income, and immigrant families. This deployment study provided me with empirical data that is relevant to the broader questions of **RQ1** (what are the feasible ways technology can reach families for regular child development tracking?) and part of **RQ2** (how does technology impose burdens on its users and how can we measure them?) and **RQ3** (how does technology provide benefits to its users and how can we measure them?) By involving families with diverse socioeconomic statuses, I was able to understand how we might include resource-constrained populations in everyday health monitoring when they may not yet have constant Internet access or sophisticated smartphones. This work became a foundation of longitudinal study I present in Chapter 7. I worked with 4 collaborators on this investigation.

4.1 PRELIMINARY WORK & SYSTEM DESIGN

4.1.1 Design Methods

To explore the use of technology in helping to reach underserved populations in Washington State in collaboration with the state's Department of Health, I conducted a user-centered design process (see Figure 7) and engaged with key members. In this preliminary user research, I focused specifically on Latino populations in Washington State, as they represent a large percentage of underserved populations and experience a number of challenges that may affect other underserved populations, such as access to health insurance, access to mobile technology, cultural and language barriers [54], and kindergarten readiness. However, the intended users of Baby Steps Text are broader than just this specific population. This study was reviewed and exempted from IRB review by the University of Washington's human subjects review board.

User Research

In the initial user research, I engaged with members from the community who work with Latino families, such as healthcare professionals and social workers, those who work at non-profits which provide services to Latino families, and experts on the culture. I also worked with families themselves to understand their needs and their access to technology. I conducted semi-structured interviews with five

stakeholders, including one parent of a child with a developmental disability, two health care providers, and two Latino community health advocates. The interviews aimed to understand how parents engage in healthcare for their children and participate in developmental screenings, how parents learn about health topics, and how health information is communicated. The research team members also visited a Latina health fair located in a predominantly working-class Latino neighborhood in Seattle. The fair was attended by Latina mothers, their families, and community leaders. I administered short surveys to 15 mothers and asked them questions related to technology ownership, technology use, the number of children in their family, and their language use.

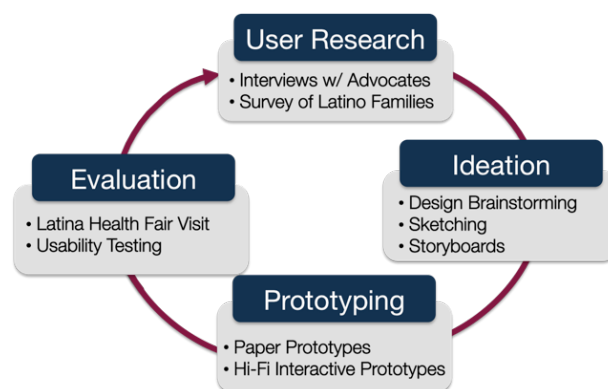


Figure 7. User-centered design process contributed to the design of Baby Steps Text

From the interviews, I was able to learn that Latino households are commonly multi-generational – consisting of children, their parents, and their grandparents – and the responsibility of caring for children is shared with members of the extended community. Designing an appropriate solution would also need to take into consideration the different technologies that community members have available to them. When asked to describe what sorts of technologies that members of the Latino community would most likely have access to, the most common answer was a mobile phone with basic capabilities. They suggested that home computers or home Internet access would not be very common, but that members might use computers in public access venues, such as at a public library. Findings from the survey supported several findings from the interviews. In particular, mobile phone use was prevalent

(93% of respondents), as was text messaging (87%). Only 6 participants (40%) used their phone to access the Internet and only two (13%) downloaded applications, which meant an entirely text message-based system would be more accessible than a mobile website or app. Home computer and home internet access were also not as prevalent (60% and 53% respectively). Another finding was how technology was used differently by mothers and fathers within a household. Two mothers (13%) mentioned that they did not use services like mobile Internet or Facebook, but that their spouses did.

I also learned from the providers and health advocates that they serve a community that varies widely in literacy skills. Members of their community might range from highly-educated households that can speak both Spanish and English at a high level to blue collar families where Spanish is primarily spoken and there are moderate literacy rates to undocumented immigrant households in which Spanish is spoken, but literacy is low. The Latina fair survey confirmed these numbers, with the majority of participants preferring Spanish, but nearly half also having English proficiency.

To summarize the preliminary user research, the research developed a set of 9 personas representing stakeholders of the potential system, including three different family caregiver types; community members, healthcare providers, and advocates in the state (see Figure 8 for an example persona). These personas were used during the ideation and design process to ensure the needs of stakeholders were represented in the design of the tools.

SARITA THE SPECIAL NEEDS PARENT

"I wish I had known sooner. It want to do everything I can to see where my child stands and help him grow."



PERSONAL INFORMATION

Age: 31 **Income:** \$25k
Profession: House cleaning
Children: Elias (5) and Ramon (2)
Technical Proficiency: Computer use at library, mobile phone at home
Country of Origin: Mexico
Languages Spoken: Spanish, some English

HEALTHCARE INFORMATION

Health Insurance: Yes, but limited coverage
Doctor Visits: About once monthly for herself or her children. Visits local health clinic.
Healthcare Background: No professional medical training

GOAL: To understand where Elias stands developmentally and connect him with treatment services, and to watch for warning signs in Ramon.

Sarita is a working mother of 2 children. Her older son, Elias, had difficulty speaking when he was nearly three. She did not notice as much at first, but it became apparent when her sister's family moved in with them, and Elias's similarly aged cousin could speak much more than him. She mentioned this to the local *hushero*, or community healer. She suggested that Sarita have her child evaluated by a local early intervention specialist. Elias was diagnosed with speech delay and referred to a speech-language pathologist, who has been seeing Elias for the last two years, and he now speaks at a developmentally appropriate level for his age.

Because early intervention worked so well for her son, Sarita is vigilant about looking for the early warning signs for her younger son Ramon. She also continues to track Elias's progress to make sure his therapies are keeping him on target.

SUPPLEMENTAL PERSONA: Parents of children diagnosed with special needs are a smaller portion of target users.

Figure 8. Example persona for a parent of a child with special needs.

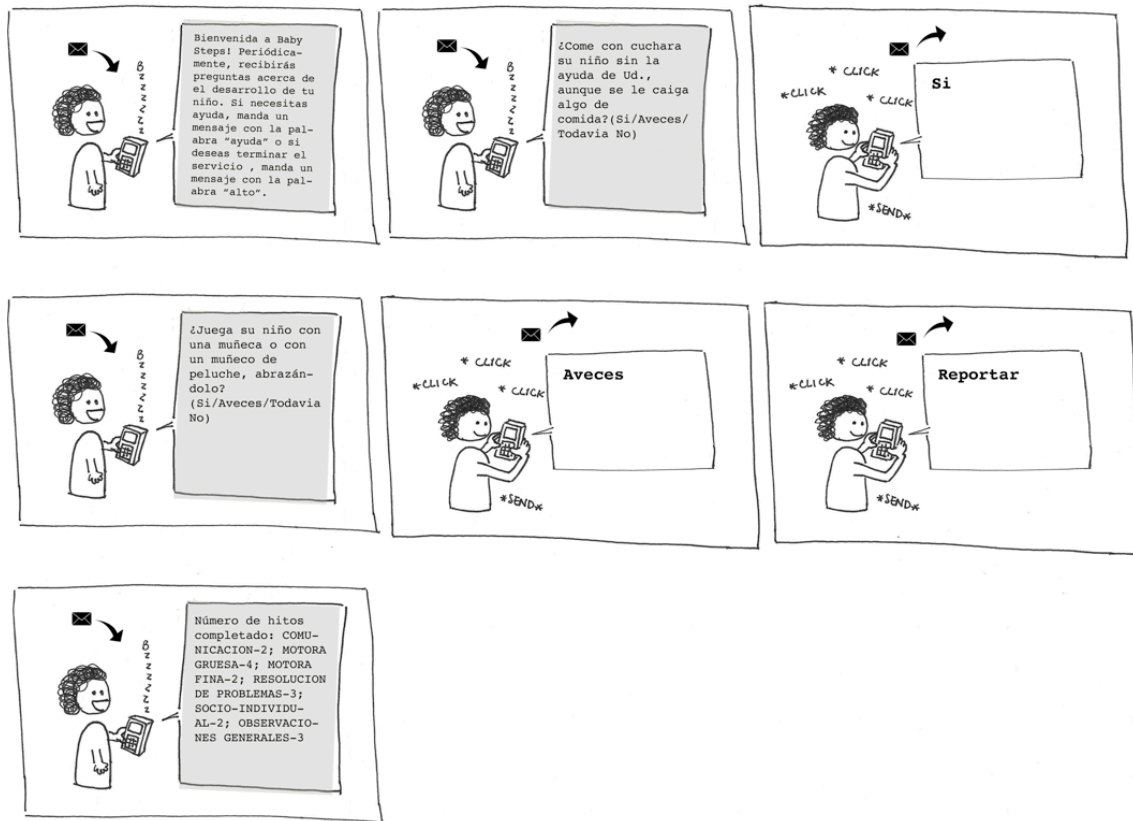


Figure 9. Storyboard of Baby Steps Text used at the Latina health fair (in Spanish)

Ideation & Prototyping

After the initial user research phase, the research team conducted a design ideation brainstorming session and generated 63 unique ideas for potential technologies to reach Latino families in the state with developmental screening. We kept copies of the personas present during the brainstorming to ensure designs were grounded in the needs of the stakeholders. From there, we narrowed down design ideas into the most promising and feasible three ideas for which we developed storyboards in both English and Spanish suitable for user feedback (Figure 9). The three ideas included text messaging, a website, and a public kiosk. We also developed high-fidelity interactive prototypes of a public kiosk and a website using the Axure prototyping tool (<http://axure.com>) and a Wizard-of-Oz prototype [85] of the text messaging tool using Google Voice to send messages to participants' phones and a detailed script of messages.

Evaluation

We reviewed storyboards and prototypes with stakeholders, including advocates and families attending the Latina health fair. When reviewing the storyboards with potential users, the text messaging-based system seemed to be the most promising approach with the furthest reach. Stakeholders and parents all responded positively to a text-messaging based system, as long as the costs were kept low. Participants were unsure if they would use or have consistent access to a public kiosk, and though a website would be helpful and provide a richer experience, the lack of a home computer or data plans on their phones limited potential access.

Task 3: Responding to Milestone Questions Participant task: You will soon receive a message with a question about your child. When you receive the message, respond to it based on your child's experience using the instructions. Wizard response: <i>Send message: "Can NAME drink water, juice, or formula from a cup while you hold it? (Reply Y for "yes," S for "sometimes," N for "not yet")"</i> <i>If response is anything but Y, S, or N, reply back with:</i> ○ Reply to ?'s with "Y" for yes, "S" for sometimes, or "N" for not yet. New msg sent after you reply. Txt "stop" to cancel. Txt "report" for NAME's progress.	Task 11: Requesting a Report Participant task: Request a progress report on how many milestones your child has completed so far. Wizard response: <i>Expect to receive the word "report" from parent. Send back:</i> <i>NAME's Progress, Number of milestones completed:</i> <i>Personal-Social – 34</i> <i>Gross Motor – 23</i> <i>Fine Motor – 22</i> <i>Communication – 40</i> <i>Problem Solving – 22</i>
---	---

Figure 10. Example script for "Wizard" in Wizard-of-Oz prototype for two of the twelve tasks. Wizard instructions are in italics and BOLD CAPS indicate parts customized by the Wizard for the user's child.

After deciding to move forward with the text messaging design, we iterated on the design and wording of the Baby Steps Text system through initial Wizard-of-Oz usability studies with 10 parent participants. In these studies, we had two researchers simulate the experience of using Baby Steps Text for the participant. Participants used their own mobile phones to send and receive text messages from the researcher playing the part of the “wizard,” who used the Google Voice interface on a laptop, along with a script (see Figure 10), to quickly customize responses for the users during testing. We had a total of 12 tasks that covered all aspects of the system, including registering, responding to several milestone questions, receiving a report, requesting help, and stopping the service. Participants had a printed sheet for each task, and a study coordinator guided them through the process while the “wizard” read and responded to the system.

Participants offered suggestions on improving the content and process for receiving and responding to text messages. In particular, our original design for the progress report only included the number of questions answered and no real information about how the child was progressing in each category. I thus added more information about the child’s progress and added more information about what to do if the child’s result in any given category of development was “More evaluation needed.” The registration component was also confusing, as it required the child’s name, birthdate, and gender to be separated by spaces. I also found that participants could easily remember the “y” “s” and “n” responses after a few milestone prompts and thus could get by with only simple instructions rather than the full instructions with each message. I iterated on the design several times to ensure the system was understandable and made improvements on the wording and language used before moving to the implementation and feasibility testing stage, which I describe in the following section.

4.2 SYSTEM DESIGN & IMPLEMENTATION

In this section, I describe design details for Baby Steps Text, which were based on the technical constraints of SMS and MMS and findings from the preliminary user research.

4.2.1 System Design

Personalized Milestone Questions

One of the primary advantages of using mobile technology is that it is personal. Because each message

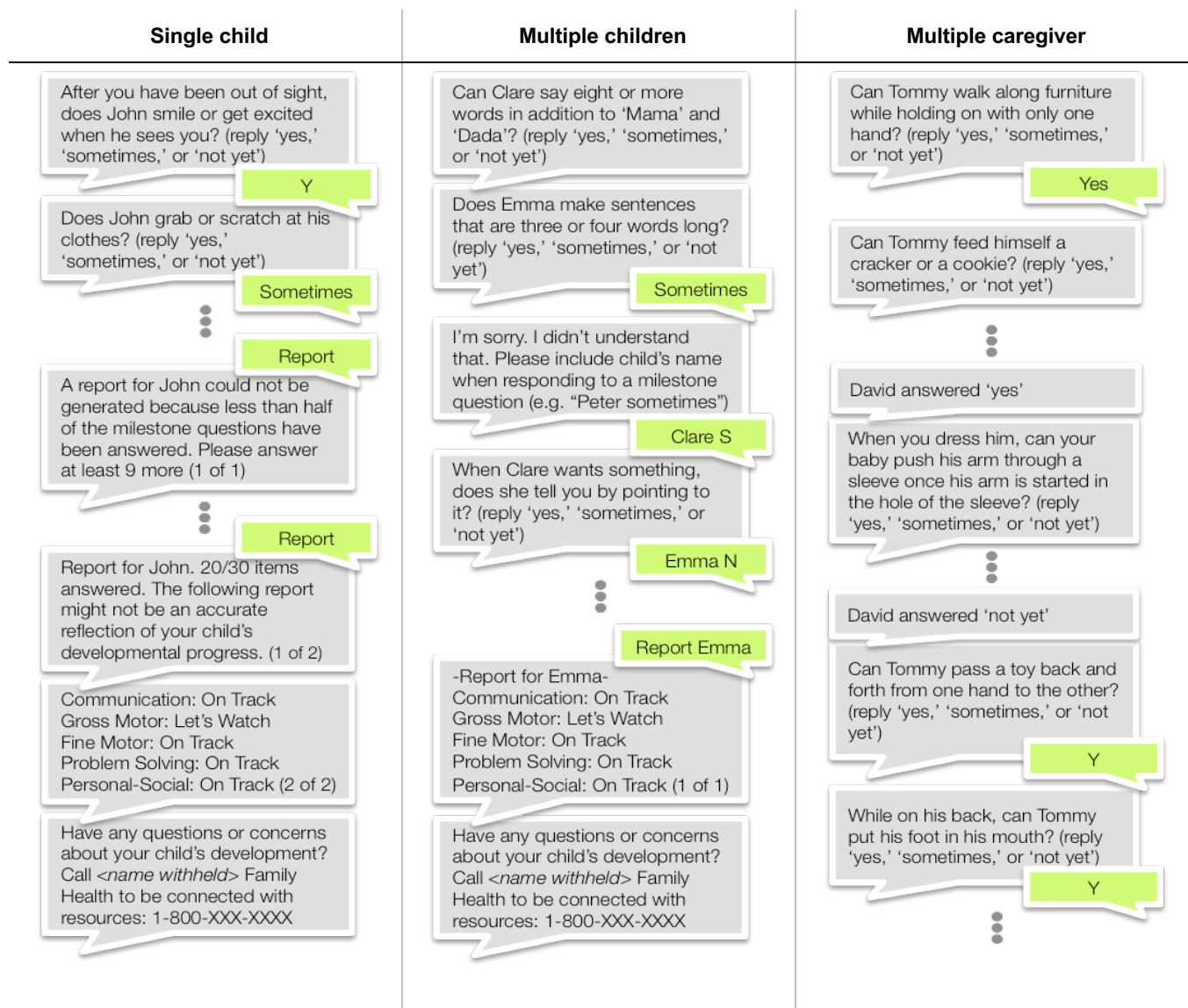


Figure 11. Example of Baby Steps Text syntax usage. Gray (left) indicates messages sent by Baby Steps and green (right) indicates messages sent by caregiver

is sent to an individual, the system can personalize the content of the messages. Using information from enrollment such as child birthdate, name, and gender, Baby Steps Text can send age appropriate messages and can include the child's name and gender-specific pronouns. To automate this process, placeholders were used for words that needed tailoring. Overall, Baby Steps Text consists of 660 distinct milestone questions that are tailored for each child and delivered based on age.

Simplified Response Structure

To make Baby Steps Text system fully automated, I made simplified structures for responses (see Figure 11). There are four response options to milestone questions: 'Yes,' 'Sometimes,' 'Not Yet,' and 'Revisit' to skip to next question and respond to it later (all case-insensitive). To reduce typing, parents could also use first letter of each response type (e.g., Y for 'Yes' and R for 'Revisit'). If the user had multiple children registered with Baby Steps Text, they needed to also include their child's name along with the question response. Beyond these required elements, users could add free text in their response (e.g., "Yes, Tommy can drink milk from a cup!") like natural text conversation, which does not interfere with the system's automated parsing mechanism. I wanted to repeat instructions for using "Y" "S" and "N" with each milestone prompt to remind participants, but some questions were too long to include the full instructions within the 160-character limit. Thus, for shorter milestone questions, the milestone question text included the full instructions ("reply 'y' for "Yes", 's' for "Sometimes", or "n" for "Not Yet") and simplified instructions (reply 'y', 's', or 'n') for longer questions. While modifying the screener content to shorten it, I ensured that the longest message was a maximum of 130 characters to ensure at least 30 characters were remaining for instructions.

Progress Report

If parents text "Report" to the Baby Steps Text system, it generates and sends out a child's progress report based on parents' responses to individual questions (e.g., whether they are developing on schedule or if they should get in touch with a provider for additional assessment). To fit the progress report into a single text message, I used three simple reporting levels ("On track," "Let's watch," and "More evaluation needed") for each development category (e.g., Communication, Fine Motor, Gross Motor, Problem Solving, Personal & Social). I aimed to carefully convey progress to reduce unnecessary parental anxiety by working on the exact wording with experts a local non-profit organization that provides a free statewide program for family health that serves many low-income families. If a parent requests a progress report before they answer half of the questions, the system responds to inform they need to answer more questions first. If parents answer more than half but not all 30 questions, the system provides a progress report but also sends an additional message that the progress report they receive

may not be an accurate reflection of their child's development until all questions are completed. If a child is not 'On Track' in any development categories based on parents' responses, Baby Steps Text offers a toll-free hotline number for a local non-profit organization in a message following the progress report so that parents have contact points for questions or concerns about child development.

Intervals of Screening Questions

Because initial user research indicated that mobile phone plans used by lower income families may vary and be limited (e.g., pay-per-use plans), I had to find ways to deliver complete screening questionnaires within a certain time frame so that those screening questions were still valid to the child's age. I originally intended to send one screening question per day to avoid information overload, but due to this finding, I designed Baby Steps Text to send the next question immediately after they respond to one because it signals that 1) they are active on mobile texting, and 2) they are able to send and receive messages at the current moment. If the parent does not respond to a question within 48 hours, the system resends the same question again.

Supporting Multiple Caregiver Engagement

Child development happens in daily life, and it is not always one parent who observes when their child hits certain milestones. During user research, I found that often family members shared child-raising responsibility with aunts and uncles and grandparents, and not every member of the family had consistent access to mobile phones. To coordinate monitoring and tracking child development across multiple caregivers, Baby Steps Text can register multiple users' phones for one child. With that, each question is sent to all caregivers' mobile phones, and if one answers a question, the others receive a message saying "[name of other caregiver] answered [response type]" and all will receive the next milestone question to answer. This allows all caregivers to follow their child's development and coordinate responses.

4.2.2 System Implementation

Baby Steps Text is a custom-built Python application and associated MySQL database, which uses the Twilio cloud communication platform (<https://twilio.com>) for SMS and MMS handling. Incoming

messages sent by users to a Baby Steps Text phone number are received by Twilio, which then relays the data to the system via HTTP request to be parsed for keywords. Similarly, by sending HTTP requests to Twilio, Baby Steps Text can send both SMS and MMS messages to users programmatically. When the system receives an incoming message, the content is parsed to determine if it is a milestone response or another set of recognized keywords (REPORT, START, STOP, HELP); responses are stored in the relevant child's database entry and the system sends an appropriate error message in response to any message that does not fit into the response structure. Outgoing messages are sent in two distinct scenarios. When a user answers a milestone question, the next question is sent immediately. They can also be sent via a daily check (handled by a cron job on the server), which determines on a child-by-child basis if it is time to either send the first question of a new milestone set (if the child has entered a new age range) or resend the last outgoing message (if 48 hours have elapsed without a response).

4.3 METHODS

In winter 2015, I conducted a month-long deployment study with 14 participants to test the feasibility and usability of Baby Steps Text. Through this, I was able to collect usage logs and qualitative data from pre- and post-study surveys. The study was reviewed and approved as minimal risk research by University of Washington's human subjects board.

4.3.1 Participants

Although I did not exclude participants based on income, I aimed to recruit resource-constrained families in the study through the waiting room of a local pediatric clinic that serves a high number of underserved, low-income, and immigrant families. Because the current prototype only has English-based questions, I required participants to have basic English literacy. The clinic receptionist introduced eligible parents of children under five years old to the study at check-in and completed sign up for the study and completed the pre-study survey in the waiting room.

I recruited 14 participants from a variety of backgrounds. The participants were caregivers (M: 2, F:12) of at least one child under 5 (Table 5). Despite my aim to recruit a balance of genders, in all but three cases (two fathers and one aunt), it was the mother who checked in at the clinic. The majority were

aged between 33 and 40 years old (n=8), and the total children in the household varied from zero (aunt) to seven. All owned smartphones and were active text users. Half (n=7) were single and half were married. Child participant ages ranged from 4 months to 60 months. One participant (P11) used Baby Steps Text with two children.

Table 5. Participant details

ID	Relation to child	Age	Race*	# of child	Marital Status	Education level	Household Income	# of texts per day	Child age (in months.)	Child sex	Post-study survey	Interview
P1	Father	33-40	A	1	Married	Some college	\$35,000-\$49,999	1-10	20	F	o	
P2	Mother	33-40	H	7	Single	8 th grade or less	< \$15,000	1-10	30	M		
P3	Mother	33-40	W	3	Married	College degree	> \$100,000	1-10	36	F	o	
P4	Father	33-40	B	3	Married	Graduate degree	\$50,000-\$74,999	1-10	20	F		o
P5	Mother	25-32	B	-	Single	High school or GED	Under \$15,000	1-10	54	M	o	
P6	Mother	18-24	B	1	Single	High school or GED	\$15,000-\$24,999	101-500	27	F		o
P7	Mother	33-40	B	6	Married	High school or GED	< \$15,000	1-10	14	M	o	o
P8	Aunt	33-40	A	0	Single	College degree	\$50,000-\$74,999	11-30	20	F		
P9	Mother	33-40	W	1	Married	Graduate degree	\$75,000-\$99,999	1-10	4	F	o	o
P10	Mother	25-32	H	2	Married	High school or GED	\$25,000-\$34,999	101-500	60	F		
P11	Mother	25-32	H	3	Married	High school or GED	\$25,000-\$34,999	1-10	30, 4	F, F	o	
P12	Mother	18-24	B	1	Single	High school or GED	< \$15,000	1-10	27	M		
P13	Mother	41-50	B	2	Single	High school or GED	< \$15,000	1-10	48	M		o
P14	Mother	33-40	-	3	Single	College degree	Prefer not to say	1-10	4	F		

(*A: Asian, B: Black or African American, H: Hispanic or Latino, W: White)

4.3.2 Study Procedure

Once I explained the study details to parents, participants signed a consent form and completed a pre-study survey on family demographics, current mobile phone use, and parenting experience. I also provided a one-page instruction sheet that explained how to use the system. They then were registered with Baby Steps Text, and participants started receiving screening milestone questions. After the first visit, the research team did not interact with participants until one month had passed. After one month, I mailed an exit package that included a study completion brief, compensation, and a post-study survey with a self-addressed, pre-stamped return envelope. Participants received \$20 for their study participation and \$10 each for the pre- and post-study survey completion (the post-study survey had to be mailed back). Participants without unlimited text messaging plans were also compensated an additional \$20 at their sign up to cover any texting fees incurred using Baby Steps Text.

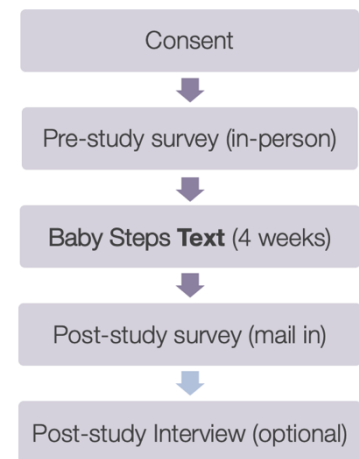


Figure 12. Study Procedure

4.4 RESULTS

Here I describe the results of the feasibility study including how Baby Steps was used, the ability for participants to understand the syntax, their overall experience, and their use of the system after the study was complete.

4.4.1 Baby Steps Text Usage

Among the 14 parent participants, one (P2) withdrew from the study in the early stage by replying “STOP.” The remaining 13 participants successfully used text messages to track their child’s development and completed the study. Over their one month-long participation, Baby Steps sent 557 messages and received 520 SMS messages in response from participants. On average, each participant who completed the study received 42.3 messages ($SD = 18.6$) and sent 39.4 messages in response ($SD = 19.7$). The high standard deviation is due to three participants whose child transitioned from one milestone question set age range to another during the course of study (therefore answering 60 questions)

and among them, one participant used Baby Steps Text for two children and responded to 90 questions. Four participants (P1, P5, P6, P8) responded to all 30 milestone questions almost in one sitting (Table 6). I did not see a completion rate difference between the two strategies (e.g., responding to all at once vs. periodically).

Table 6. Overall text activity by 14 participants. The rightmost cell shows a sparkline plot of the month-long timeline of texting activity.

ID	# of Outgoing texts	# of incoming texts	# of active days	# of texts per active day	Texting activity across study duration
P1	43	33	2	16.5	
P2*	8	7	1	7	
P3	36	32	5	6.4	
P4	36	32	5	6.4	
P5	30	30	1	30	
P6	60	60	2	30	
P7	32	33	4	8.3	
P8	30	30	2	15	
P9	62	62	3	20.7	
P10	31	30	4	7.5	
P11	92	91	5	18.2	
P12	33	31	4	7.8	
P13	27	18	3	6	
P14	37	31	6	5.2	
Total	549	520	24	21.7	

(*P2 withdrew from the study)

4.4.2 Use of Baby Steps Text Syntax

For Baby Steps Text to recognize and store user responses in the database automatically, it required participants to use one of the four response options. Although simple, this is not the most naturalistic way of text communication. Thus, I provided an instruction sheet when participants enrolled in the study and continuously included these instructions within the question message itself. As a result, 505 of the 520 user messages were correctly formatted, yielding a very low error rate of 2.88%, and participants were able to self-correct errors early due to an automated response that informed them of syntax errors. About 15% of the text message responses used shortened response types such as Y/S/N, but there was no particular pattern as many participants used both full and shortened responses interchangeably. Due to the low error rate, it appears that the strategy of only repeating full instructions with shorter milestones and using simplified instructions for longer milestones was successful. On the contrary, I only included the “Revisit” feature in the instruction sheet, and it was not included as a response type in milestone question messages. It turned out none of the participants used this feature during the study. One post-study survey respondent mentioned she did not know about the option but would appreciate having it, indicating the importance of including this instruction in the text message.

4.4.3 Participants' Experience with Baby Steps Text

To understand the participants' experience with Baby Steps Text, I sent an exit package with a post-study survey via postal mail. Six out of thirteen participants who completed the study mailed back their survey. While the response rate was low, this is not unexpected for mail-in surveys with lower-income populations. All survey respondents agreed that milestone questions asked by Baby Steps Text were understandable and the system was easy to use, respectively scoring 4 and 4.3 where 5 is strongly agree. They also agreed the total number of milestone questions (30 per 2-6 month age range) asked was appropriate. However, P9 was not as satisfied with receiving one message immediately after the other on a single day: *“the constant stream of texts on the 1st day was a little overwhelming.”* This may have happened due to not clearly explaining that they could take time between questions and that Baby Steps Text would send reminder messages if they did not respond in 48 hours. When asked about message frequency, participants were mixed. Half the survey respondents (n=3) preferred to receive the next

question immediately, indicating their preference to answer in one sitting if possible, and the other half preferred to receive questions periodically. When asked an open-ended question about what they liked about Baby Steps Text, P3 and P9 reported that it was easy, quick, and convenient to use. Several participants also liked receiving useful information that can be helpful to their child in a timely manner without needing them to look for other resources (P5, P9 and P10). P10 noted, *“I really enjoyed participating in this study. I think the texting format is really great and unique. It’s nice to receive personalized texts and not have to go to the Internet to look things up!”*

4.4.4 Continued Usage After Study Completion

In the exit package, I explained to participants that they could choose to remain in the system beyond the study and continue to receive milestone questions. P10 finished the 60-month milestone questionnaire during the study, which was the last questionnaire of the Baby Steps Text system, thus there were no more milestone questions to send. None of the remaining twelve participants who completed the study explicitly opted out of Baby Steps Text. Moreover, 9 out of 12 were still responding to Baby Steps Text three months after the study was completed. Due to a system glitch, Baby Steps Text could not send messages longer than three months. During those three months, we received an additional 385 messages from participants, averaging 42.8 messages. This is a promising result, which may indicate Baby Steps Text system is engaging, and participants find the system worth using outside of study compensation.

4.5 DISCUSSION

As a result of the overall design process and feasibility study, I have uncovered a number of insights for both improving the design specifics of Baby Steps Text, as well as some limitations.

4.5.1 Recommended Changes to Baby Steps Text Design

This feasibility study uncovered several ways to improve the usability and user experience of Baby Steps Text.

Detailed Progress Report

To fit the content of the progress report into a single text message, I had to make it brief, as described earlier. Although the report design of Baby Steps Text worked well for the final set of usability study participants for simulated tasks, when it was sent to the study participants, they were not as satisfied with the minimal data. P3 noted in post-study survey that, *“I felt the report was too high level and vague. Didn’t tell me much of anything.”* To reflect this feedback, the next version of Baby Steps Text will need to convey more information on interpreting each item in the report. One solution could be parents being able to text “more” after a report to receive text-based explanations describing each developmental category, what the child’s results mean, and what they can do besides connecting to the local family health hotline. For users with Internet access, providing a link to a web-based report with more detail could be useful, as was done for @BabySteps in Chapter 3.

Providing Ideas for Activities to Encourage Development

To increase information for parents on how they can support development beyond tracking, especially for those who have children who are not “on track,” I plan to implement a feature to provide activity ideas that are specific to a child’s current age. Because some parents do not have unlimited texting or may get overwhelmed by too many messages, this feature would be opt-in. After a parent finishes responding to 30 questions in one milestone set, the system can send a message asking if they want to receive activity ideas they can do with their child. If they respond yes, the system could send age-appropriate items automatically (e.g., “Activity Idea: Put dry rice inside an empty bottle to encourage shaking and turning behaviors”). Because activities are suggestions rather than screening, no response would be required.

Sentimental Records along with Developmental Records

Previous research indicates that tracking sentimental childhood records, such as memories and “firsts and favorites,” along with developmental records can increase parent engagement [102, 103]. Because most mobile phones are equipped with cameras that allow users to take photos and videos, I see this as an opportunity to help parents keep memories via Baby Steps Text. I asked in the pre-study survey about difficulties participants had in recording memories, and multiple participants stated lack of time as their

number one constraint. P10 said, *“Time is everything to me especially when my day gets crazy with having my oldest in school. So finding time to just to note things down.”* I plan to expand Baby Steps Text to make the process of keeping sentimental records as quick and easy as what they currently do to share photos and videos of their child with family and friends. The new version will allow parents to send photos and videos of their child in MMS with the keyword “#memory” to save those in database automatically. That way, parents who have access to the Internet can view their records on a companion website.

4.5.2 Limitations

Due to the 160-character limit of SMS, we needed to shorten the validated screening tool and thus the validity of content needs to be reevaluated. I also did not follow up with the family health hotline to determine if any of the participants called to request access to services for early intervention, and only tested an English-language version of the tool. As this was an initial feasibility study, participants used Baby Steps for a brief period and I do not have sufficient data to tell how Baby Steps Text impacts parents in the long term (e.g. how it changes parental efficacy, their knowledge on child development, etc.) Therefore, I plan to conduct a longer study to answer questions about long-term engagement, identify ways to combine developmental tracking with sentimental record-keeping, and compare it to other approaches to technology-based developmental screening, which I will present in Chapter 7.

4.6 CHAPTER 4 SUMMARY

In an attempt to reach more underserved populations with child developmental screening, I have developed Baby Steps Text, a tool that prompts parents to track and review developmental milestone data and connect them to resources using only SMS. Baby Steps Text is a completely automated, personalized, and self-contained two-way communication tool, which provides a way to understand how we can include resource-constrained populations who may not yet have constant Internet access or sophisticated smartphones. In this chapter, I presented the design, development, and feasibility study of the Baby Steps Text system. The initial month-long deployment study showed that text messaging is a feasible tool for supporting parents in monitoring and screening their child. Continued use beyond the study participation also indicates that the approach is promising in encouraging parent engagement.

Empirical data from this research helped evaluate **T1** by examining the questions of how simple everyday technology that is widely adopted can be used to support child development monitoring. There are two main contributions of this research: 1) Baby Steps Text system (artifact contribution) brings SMS-based mHealth interventions to a new domain, childhood developmental screening; and 2) the results of the deployment study help expand knowledge about how parent-generated health data can be collected in a simple but thoughtful way, thus expanding the reach of its benefits (empirical contribution). I build on these findings in Chapter 7, in combination with the learnings from Chapter 3, by re-designing and developing the Baby Steps ecosystem to answer questions about long-term engagement in developmental tracking.

Chapter 5.

Understanding User Burdens in Technology

The growth of computing systems over the last few decades has been tremendous. Wearable sensors, smartphone-based applications, websites, and more are both readily available and affordable to end-users. Designers and developers have created new systems targeting applications in domains as varied as health, finance, social networking, productivity, and entertainment. The Human-Computer Interaction community has been instrumental in developing ways to evaluate these types of systems and assess them along a number of dimensions, including usability, user experience, and usefulness. Despite these efforts, many people still fail to adopt systems that could benefit them greatly or abandon systems such as health technologies after very little use, even if they are seeing a benefit. They may also use a

system out of necessity, but it could still have a negative impact on their lives. I suspected one possible cause of these issues is that the burden these systems place on the user is too high. Although there were previous attempts to understand negative impacts of technology on its users from various angles, most studies focused on one aspect and lacked holistic understanding. Also, an umbrella term like user burden was yet to be defined. In this chapter, I introduce the concept of user burden and seek to explore the following to understand it better:

- What is user burden? How does technology impose burdens on its users?
- How can we evaluate user burden systematically yet quickly?

To answer the first question, I conducted literature review, and a number of rounds of discussion among experts in user-centered design of computing systems, followed by interviews. When the initial user burden types were outlined, I then moved to create a scale that can measure user burden systematically through steps outlined later in the chapter. Exploring these questions helped me answer **RQ2** (how does technology impose burden on its users and how can we measure them?) and partially evaluate **T2**. I worked together with 3 collaborators on this investigation.

5.1 BACKGROUND

I defined *user burden* as the negative impact that computing systems might place on the user. While burden includes issues with usability and user experience, it can also include other aspects that may be more subjective in nature and more dependent on individual differences. The concept of user burden as I define it is unique in that it focuses primarily on all the aspects of a system that may negatively impact a user's ability to use and tolerate it. It is important for designers to be able to understand and assess the amount of user burden in each of these areas in their designs so that they can make an effort to reduce as much of it as possible. Different types of user burdens can make it difficult for people to initially adopt or continue to use a system and may have a negative impact on the overall user experience. In addition, user burden may be present even when the user has continued to adopt and use a system, but it may be decreasing their overall user experience and engagement.

There are a few user experience methods that touch on these aspects in an indirect way (e.g., by assessing error rates in a usability test or task load), however, assessing user burden is not something that is part of standard usability practice currently. In addition, although user burden may be measured by more objective measures, whether a system places a burden on a particular person needs to be assessed on an individual level. For example, if the cost of a smartphone is fixed at \$300 with a \$40 per month subscription fee, this may be an easy expense for someone who is wealthy, but very burdensome for someone from a resource-constrained background. Likewise, a wearable technology shirt may be too heavy and bulky for someone with a smaller frame, but quite comfortable and unnoticeable for someone with a larger build. Because of these individual differences, gathering feedback from a large number of diverse users is important. Many existing methods of assessing user burden may not scale well, and thus I have sought to develop a validated measure of user burden through the form of a questionnaire, called the User Burden Scale (UBS).

5.2 INITIAL DEFINITION OF USER BURDEN

Based on a review of the literature, and a number of rounds of discussion among experts in user-centered design of computing systems, I determined an initial eight types of user burdens, which included access, emotional, financial, mental, physical, privacy, social, and time-based burdens. I defined each of those initial categories as follows:

- **Access Burden:** The system does not fit with the abilities or cultural background of the user.
- **Emotional Burden:** The system makes the user feel bad or unnecessarily worried.
- **Financial Burden:** The system costs a significant amount of money for the user to initially purchase or to maintain use.
- **Mental Burden:** The system requires significant attention, concentration, or is distracting.
- **Physical Burden:** The system makes the user physically uncomfortable.
- **Privacy Burden:** The system risks revealing information about a user that he or she would prefer not to share.

- **Social Burden:** The system may disrupt the user's ability to create and sustain social relationships.
- **Time Burden:** The system requires frequent use or a significant amount of time to use.

5.3 DEVELOPING USER BURDEN SCALE

I aimed to design a questionnaire that could be used by designers and developers in the field to evaluate existing technologies that have been used in real world situations. Following the requirements and guidelines of scale development procedures used by Yarosh et al. [175], I came up with several requirements for UBS:

- Measure different categories of perceived user burden in using technologies.
- Refer to a specific system but be generic to be applicable to a wide range of technologies.
- Be quick to administer.
- Demonstrate reliability and validity on multiple metrics.
- Be sensitive enough to detect differences between technologies.

5.3.1 Refining Initial User Burden via Interviews

To refine my understanding of the initial definitions of the user burden categories I developed and to refine them further, I conducted an hour-long, semi-structured, in-person interview study with 12 participants (6 male, 6 female). I recruited participants who were at least 18 years old and use computing technologies on a daily basis via university and group email lists of faculty and students, and by word of mouth. Participants provided a list of five technologies they frequently used, and I selected which two technologies to interview and survey about to ensure a broad spectrum of applications. From 12 participants, I collected data about 24 different computing systems. Because the purpose of the interview was to explore participants' experience with computing technologies and the user burden associated with it, interview questions were semi-structured and open-ended to facilitate interviewees in bringing up new ideas during the interview. Interview methods were patterned after Weiss' [156] techniques. As a token of appreciation, participants received \$25 in gift cards for their participation. All interviews were audio recorded and transcribed.

I analyzed interview transcripts using thematic analysis [152], creating thematic connections of interview data. Several iterations of this process produced refined and distinct themes. In addition to this, all interview statements were examined and for the statements implying or identifying user burdens, two authors assigned them to one of corresponding eight user burden scales (Table 7). The results of these interviews indicated that the initial categories I defined were consistent with expressions of user burden from end users, and that I could move forward to a more formal validation of the categories into constructs and the definition of a scale.

5.3.2 Identification & Development of Scales

After defining the initial eight categories of burden, the research team brainstormed and developed preliminary scales for each category of user burden. This process yielded 15 items in each of eight categories. Review and discussion by three experts led to eliminating and re-writing a significant portion of the questions, resulting in a draft of the survey with 64 items (8 items on each scale). Through 5 rounds of pilot studies with over 922 participants using Amazon Mechanical Turk (MTurk), I continued eliminating items to make it brief and concise and to remove questions that were confusing, resulting in 8 scales with 26 items. At this point, the items were measured using a five-point Likert scale ranging from strongly disagree (1) to strongly agree (5), with a mix of positively and negatively framed questions. Throughout pilot studies, tasks given to participants were always completing the most recent user burden scale questionnaire at the moment. In the first round (15 items in each of eight categories), I asked participants to complete the survey on a single, specific system: Facebook. However, to make UBS generalizable to any interactive systems, each of the subsequent rounds of testing asked participants to choose an interactive system that they use frequently or one that they have previously used that they have now abandoned. Based on these preliminary results and discussion amongst the research team, I decided to convert all of the questions to be negatively framed (since user burden was considered a negative experience) and use two different 5-point scales to add additional nuance beyond a simple Likert scale. The two final response types are as follows:

- **Response Type 1 (Frequency/Occurrence):** 0 = Never; 1 = A little bit of the time; 2 = Sometimes; 3 = Very often; and 4 = All of the time.

- **Response Type 2 (Degree/Magnitude):** 0 = Not at all; 1 = A little bit; 2 = Somewhat; 3 = Very much; 4 = Extremely

Table 7. Sample interview prompts and sample coded phrases for each initial category of user burden

Initial Category	Sample Interview Prompt	Interview Participant Statement
Access	[Steam] Have you ever felt difficulty in seeing, hearing, or manipulating the system?	"Sometimes the graphics requirements are too high for my laptop or sometimes it will just mess up. There will be some kind of lag or something, and I can't see properly, and then I'll die."
Emotional	[MS Word] Has using the system ever made you feel bad, other than the frustration that you mentioned already?	"Yeah, well I mean, general software crashes. I've definitely lost information. That makes me feel bad. Sometimes the auto-save doesn't catch it all."
Financial	[MapMyRun] Was the amount that you spent on the system the amount that you expected to pay?	"No. I was very disappointed actually. Of all the apps that I paid for, and I love MapMyRun, but the features you pay for really aren't worth \$6.00 a month."
Mental	[Garmin] What did you have to do to learn how to use the system?	"The Garmin was a little more challenging and there was a lot of trial and error and looking things up a little bit. I had a manual because some of the options and features weren't quite intuitive. But the Google Maps navigation is pretty straight forward."
Physical	[FaceTime] Have you ever felt physically uncomfortable while using the system?	"Yeah. It actually heats up very fast, so it becomes very hot. Also, as with many Apple products, it is rounded, and it's very thin, so it's very hard to set down."
Privacy	[Netflix] Were there any steps that you have taken to ensure your privacy?	"Not really with Netflix, since it's pretty easy to use. I would actually rate movies actively so that it would recommend something that I want to see. ... I don't worry about Netflix using my personal information."
Social	[MapMyRun] How does using this system impact your relationship with others?	"I think... My dad, for example, really quite likes it because when it goes to my Facebook, he always likes it because to him it's probably like because I'm from England so he's in England. He sees when I go to a new place and go for a run. I think he quite likes that. He can see I'm in Seattle or Vancouver or whatever."
Time	[Reddit] Does the time that you spend on Reddit match your desire to use to Reddit?	"No, I'd love to use it a lot less... but I'd usually do it during times of boredom, which means I'm bored a lot."

5.3.3 Principal Component Analysis

To explore whether the initial 8 categories of user burden held up as constructs and to reduce the number of questions with statistical analysis, I deployed the 26-item survey (all negatively phrased with two response types) via Amazon's Mechanical Turk. 300 participants completed this version of the UBS on an interactive system that they frequently use or one that they have previously used but they have now abandoned. A total of 274 responses remained after filtering by location and survey completion time of less than 60 seconds. For the remaining 274 responses, the principal component analysis was used to extract the components, and this was followed by a varimax (orthogonal) rotation. After principal component analysis of factor extraction, only the first six components displayed eigenvalues greater than 1, not eight. The results of a scree test also suggested that the first six components were meaningful. Therefore, the first six components were retained for rotation. Combined, components 1 to 6 accounted for 67% of the total variance.

Questionnaire items and corresponding factor loadings are presented in Table 8. With these results, the research team carefully selected items to be included in the final question set. We chose the 4 items of highest loading from each factor. The *Privacy* and *Financial* constructs had less than 4 items, and we chose all of them. One exception was made for the *Physical* construct, where we chose to include the two items with highest factor loading and then included the 5th item which had slightly lower factor loading but seemed more relevant to the overall construct. In addition to eliminating items, the PCA also resulted in changes in the user burden constructs. Through the PCA, *Mental and Emotional* burden categories were combined into one construct, as were *Time and Social* burdens. We also developed a new construct, *Difficulty of Use* burden, and eliminated the *Accessible* burden construct as it was covered in the other constructs. Overall, these analyses indicated that six distinct factors were underlying the model of user burden and a total of six items were eliminated because they did not contribute to a simple factor structure. The overall identification and development process of the scale is summarized in Figure 13.

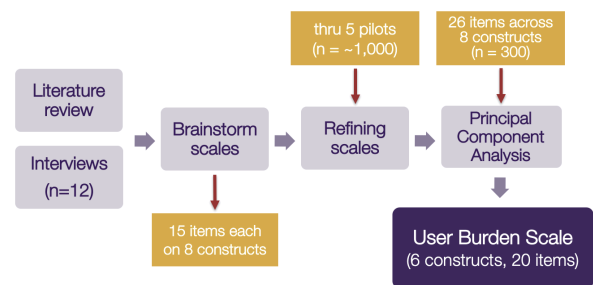


Figure 13. Identification and development process of UBS

Table 8. Factor loadings and communalities based on a principal component analysis (PCA) with varimax rotation for 26 items (N = 274). Items included in the final questions set are highlighted

	Initial user burden group*	Rotated Factor Pattern					
		Difficulty of use	Physical	Time & Social	Mental & Emotional	Privacy	Financial
I need assistance from another person to use [X].	A	.78					
[X] demands too much mental effort.	M	.77					
It takes too long for me to do what I want to do with [X].	T	.69					
[X] is hard to learn.	M	.69					
I get frustrated when using [X].	M	.68					
Information, such as visual cues or sounds, from [X] is hard to understand.	A	.60					
The value of [X] is not worth the cost to me.	F	.47					
Using [X] too much creates physical discomfort.	Ph		.74				
[X] has made me feel physical pain.	Ph		.67				
[X] is not appropriate for my cultural background.	A		.62				
I don't want others to know that I use [X].	S		.61				
Use of [X] is too physically demanding.	Ph		.53				
I spend too much time using [X].	T			.85			
I use [X] more often than I should.	T			.85			
[X] distracts me from social situations.	S			.73			
Using [X] has a negative effect on my social life.	S			.61			
[X] requires me to remember too much information.	M				.66		
[X] presents too much information at once.	M				.65		
Using [X] makes me feel like a bad person.	E				.64		
I feel guilty when I use [X].	E				.64		
[X] forces me to make changes to how I normally use digital technologies.	A				.56		
I am worried about what information gets shared by [X].	Pr					.88	
[X]'s policies about privacy are not trustworthy.	Pr					.83	
[X] requires me to do a lot to maintain my privacy within it.	Pr					.78	
[X] is too expensive.	F						.88
The upfront cost to using [X] is too high.	F						.81

(* A: Access, E: Emotional, F: Financial, M: Mental, Ph: Physical, Pr: Privacy, S: Social, T: Time)

5.4 FINAL USER BURDEN CONSTRUCTS & SCALE

Based on initial exploration and the construct validity using principal component analysis, I refined the initial proposed categories into a model consisting of six constructs, as well as definitions for each construct.

5.4.1 Definitions of User Burden Constructs

Below I provide a definition of each of the final constructs and two or three examples of systems that I considered to placed a high burden on the user in that area.

Difficulty of Use Burden: The system does not fit with the abilities of the user and is difficult to use.

Example systems: i) A photo editing software package with a steep learning curve; ii) A website that is not compatible with a blind user's preferred screen reader

Physical Burden: The system makes the user physically uncomfortable. *Example systems:* i) A body-worn sleep sensor that gives the user a rash if worn too long; ii) A text-entry system that causes repetitive stress injuries in the wrist due to over use

Time & Social Burden: The system may require a significant amount of time to use or disrupt the user's ability to create and sustain social relationships. *Example systems:* i) A mobile food diary that requires several minutes to enter each item of food consumed throughout the day; ii) A mobile application that makes noise and annoys others who are working nearby the user; iii) A system that automatically sends emails to everyone on a user's contact list

Mental & Emotional Burden: The system requires significant attention, concentration, or is distracting, or makes the user feel bad or experience unnecessary worry. *Example systems:* i) An exergame that shames an overweight person who plays it if they are too heavy; ii) A phone-based news application that constantly sends the user disruptive reminders; iii) a system that overwhelms the user with a confusing visual display

Privacy Burden: The system risks revealing information about a user that he or she would prefer not to share. *Example systems:* i) A weight scale that by default automatically posts a user's age and weight to

their social media accounts; ii) A social networking system that reveals personal information to others without the user's consent

Financial Burden: The system costs a significant amount of money for the user to initially purchase or to maintain use. *Example systems:* i) A bicycle GPS system that has a high initial cost and is expensive to replace if damaged or stolen; ii) A video streaming service that requires a costly monthly fee

5.4.2 Final Scale and Question Set

Throughout the questionnaire design, UBS aimed to cover a wide range of user burdens associated with using technologies. After multiple iterations of pilot studies and analysis, the final UBS consisted of 6 subscales representing the 6 constructs defined above with 20 total items. To be able to use UBS for both systems that people currently use and systems they used once but have abandoned, UBS has both a past tense as well as a present tense version. The full text of the scales and questions are presented in Table 9.

Table 9. User Burden Scale items and categories. [X] is the name of the system being investigated. The order of item was randomized and two response types were used as appropriate. Of the 20 items, 13 used Response Type 1 and 7 used Response Type 2.

	#	Final User Burden Scale Question Set and Subscales		RT
		Currently using (present tense)	Abandoned (past tense)	
Difficulty of Use	1	I need assistance from another person to use [X].	I needed assistance from another person to use [X].	1
	2	[X] demands too much mental effort.	[X] demanded too much mental effort.	1
	3	It takes too long for me to do what I want to do with [X].	It took too long for me to do what I wanted to do with [X].	1
	4	[X] is hard to learn.	[X] was hard to learn.	2
Physical	5	Using [X] too much creates physical discomfort.	Using [X] too much created physical discomfort.	2
	6	[X] has made me feel physical pain.	[X] had made me feel physical pain.	1
	7	Use of [X] is too physically demanding.	Use of [X] was too physically demanding.	1
Time and Social	8	I spend too much time using [X].	I spent too much time using [X].	2
	9	I use [X] more often than I should.	I used [X] more often than I should have.	1
	10	[X] distracts me from social situations.	[X] distracted me from social situations.	1
	11	Using [X] has a negative effect on my social life.	Using [X] had a negative effect on my social life.	1
Mental and Emotional	12	[X] requires me to remember too much information.	[X] required me to remember too much information.	1
	13	[X] presents too much information at once.	[X] presented too much information at once.	1
	14	Using [X] makes me feel like a bad person.	Using [X] made me feel like a bad person.	1
	15	I feel guilty when I use [X].	I felt guilty when I used [X].	1
Privacy	16	I am worried about what information gets shared by [X].	I was worried about what information got shared by [X].	2
	17	[X]'s policies about privacy are not trustworthy.	[X]'s policies about privacy were not trustworthy.	2
	18	[X] requires me to do a lot to maintain my privacy within it.	[X] required me to do a lot to maintain my privacy within it.	1
Financial	19	[X] is too expensive.	[X] was too expensive.	2
	20	The upfront cost to using [X] is too high.	The upfront cost to using [X] was too high.	2
Reporting Value		Response Type 1	Response Type 2	
0		Never	Not at all	
1		A little bit of the time	A little bit	
2		Sometimes	Somewhat	
3		Very often	Very much	
4		All of the time	Extremely	

5.5 VALIDATING USER BURDEN SCALE

To provide evidence regarding reliability and validity of the refined scales, I deployed the survey on Mechanical Turk and conducted three different analyses: 1) inter-item reliability; 2) convergent validity with existing instruments; and 3) the sensitivity of the instrument.

5.5.1 Methods

To test the User Burden Scale with a large sample, I deployed an online version of UBS again on Mechanical Turk. I asked participants to complete the new 20-item version of the UBS, the NASA Task-Load Index (NASA TLX), and the System Usability Scale (SUS) two times each: one for a computing system they currently use and one for a computing system they once used but have now abandoned (hence also the test of present and past tense). The order of answering on using vs. abandoned technologies was counter-balanced while the order of UBS, TLX and SUS was always in the order written. To be able to take the survey, I required that participants be at least 18 years old, reside in the United States, and be frequent users of any type of computing system. Qualified participants took the survey on a computing system of their choice. Some popular systems participants named are in Table 10.

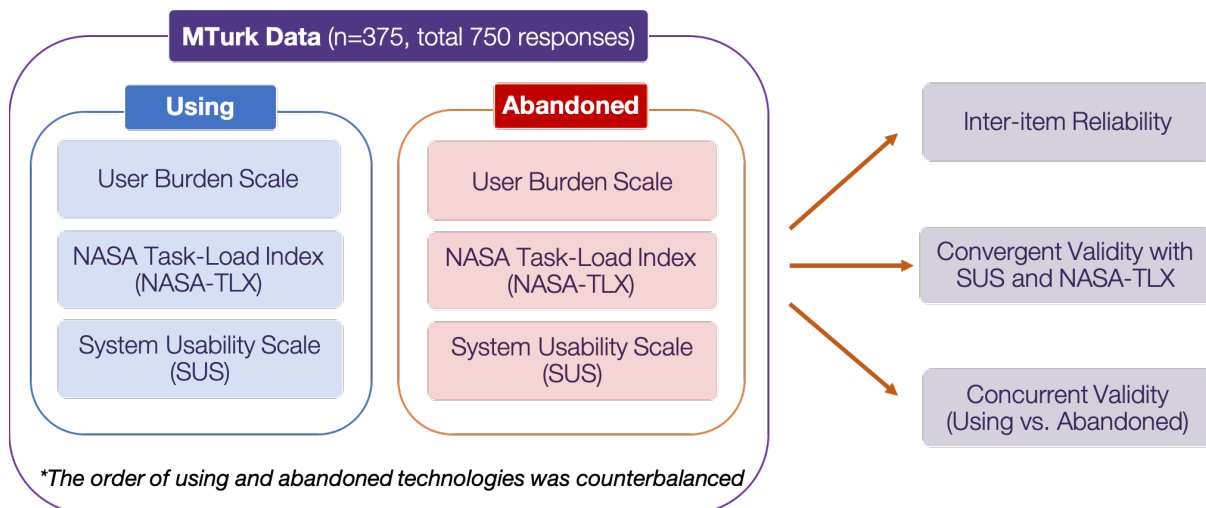


Figure 14. Validation process of UBS

Table 10. List of systems participants rated using UBS

Using System	# of responses	Abandoned System	# of responses
Facebook	78	Skype	65
Netflix	65	Facebook	55
YouTube	42	Wii	31
Gmail	38	Dropbox	30
PayPal	18	Netflix	23
iPad	17	PayPal	22
Kindle	16	Kindle	15
Skype	12	Fitbit	13
Fitbit	10	Gmail	11
iPhone	8	MS Word	9
Other	71 (45 distinct)	Other	101 (50 distinct)
TOTAL	375 responses (55 distinct systems)	TOTAL	375 responses (60 distinct systems)

There were a number of systems currently used by some participants but which had also been abandoned by others. Participants were compensated \$0.80 through Mechanical Turk for their task. A total of 396 participants completed all the surveys. I filtered out some responses (i.e., 9 participants outside the United States and 12 who took too short of time to complete (less than 3 minutes for the entire survey)). The remaining 375 responses were analyzed to examine the reliability and validity of the questionnaire.

5.5.2 Results

In this section, I report the results of the User Burden Scale validation process in terms of inter-item reliability, convergent validity with existing instruments, and concurrent validity sensitivity in detecting differences between technologies.

Inter-item Reliability

With 750 survey responses from 375 participants (two technologies per participant), I calculated the inter-item reliability metrics of the UBS (Table 11). All of the subscales had good internal consistency (> 0.8), with the exception of the *Mental and Emotional* burden subscale, which was still in the acceptable (> 0.7) range [96]. This is promising because it is generally known that Cronbach's alpha is dependent on the

Table 11. Subscale reliability using Cronbach's Alpha coefficient for UBS

Burden Subscale	# of items	Cronbach's alpha
Difficulty of Use	4	0.817
Physical	3	0.814
Time and Social	4	0.862
Mental and Emotional	4	0.728
Privacy	3	0.890
Financial	2	0.891
OVERALL	20	0.883

number of items: fewer items will likely lead to a low alpha, whereas many items will increase reliability. So, it is common for scales with only a few items per construct (3 to 6) to yield a lower alpha [32, 105]. The questionnaire achieved good alpha values across even a small number of items within each subscale. The entire UBS also attained a good Cronbach's alpha of 0.883, indicating the survey has good internal consistency as a whole.

Convergent Validity with Existing Validated Instruments

In addition to the UBS, I asked participants to complete the NASA Task Load Index (NASA-TLX) scale and the System Usability Scale (SUS) to see the relationship between the perceived workload of the system and user burden and that of the system's usability and user burden. I hypothesized that those participants who reported having greater user burden with certain technologies would report higher NASA-TLX scores and lower SUS scores. To test these hypotheses, I ran a Spearman's rank-order correlation. In accordance with my assumption, there were significant positive correlations between the UBS and the NASA-TLX ($r(748) = 0.506$, $p < 0.001$). Correlation between the UBS and SUS was significantly negative ($r(748) = -0.366$, $p < 0.001$). This result indicates that although UBS evaluates technologies on different dimensions than the NASA-TLX and SUS, it has good convergence with it.

Concurrent Validity for Technologies Used or Abandoned

I also wanted to determine if the UBS is sensitive enough to detect differences between technologies with lower user burden (those likely to be still in use) than those with higher user burden (those likely to be

Table 12. Comparing different computing systems based on UBS score using Independent Samples t-Test (n=375 each on using and abandoned technologies)

Burden Subscale	Using	Abandoned	Conditions
Difficulty of Use	M = 1.22 (SD = 1.832)	M = 3.83 (SD = 3.759)	t(748) = -12.102 (p < 0.001)
Physical	M = 0.52 (SD = 1.375)	M = 0.99 (SD = 2.038)	t(748) = -3.696 (p < 0.001)
Time & Social	M = 4.57 (SD = 3.791)	M = 3.14 (SD = 4.009)	t(748) = 4.997 (p < 0.001)
Mental & Emotional	M = 1.39 (SD = 1.940)	M = 2.37 (SD = 3.112)	t(748) = -5.183 (p < 0.001)
Privacy	M = 2.59 (SD = 2.780)	M = 3.66 (SD = 3.926)	t(748) = -4.326 (p < 0.001)
Financial	M = 0.93 (SD = 1.684)	M = 1.38 (SD = 2.132)	t(748) = -3.231 (p = 0.001)
OVERALL	M = 11.21 (SD = 8.979)	M = 15.38 (SD = 13.132)	t(748) = -5.074 (p < 0.001)

abandoned). To compare survey responses on two technologies (using vs. abandoned) by 375 participants, I ran an independent samples *t*-test, where the test variable was the sum of each scale and the grouping was *using* vs. *abandoned*. For all burden scales and the overall questionnaire, the UBS score difference between the two technologies was statistically significant ($p < 0.001$) (Table 12). This suggests that the UBS is sensitive to detect differences between used and abandoned technologies. One thing to note is that for all subscales except *Time and Social*, the user burden was higher for abandoned system than for systems still in use but for *Time and Social*, it was the opposite. It may be that *Time and Social* burdens are not as good predictors of abandonment than the other constructs. One reason for this may be that technologies often become abandoned when they are no longer used, which would mean they were not imposing a time burden or interfering socially.

5.6 DISCUSSION

In this section, I provide guidelines for administrating the User Burden Scale, including how to score and interpret the results and general usage guidelines. I also discuss its limitations and future directions.

5.6.1 UBS Scoring and Analysis and Guidelines for Use

The User Burden Scale uses two 5-point scales (ranging from 0 to 4). This allows the scale to have a higher score resulting in a higher level of user burden. Given that there are 20 total questions, the maximum score is 80 and the minimum score is 0. So that the score is comparable across systems and across users, I recommend that participants be required to answer every question. If a participant feels that they cannot answer a particular question or that it is not applicable, they should be instructed to choose the 0 value for that item since it is likely that the item is not a burden. Because each subscale had a good alpha value, survey administrators may also choose to only administer the subscales relevant to the system of interest. If the administrator chooses, he or she can calculate the score for each subscale by calculating the mean for items within each burden category to determine which constructs of user burden seem to be contributing the most burden.

Figure 15 shows UBS score distribution from the initial 750 data points. Based on this, I present a score guideline with letter grades in Table 13 (A: top 15%, B: next 30%, C: next 40%, D: next 10%, and F: bottom 5%.) I provide cutoff scores for each subscale as well as overall UBS.

The validity tests described above were administered using Survey Gizmo (surveygizmo.com), where the first question asked participants the name of the system they were evaluating. I then piped the system name into all of the questions indicated by [X] in Table 9 and had the survey tool randomize the

Table 13. Score guidelines for UBS

Burden Subscale	# of items	User Burden Score Range				
		A top 15%	B next 30%	C next 40%	D next 10%	F bottom 5%
Difficulty of Use	4	0	0.25	0.5 - 1.5	1.75 - 2.25	2.5 - 4
Physical	3	0	0	0.33 - 0.67	1-1.33	1.67 - 4
Time & Social	4	0	0.25 - 0.5	0.75 - 2.25	2.5 - 2.75	3 - 4
Mental & Emotional	4	0	0.25	0.5 - 1	1.25 - 1.75	2 - 4
Privacy	3	0	0.33	0.67 - 2.33	2.67 - 3.33	3.67 - 4
Financial	2	0	0	0.5 - 1.5	2 - 3	3.5 - 4
OVERALL	20	0 - 0.15	0.2 - 0.45	0.5 - 1.2	1.25 - 1.7	1.75 - 4

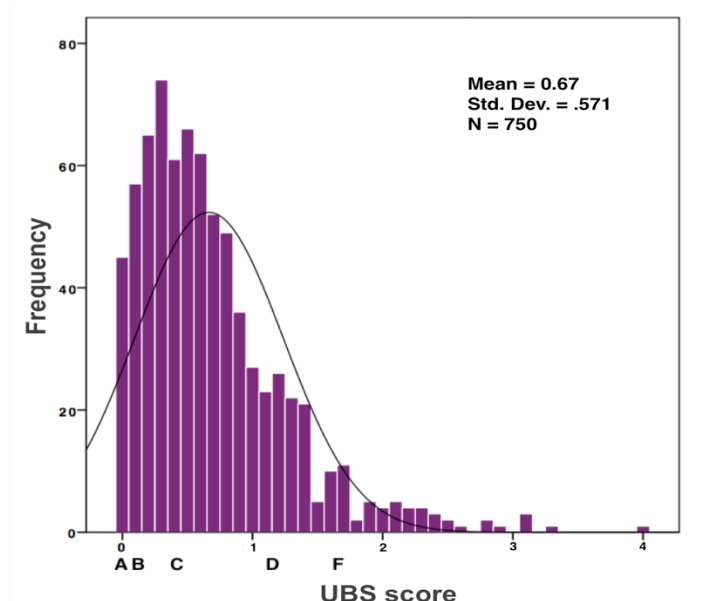


Figure 15. UBS score distribution (n=750)

order of the questions for all participants. I believe that the scale could also be administered on paper if the system name was completed ahead of time (or [X] was replaced more generally by “the system”) and each sheet printed had the questions in random order, but I have not yet validated the scale for use on paper. The 20-item version of the scale took approximately 2 to 3 minutes to complete. I also encourage administrators to consider whether participants are currently using the system of interest and use the present tense version, or if they are evaluating a technology that has been abandoned and use the past tense version (see Table 9). I should also note that the scale assumes that the user has been using the system for some unspecified amount of time, and thus is intended more for evaluation of fully developed systems or later stage prototypes that are fully deployable, rather than early mockups or low-fidelity, non-functional prototypes.

Another consideration for use is interpreting *why* the system is causing the user to be burdened. During interviews, there were several instances when people had a hard time distinguishing between the burden of the system based on its design or its content. For example, it would be as if someone were evaluating Gmail, and they were asked to respond to questions about times when it had made them feel badly, they then mentioned the content of specific emails or the quantity of emails as the source of bad

feelings in addition to issues with the interface itself. A few indicated that they were not sure that they would fault Gmail for making them feel bad because of this. The UBS does not currently distinguish between interface design and content, and so the user burden scores should be interpreted as a combination of the two. If further discernment is required, I recommend follow up studies with survey respondents to probe exactly what about the system is causing the user to feel burdened. Administrators could consider adding open-ended questions after administering the survey to allow participants to elaborate on the specific causes of user burden.

I believe that the UBS can be useful in helping designers to determine different types of trade-offs in their design. While it is good to reduce the overall level of user burden across all categories for a given system, it may not be possible to reduce all of them. In addition, it may be that by reducing one burden, another one increases. For example, to help reduce the privacy burden, a designer may add in a rich set of privacy control features. However, this may increase the difficulty of use burden by requiring the user to spend more time trying to understand and maintain their privacy settings. Because the User Burden Scale has valid subscales, designers can look at individual scores for each subscale to determine which aspects of the system are contributing most to user burden. Because there are an unequal number of questions within each subscale, however, average scores should be used rather than total scores when comparing across burden type.

5.6.2 Limitations

Although I obtained good validity results for the UBS, I acknowledge a few limitations at this time. UBS is still in its early stage of validation and has not yet been tested on a large scale with large Ns for researchers or practitioners aiming to improve their technology design. Validating the questionnaire in practice with designers or researchers could provide more insights on how to best use the UBS. I encourage researchers and practitioners to use UBS and provide feedback on the usefulness of the results so that I can continue to develop and refine the scale for broader ranges of use. In addition, the scoring guidelines we provide in this paper are based on 750 responses on ~150 technologies. I expect scoring guidelines may change as people use it more and the community builds a larger database.

There are other aspects of validity that still need to be tested. For one, I did not conduct a test-retest validation due to the difficulty of following up with online, anonymous participants. I also have not yet run predictive validity to determine if the UBS can predict if someone might abandon a technology later. There are also some general issues and limitations with numerical scales and subjective measures [162, 81]. Although user burden is intended to be inherently subjective, I do encourage administrators to use the UBS in combination with other more objective tests to gain an overall complete picture of user experience. Finally, although the user population on Mechanical Turk, which I used for tests of validity, is relatively diverse for an Internet sample [136], it would still be prudent to test whether the UBS holds across different populations and different cultures.

5.7 CHAPTER 5 SUMMARY

In this chapter, I described the design and validation of a new scale for assessing user burden in computing systems, called the User Burden Scale (UBS), in an attempt to answer **RQ2**. User burden is a model for characterizing the ways that computing systems might have a negative impact on the user across six different constructs: 1) difficulty of use, 2) physical, 3) time and social, 5) mental and emotional, 5) privacy, and 6) financial. I believe that user burden is a unique but important view of overall user experience that has not yet been supported through specific, lightweight measures.

This research makes methodological and theoretical contributions. In an effort to address **T2**, I have defined the concept and constructs of user burden. Also, by creating a quick and easy measurement assessing user burdens in using technology, I provide useful methods that can provide reliable, valid results to make a methodological contribution to human-computer interaction researchers. I hope that this scale can be useful for researchers and practitioners alike in understanding the ways that computing systems can have an impact on the user's lives beyond just issues of usability and enjoyment.

Chapter 6.

Understanding User Benefits in Technology

Since the human-computer interaction community started in 1982 with the first CHI conference, the role of computers in people's lives has drastically changed. While computing systems began primarily as tools for improving productivity at work, they have become so much more. People use technology of many different types as sources of communication, information, social connection, entertainment, comfort, experience, income, and health and wellbeing, just to name a few. As computing's role and people's goals for using it have shifted over the years, so has the need to evaluate how well computing systems have been designed to fill those roles and achieve those goals. Simple usability, ease-of-use, and efficiency metrics were sufficient for productivity-driven apps such as e-mail and spreadsheets, but now that people are turning to computing for broader life goals such as improving health or seeking

relaxation, new metrics are needed to expand our traditional notions. The field of user experience has recognized this need to evaluate computing systems beyond simple usability metrics, and thus has begun to look at metrics such as engagement, experience, enjoyment, and aesthetics.

In the previous chapter, I developed a model and validated a measure of user burden, which focused on understanding the ways that computing might be having a negative impact on people's lives. While investigating user burden, I came to realize it is also important to understand perceived benefits of technology, because the theme of 'continued use despite high user burden because I benefit from it' kept on emerging throughout. Thus, to build on that work and to complete my investigation of **T2**, I developed a complementary model and measure that focuses on positive aspects of computing systems that bring benefit to their users, *user benefit*. In this chapter, I describe the development and validation process of the User Benefit Scale (UBeS), a new 25-item scale with seven subscales designed to measure the user benefit along the seven constructs of the user benefit model in a wide variety of computing systems. This investigation helped me answer **RQ3** (how does technology provide benefits to its users and how can we measure it?)

6.1 INITIAL DEFINITION OF USER BENEFIT

Based on literature review and through rounds of brainstorming and discussion among experts in user-centered design of computing systems, I determined nine initial types of user benefit: accessibility/usability, emotional, experience, financial, health/well-being, informational, reliability, social, and time benefits. I defined each initial benefit type as follows:

- **Accessibility/Usability Benefit:** The system fits with the user's abilities and background.
- **Emotional Benefit:** The system provides emotional support or evokes positive emotions when using it.
- **Experience Benefit:** The system creates new and/or satisfactory experiences for its users.
- **Financial Benefit:** The system provides a good value for its price.
- **Health/Wellbeing Benefit:** The system helps the user achieve their health goals and become healthier.

- **Informational Benefit:** The system provides information or awareness about phenomena both externally and internally.
- **Reliability Benefit:** The system is not prone to failure and performs and maintains its functions continuously.
- **Social Benefit:** The system helps the user connect with others and provides a sense of belonging.
- **Time Benefit:** The system helps the user save time, accomplish tasks efficiently, or improve their productivity.

6.2 DEVELOPING USER BENEFIT SCALE

I wanted to develop a validated questionnaire that is quick to administer by researchers and practitioners in the field who want to maximize the user benefit of their products, while covering diverse types of user benefit in technologies. To develop the scale, I followed a questionnaire design process I used for User Burden Scale, which was adopted from Yarosh et al. [175]. My goals for the scale were as follows:

- Measure different types of user benefit in a variety of interactive computing systems.
- Refer to a specific system but be generic enough to be applicable to a wide range of technologies.
- Be quick to administer.
- Demonstrate reliability and validity on multiple metrics.
- Be sensitive enough to detect differences between technologies.

6.2.1 Refining User Benefit via Interviews

To further understand user benefit in technologies and refine initial user benefit types, I conducted semi-structured interviews with 17 participants. I recruited participants via the university mailing list, and each participant shared their experiences with two technologies they liked and appreciated. To ensure I interviewed participants on a diverse spectrum of technologies, I asked interested participants to provide a list of five technologies they liked when they signed up via email, and I selected which two technologies to focus on the interview to minimize overlapping. As a result, I was able to investigate 29 different systems, ranging from mobile/desktop applications to smart home devices.

Table 14. Sample interview prompts and coded phrases for each initial category of user benefit

Initial Category	Sample Interview Prompt	Sample Quotes
Accessibility	[Amazon Echo] What changes have you made to your preferred technology practices to use the system?	"I guess I wasn't used to using it at first because I never had to talk out loud to do certain tasks. But then I guess I got used to it so it's pretty straightforward to me."
Emotional	[Snapchat] Please describe the emotions you feel when you use the system.	" I'm happy. Like just being able to talk to my friends... like we say really funny and stupid stuff on snapchat sometimes."
Experience	[Lyft] Please tell us your experience with the system.	"First time it felt weird cause I've never ride in a car with a stranger. But it just became part of the things you know. My overall experience is positive. It's just convenient. "
Financial	[Spotify] How much do you think the system is worth compared to what you paid for it?	"Yeah I think so. I'm on the family plan. I used it for 6 months and I paid 20 dollars in total. It's really worth it for me and I don't have ads."
Health	[Fitbit] How does the system motivate you to live a healthy life?	"I think it reminds me to work out more. Having that awareness is helpful. I feel like I really had no idea before [I started using Fitbit]. But it made me more aware of how much I exercise in my life. "
Informational	[Yahoo News Digest] How does the system help you feel more informed?	" If something significant happened it would always pop up in Yahoo News Digest. Like what are the ten most significant things you should know about? They did a really great job of selecting those."
Reliability	[Google Maps] How reliable do you think the system is?	"It's reliable. I don't expect the system to fail or show me some incorrect information. They are really obsessed about accuracy."
Social	[Facebook] What are the changes the system brings to your social life?	" I stay connected with people that I don't see on a regular basis. So, when I do see them I kind of have an idea about what's been going on in their lives and they have an idea about what's been going on in mine. It's really funny that I'll talk to people that I haven't talked to in a while, and they are like 'oh yeah you've been biking a lot'."
Time	[One Bus Away] Do you feel more productive/efficient because you use the system?	"Yeah I do. I just feel that I was able to predict my schedule a little bit more, not like in a huge way that it would change my entire day... but it helped me to save a few ten minutes or so here and there. Because One Bus Away... like I mentioned I would end up at a bus stop way earlier [before I started using One Bus Away.]"

Because the purpose of this interview study was to explore user benefit beyond the literature and discussion among a research team, interview questions were semi-structured and moderately broad to solicit unbounded new ideas from participants. The average length of interviews was about 48 minutes, excluding greeting and brief introduction. I gave participants \$20 Amazon.com gift cards to thank them for their participation. All interviews were audio recorded for future reference. I did not fully transcribe audio recordings, but wrote a summary of the interview, selectively transcribed representative quotes and shared other notable things with the research team. Table 14 presents sample interview prompts and quotes.

As a result of the interview study, I was able to broaden and enhance the definitions of benefit types. For example, one of the participants (P14) mentioned that she likes Lyft because it helps other people make money. That led me to expand the meaning of social benefit beyond one's network and consider impact technology on society.

6.2.2 Identification & Development of Scales

After defining and refining the initial nine user benefit types, the research team brainstormed 15 questions on each benefit type, which reflected various aspects of benefit we previously learned through the literature review and interviews. For example, the Experience Benefit included questions related to cultural experiences, new and past experiences, and UX/UI related questions.

To reduce the number of questions and create a concise questionnaire, I ran several rounds of surveys with participants from Amazon's Mechanical Turk (AMT). I asked participants to complete the survey on a technology they liked and appreciated the most. Questionnaire items were measured using a five-point scale ranging from strongly disagree (0) to strongly agree (4). In addition to this, I asked participants to indicate whether each question was 1) understandable and 2) applicable to the system they chose when there were 15 items in each benefit type. I used this information to exclude items that were not largely applicable or were confusing. I also included several attention tests in the survey (e.g., "I am reading this and will choose neutral," "Choose strongly disagree if you have used [X]" where [X] is the system they were answering the question about) and measured time spent on each page of the survey and total time spent on the survey to filter inattentive responses.

With data from 100 participants, I gradually eliminated items and reduced the number of questions on each benefit type from 15 to 8 and then to 4.

6.2.3 Principal Component Analysis

When I reduced the questionnaire to 4 items on each of 9 benefit types, I deployed a 36-item survey on AMT with 400 participants. I conducted a factor analysis (principal component analysis with varimax rotation) with the remaining 373 responses after filtering to extract the primary constructs from the user benefit types I initially defined.

The result of principal component analysis of factor extraction suggested that there were eight components with eigenvalues greater than 1, not nine as I had originally defined them. The results of a scree test also showed that the first eight components were meaningful. However, because Factor 8 seemed more loosely bonded than other factors and because it only had two items, I decided to exclude it. In addition to this, the analysis resulted in several other changes to the initial benefit types. Emotional and Experience benefits were merged into a single category, resulting in a single *Experience* construct in the final user benefit scale and a new construct, *Ease of Use*, emerged, embracing several different initial benefit types including Accessibility/Usability and Reliability. Combined, Factors 1 to 7 explained 60% of the total variance. Table 15 details the 36 survey items used for factor analysis and their corresponding factor loadings. I chose the three to five items with the highest loading from each factor. Items included in the final User Benefit Scale are highlighted. Figure 16 summarizes the process of identifying and developing User Benefit Scale.

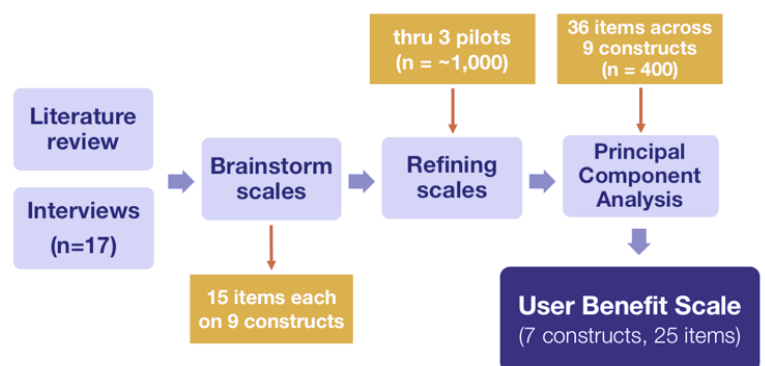


Figure 16. Identification and development of UBeS

Table 15. Factor loadings and communalities based on principal components analysis (PCA) with varimax rotation on 36 items. Highlighted are items included in the final scale.

	Initial user benefit group*	Rotated Factor Pattern							
		Ease of use	Time	Experience	Health	Financial	Social	Informational	
I expect [X] will function properly when I use it.	R	.699							
The content of [X] is well-organized.	R	.674							
[X] can be used by people of my age.	A	.647							
[X] functions in a way that is expected.	I	.637							
The look and feel of [X] is well done.	Ex	.625							
I can use [X] without assistance from other people.	A	.598							
Visual cues and sounds of [X] are easy to understand.	A	.485							
[X] increases my productivity.	T		.865						
[X] helps me manage my time better.	T		.840						
Using [X] helps me to be efficient.	T		.837						
[X] enhances my work effectiveness.	T		.795						
I have saved money as a result of using [X].	F		.539						
I feel relaxed when I use [X].	Em			.746					
Using [X] helps me relieve stress.	Em			.693					
Using [X] makes me feel good.	Em			.625					
I'm excited by the new experience [X] provides.	Ex			.609					
[X] gives me peace of mind when I use it.	Em			.552					
[X] is innovative and provides new experiences.	Ex			.495					
[X] keeps my data safe without me having to worry.	R			.358					
I find it easy to use [X] in support of my health.	H				.857				
[X] supports my goals for leading a healthy lifestyle.	H				.854				
[X] makes a positive impact on my health.	H				.835				
Health-related aspects of [X] are useful.	H				.827				
[X] is reasonably priced.	F					.870			
[X] is affordable to me.	F					.823			
[X] provides a good value for its price.	F					.774			
Using [X] makes me feel more connected to society.	S						.653		
[X] contributes to our society.	S						.632		
[X] is generally benefiting people in the society.	S						.541		
Using [X] has a positive effect on my friends and family.	S						.506		
[X] provides fresh perspectives on daily things.	Ex						.390		
Information I find with [X] is useful.	I							.744	
I rely on [X] for information on certain topics.	I							.670	
Information I find with [X] is reliable.	I							.665	
[X] works with my preferred technology practices (e.g., cross-device, screen reader, etc.)	A								0.726
[X] provides similar experiences across the devices.	R								0.583

6.3 FINAL USER BENEFIT CONSTRUCTS & SCALE

Through the aforementioned process, I have developed a finalized version of the User Benefit Scale (UBeS) consisting of 25 items across 7 user benefit subscales relating to the key constructs of user benefit. Below is a definition of each final benefit construct with examples supporting its definition.

6.3.1 Definitions of User Benefit Constructs

I provide a definition of each of the final constructs and two or three examples of systems that I considered to provide a high benefit to the user in this area.

Ease of Use Benefit: The system provides the user with a good experience regardless of their age and abilities and works consistently and reliably. *Example systems:* i) An e-book reader that gives the user flexibility to adjust the display based on their reading ability; ii) A communication system that can be used by people across cultures and generations; iii) An online cloud document editing system that auto-saves the document and reduces the risk of data loss

Experience Benefit: The system creates new and /or satisfactory experiences to its users or evokes positive emotions. *Example systems:* i) A virtual reality experience that provides the user with a new and interesting experience; ii) A mobile game with cats that delights the user and helps them relax while playing it; iii) A leak detection sensor that will text message a user if there is a problem, so they do not have to worry

Financial Benefit: The system is affordable and provides a good value for its cost. *Example systems:* i) A free navigation app that shows real-time traffic information and routes to avoid expensive tolls; ii) A paid application supporting useful features that are not available in other free applications

Health & Wellbeing Benefit: The system helps the user achieve their health goals or become healthier either physically or mentally. *Example systems:* i) A food journaling app that teaches about good nutrition and motivates the user to monitor their diet; ii) A music app with good workout playlists that helps the user power through exercise; iii) A mobile application that helps users perform mindfulness exercises to destress after a difficult day

Informational Benefit: The system provides reliable and thorough information on a variety of topics of interest to the user. *Example systems:* i) A news app with accurate and reliable articles; ii) A sports app that encompasses various types of sports and provides real-time play-by-play updates; iii) A smarthome sensor that notifies the user when things in the home need maintenance

Social Benefit: The system helps the user connect with other people and/or positively contributes to society. *Example systems:* i) A social networking system that connects the user to friends and family both near and far; ii) A sharing economy service that creates new opportunities for people to earn income.

Time & Productivity Benefit: The system helps the user save time, accomplish tasks efficiently, or improve their productivity. *Example systems:* i) A team collaboration tool that helps organize, track, and manage teamwork; ii) A calendar app that organizes your schedule and reminds you of upcoming appointments; iii) A food ordering app that lets you order meals to be delivered to help save a trip to a restaurant

6.3.2 Final Scale and Question Set

I present the full text of the UBeS in Table 16. To be able to use the scale across various systems, I used [X] in each item, where [X] can be replaced by the system the UBeS administrator wants to investigate. The [X] can also be replaced more generally with the phrase “the system.”

6.4 VALIDATING USER BENEFIT SCALE

To enhance the reliability and validity of the User Benefit Scale further, I deployed the final version of UBeS and conducted three analyses: 1) inter-item reliability; 2) concurrent validity with existing instruments; and 3) the sensitivity of the instrument. This section describes the validation process.

6.4.1 Methods

To test reliability and validity of the scale, I conducted another series of tests on Amazon’s Mechanical Turk. In this validation process, I had two sets of surveys, with each set including the UBeS, System Usability Scale (SUS) [91], and Perceived Ease-of-Use (PEOU) from the Technology Acceptance Model (TAM) [51]. I asked participants to complete one set of surveys on a technology they really liked and the

Table 16. User Benefit Scale categories and questions ([X] is the name of the system being investigated)

Subscale	#	User Benefit Scale Questions
Ease of use	1	I expect [X] will function properly when I use it.
	2	[X] can be used by people of my age.
	3	The content of [X] is well-organized.
	4	[X] functions in a way that is expected.
Experience	5	I feel relaxed when I use [X].
	6	Using [X] helps me relieve stress.
	7	I'm excited by the new experience [X] provides.
	8	Using [X] makes me feel good.
	9	[X] gives me peace of mind when I use it.
Financial	10	[X] is reasonably priced.
	11	[X] is affordable to me.
	12	[X] provides a good value for its price.
Health	13	[X] supports my goals for leading a healthy lifestyle.
	14	I find it easy to use [X] in support of my health.
	15	[X] makes a positive impact on my health.
Informational	16	Information I find with [X] is useful.
	17	Information I find with [X] is reliable.
	18	I rely on [X] for information on certain topics.
Social	19	Using [X] makes me feel more connected to society.
	20	[X] contributes to our society.
	21	[X] is generally benefiting people in the society.
	22	Using [X] has a positive effect on my friends and family.
Time	23	[X] increases my productivity.
	24	[X] helps me manage my time better.
	25	Using [X] helps me to be efficient.

other set on a technology they did not like. Some popular technologies participants chose to review in the survey are in Table 17. I chose not to use the perceived usefulness scale of TAM and only used PEOU because the perceived usefulness scale was highly work/task-oriented. As I let participants freely choose

any technologies they wanted, it was difficult to find the correct term to substitute for every possible situation for which participants might use a certain technology.

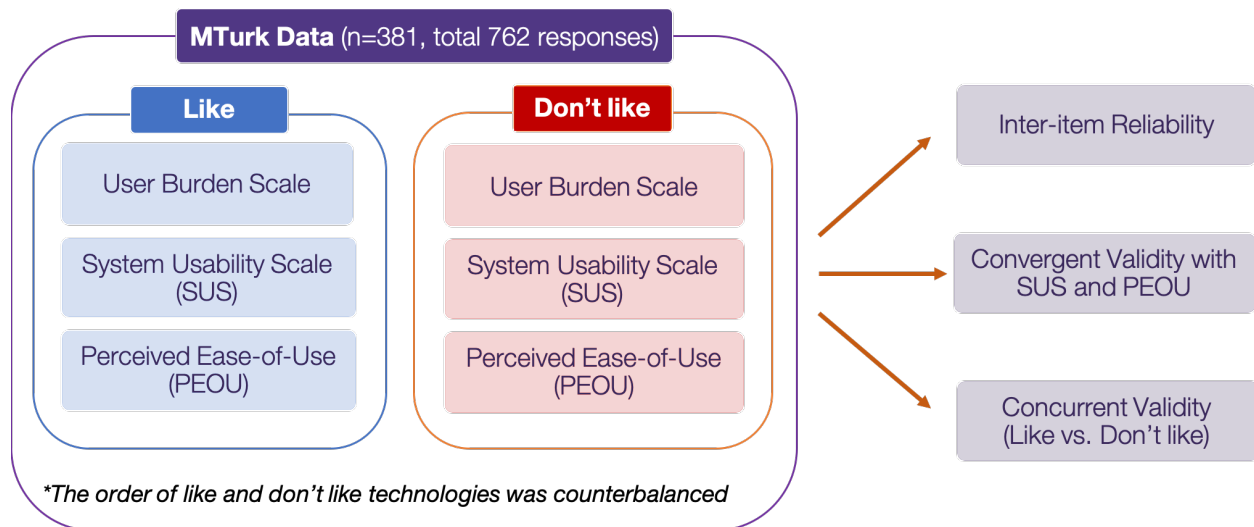


Figure 17. Validation process of UBeS

I had a set of requirements for participants on AMT to complete the validation task: they had to i) be 18+ years old; ii) live in the United States; and iii) take the survey seriously. Participants who self-reported meeting all requirements qualified to complete the task and responded to the surveys. Because

Table 17. List of popular systems evaluated with UBeS

Like		Dislike	
System	# of responses	System	# of responses
YouTube	54	Facebook	47
Facebook	43	Twitter	44
Google Maps	22	Snapchat	33
PayPal	17	Skype	24
Kindle	15	Pokémon Go	19
Instagram	12	Uber	13
Snapchat	8	Google Maps	12
Twitter	7	PayPal	10
Fitbit	7	Amazon Echo	9
Other	196 (118 distinct)	Other	170 (105 distinct)
TOTAL	381 (127 systems)	TOTAL	381 (114 systems)

Table 18. Scale reliability using Cronbach's alpha coefficient (n=762)

Benefit Type	# of items	Cronbach's alpha
Ease of use	4	.832
Experience	5	.954
Financial	3	.899
Health	3	.909
Informational	3	.831
Social	4	.857
Time	3	.920
OVERALL	25	.964

the survey diagnostics tool showed that it would take about 10 minutes to complete two sets of three surveys (UBeS, SUS, and PEOU), I compensated \$1.2 to participants. Similar to what I did during the questionnaire reduction process, I included attention tests in the survey and measured time spent on the survey to use them in filtering inattentive responses. Overall, I received 415 responses to the entire validation set of surveys. After filtering, I used the remaining 381 responses to examine the reliability and validity of UBeS.

6.4.2 Results

In this section, I report the results of three validations I conducted: 1) inter-item reliability, 2) concurrent validity with existing instruments, and 3) sensitivity of the UBeS in detecting differences between technologies.

Inter-item Reliability

To measure scale reliability and internal consistency of the UBeS, I ran reliability analysis on 762 survey responses from 381 participants (two technologies per participant). Table 18 details the result. To summarize, the alpha coefficients suggest that all user benefit subscales have relatively high internal consistency (> 0.8) [96]. The overall scale also has good internal consistency.

Concurrent Validity with Existing Instruments

To compare UBeS with previously established measures, I asked participants to complete adapted versions of the System Usability Scale (SUS) and Perceived ease-of-use (PEOU) from the Technology Acceptance Model (TAM). For adapted SUS and PEOU, I substituted 'the system,' 'this system' and 'electronic mail system' with the name of technology participants entered at the start of the survey (e.g., if participants evaluated "Snapchat," the name of the system was replaced for "this system" within each item to be consistent with the UBeS).

I hypothesized that those participants who find more user benefit in a technology (high UBeS score) would perceive the technology usable and easy to use, resulting in high SUS and PEOU scores. To test these hypotheses and determine the relationship between scales, I ran a Spearman's rank-order correlation. In accordance with the assumption, there was a strong, positive correlation between UBeS and SUS, which was statistically significant ($r(760) = .751, p < 0.001$). There also was a significant positive correlation between the UBeS and PEOU ($r(760) = .720, p < 0.001$). This result suggests that although the UBeS evaluates some distinct constructs of positive aspects of technology, compared to SUS and PEOU from TAM, it has a good convergence with them.

Sensitivity to Differences between Technologies

In addition to its internal consistency and relationship to established measures, I wanted to determine whether the UBeS is sensitive enough to distinguish technologies where the user has a different level of appreciation. To do that, I asked participants to complete one set of the survey on a technology they really liked and another on a technology they did not like. I ran an independent samples t-test, where the test variable was sum of each scale and grouping was like and dislike. The result of the t-test showed that technologies participants did not like earned significantly lower UBeS scores compared to technologies participants liked across all benefit subscales as well as the overall scale (Table 19). From this, I can verify that UBeS is sensitive enough to detect differences between technologies that the user likes and does not.

Table 19. Comparing different computing systems based on UBeS score using Independent-Samples t-Test (n=381 each on like/dislike technologies)

Benefit Type	Like	Dislike	Conditions
Ease of use	M = 18.08 (SD = 1.984)	M = 12.33 (SD = 3.091)	t(760) = 30.560 ($p < 0.001$)
Experience	M = 20.75 (SD = 2.882)	M = 10.14 (SD = 3.605)	t(760) = 44.977 ($p < 0.001$)
Financial	M = 13.35 (SD = 1.980)	M = 10.30 (SD = 3.317)	t(760) = 15.384 ($p < 0.001$)
Health	M = 9.80 (SD = 2.801)	M = 6.12 (SD = 2.292)	t(760) = 19.858 ($p < 0.001$)
Informational	M = 11.67 (SD = 2.288)	M = 7.17 (SD = 2.533)	t(760) = 25.614 ($p < 0.001$)
Social	M = 15.95 (SD = 2.522)	M = 10.37 (SD = 3.231)	t(760) = 26.576 ($p < 0.001$)
Time	M = 10.83 (SD = 3.325)	M = 6.01 (SD = 2.516)	t(760) = 22.573 ($p < 0.001$)
OVERALL	M = 100.43 (SD = 11.498)	M = 62.44 (SD = 13.992)	t(760) = 40.946 ($p < 0.001$)

6.5 DISCUSSION

In this section, I provide guidelines for administering the User Benefit Scale, including how to score, interpret the results (compared to an initial results distribution), and other general usage guidelines and considerations. I also discuss its limitations and future directions.

6.5.1 How to Administer the User Benefit Scale

I administered the User Benefit Scale using SurveyGizmo (<https://www.surveygizmo.com/>), an online survey tool. Because it provides flexibility to pipe the answer of previous questions to later questions in the survey, I first asked respondents to name the system they were evaluating and substituted all [X]'s in the UBeS text with it. I also randomized row orders to avoid clustering the same benefit types. However, if the survey tool a UBeS administrator uses does not support a similar feature or the user wants to administer the UBeS on paper, they may use a fixed row order UBeS and may also replace [X] with a more general term such as “the system” and clearly communicate what “the system” means to

survey respondents (i.e., have a good mutual understanding of which technology for which they are completing the UBeS) Also, in the case where the administrator already knows the name of the technology they want to evaluate, they can pre-insert the name of the technology in the place of [X] as well. However, I have not yet validated the scale for use on paper, and there is some evidence that paper surveys may differ from online surveys [125].

The UBeS uses a five-point scale for all items ('Strongly disagree,' 'Disagree,' 'Neutral,' 'Agree' and 'Strongly agree'), allowing respondents to express how much they agree or disagree with a particular statement. Because the UBeS was designed to measure use benefit of technology on an individual level, the administrator should advise respondents to reflect their own experience when they answer, rather than answering on behalf of others.

It is possible that not all benefit types of the UBeS are applicable to a technology or are of interest to the questionnaire administrator. Because each subscale of UBeS has a good internal consistency with high Cronbach's alpha, the administrator may choose to use only the subscales that are relevant to the system they evaluate (e.g., financial benefits may not be applicable to a system that is free to use). They can also choose to administer the entire UBeS but advise respondents to answer "strongly disagree" to items they do not feel are applicable to the system. If a benefit is not applicable to a technology, it is likely that the benefit does not exist, therefore "strongly disagree" would be the proper answer.

6.5.2 Scoring Guidelines

The reporting values for the five-point scale of the UBeS range from 0 (Strongly disagree) to 4 (strongly agree). With 25 items in the scale, the maximum possible score of UBeS is 100 whereas the minimum possible score is 0. Because all items in the scale are positively framed, stronger agreement to each item means a higher level of user benefit of that kind, and a higher overall score can be interpreted that the system provides a higher user value across user benefit types.

Figure 18 shows the initial UBeS score distribution from 762 data points I collected as part of the validation process. You can see that the score distribution of the technologies people like (top) is much higher than the technologies people do not like (bottom). Based on this initial data, I created scoring

guidelines with a letter grade in Table 20, similar to other scales such as the System Usability Scale [91] or the User Burden Scale in Chapter 5. It includes letter grades for the overall score as well as for each subscale. I expect UBeS score distribution will change with more data points and scoring guidelines will

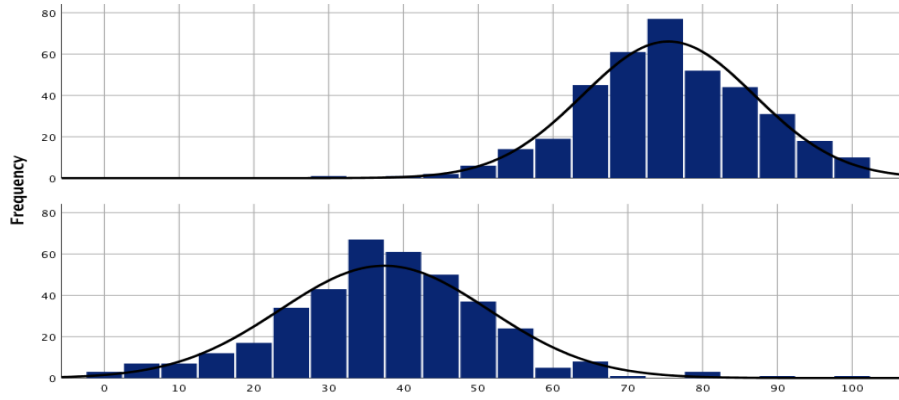


Figure 18. UBeS score distribution (n=762, top is distribution of technologies people liked and bottom is disliked)

need to be adjusted accordingly.

I expect that the UBeS can be useful to both researchers and practitioners who wish to quickly evaluate what aspects of their product users appreciate the most and what aspects can be improved. Though I mentioned in the earlier section that UBeS administrators can choose to use subscales, if they do not have particular user benefit types they are interested in investigating, it can also be good to

Table 20. Score guidelines for UBeS

Subscale	# of items	User Benefit score range			
		A top 15%	B 15 to 45%	C 45 to 85%	D bottom 15%
Ease of use	4	16-20	12-15	7-11	0-6
Experience	5	18-20	13-17	3-12	0-2
Financial	3	12	10-11	6-9	0-5
Health	3	9-12	6-8	1-5	0-1
Informational	3	10-12	7-9	3-6	0-2
Social	4	14-16	10-13	5-9	0-4
Time	3	10-12	6-9	1-5	0
OVERALL	25	82-100	63-81	31-62	0-30

administer the entire UBeS for exploration to compare overall letter grades to other types of similar systems. It is possible that they may find some user benefits they were not expecting to discover. For example, one may not expect a music app to have health and wellbeing benefit, but it may turn out that the music app plays a big role in the user's workout routine and benefits them by helping them engage in healthy habits.

6.5.3 Limitations

Despite the effort to create a rigorous scale by running different types of validations, UBeS has several limitations I hereby acknowledge. First, as this is the first UBeS paper where I introduce its development and validation process, the UBeS has not yet been tested on a large scale. As a result, I was only able to provide the scoring guidelines based on the initial results of 762 UBeS data points in this paper, but I expect the UBeS score distribution will change as it is used by more people and becomes more widely accepted. I request that researchers and practitioners using UBeS to share their results and feedback so that I can continue to refine and adjust scoring guidelines accordingly.

In addition, it is also important to acknowledge that the UBeS is measuring one part of overall user experience. I suggest administrators use the UBeS in combination with other validated scales that measure different aspects of user experience to gain a complete picture. For example, while it is good to have a high user benefit score, a high score does not mean that the technology does not impose user burden that negatively impacts user experience. It is possible that a technology may expose fairly high user burden but because a user finds greater benefit in using it, they sacrifice and endure existing burdens on use. However, when it comes to the point that the perceived user burden exceeds perceived user benefit, they may abandon and stop using the technology. It is hard to capture that through the UBeS alone. Therefore, I encourage administrators to use the UBeS with other complementary scales to obtain a more thorough understanding.

Also, because I designed UBeS to be quick to administer, UBeS only measures what types of and degree of user benefits exist in a technology and does not provide further explanation to its administrator. If the UBeS administrator wants to find out why respondents agreed and/or disagreed to each benefit and what aspects of technology is associated with it, they will need further investigation

beyond administering UBeS, such as also conducting interviews or including open-ended survey responses for more details. Finally, the design of UBeS is such that it assumes that users have already used the system being evaluated, and thus it may not be appropriate for early stage prototypes that are not yet ready for deployment.

6.6 CHAPTER 6 SUMMARY

In this chapter, I presented a model of user benefit and identified seven unique user benefit constructs: 1) Ease of use, 2) Experience, 3) Financial, 4) Health and Wellbeing, 5) Informational, 6) Social, 7) Time and Productivity. I also presented the design and validation process of the User Benefit Scale, a 25-item scale with seven subscales representing each construct.

This work makes methodological and theoretical contributions by designing, developing, and validating a scale that encompasses various types of user benefits that computing systems provide to their users. This research can help researchers and practitioners who want to evaluate and understand the benefit of their systems to improve the design and positive impact that computing systems can have on people's lives. Finally, combining findings from Chapter 5, I have identified the concept and constructs of user burden and benefit to lay the ground work for building new theory in this space.

Chapter 7.

Baby Steps Longitudinal Study

Two previous Baby Steps deployment studies (Chapter 3 & 4) were successful in showing the feasibility of several different everyday technologies in child development tracking. And through analyzing quantitative usage logs and qualitative data from surveys and interviews, I was able to reveal several design opportunities to improve Baby Steps design by enhancing benefits and reducing burdens in child development monitoring. However, participants used @BabySteps, Baby Steps Text and Baby Steps web portal for a brief period of time for each feasibility study, and I could not gather sufficient data to tell if Baby Steps is capable of engaging parents over a longer period of time. As developmental screening

involves regular checks for developmental milestones at up to 22 different intervals across a child's first 5 years of life, I next needed to explore **RQ4**, "how can we design technology to support child development tracking in the long term?" Therefore, I conducted a 20-month long study with 139 families to investigate **T3**. Through this longitudinal study, I was interested in finding out i) how the ecosystem of Baby Steps tools would be used by parents; ii) how specific design features to reduce user burden and enhance user benefit might influence developmental tracking, and iii) whether Baby Steps is successful in long-term engagement.

7.1 SYSTEM DESIGN

Before I launched the long-term deployment study, I iterated on the Baby Steps design based on themes that surfaced in the formative work I describe in Chapters 3 and 4. Also, because Twitter's user base waned over time, there were only a handful of Twitter users from the preliminary survey. For these reasons, I decided to drop the @BabySteps system from the longitudinal study. However, lessons learned from the @BabySteps study contributed to design features described below. The primary sources of redesign decisions came from: i) review of prior research on health monitoring; ii) the results uncovered from my two feasibility studies (Chapters 3 and 4); iii) understandings of user burden and benefit in computing systems (Chapters 5 and 6) and iv) iterative, informal user testing among the research team.

Design features to reduce user burdens

- **Integration of monitoring into everyday technology:** To engage parents with child development tracking more easily, Baby Steps is integrated into everyday technologies that are accepted widely by the general public and are already adopted by a majority of parents. Therefore, it is easy to use and imposes no additional *financial burden*, such as of purchasing a new device or *physical burden* of carrying another device.
- **Free-of-charge service:** Other developmental screening systems often charge a fee to use. Baby Steps system offers screening at no cost to its users because child development tracking should be universal. Also, text version was available entirely via basic SMS (short message service), causing no additional charges of using MMS (Multimedia Messaging Service) or a data plan.

This decision was intended to reduce the *financial barrier* to health care and invite families with diverse socio-economic status.

- **Sentimental memory keeping to celebrate achievements:** Though successful child development screening can give assurance of a child's health and bring peace of mind to parents, the process of developmental tracking can also put emotional burdens on parents because the screener may reveal possible signs of developmental delay. To positively frame the process of child development tracking and reduce unnecessary *emotional burdens*, Baby Steps added sentimental memory keeping to its developmental monitoring. With sentimental memory keeping, parents can record fun memories of their child and what they have done to celebrate their achievements alongside the developmental milestones. This also might help parents build sentimental attachment to the system.
- **Progress reports with careful wording and actionable next step:** To minimize *emotional burden* caused by parental anxiety further, it is important to provide a screening result with clear, actionable steps that can be taken in addition to the screening results [45]. To achieve this, I consulted with experts from WithinReach, a local non-profit organization that provides a free statewide program for child development screening and family health, on the best way to deliver screening results. With their help, I carefully curated progress reports to lessen parental anxiety while conveying developmental progress accurately. In addition, by partnering with WithinReach, the Baby Steps could provide a clear follow up action parents could take when they have concerns about their child's development.
- **Baby Steps delivering timely information:** It is well known that busy parents often forget to regularly complete paper-based or web-based developmental screening questionnaires [103], because remembering to complete questionnaires requires additional mental effort. Thus, to reduce cognitive *mental burden* and *time burden* for parents, Baby Steps delivers questionnaires to them at a convenient moment via text message, instead of requiring them need to remember to complete developmental questionnaires.

Design features to enhance user benefits

- **Accurate and reliable information:** Developmental milestone questions asked by Baby Steps are adopted from the Ages and Stages Questionnaire [36]. ASQ has been researched and validated over decades and used in pediatric clinics and adopting it increases *informational benefits*.
- **Simple and easy-to-use interface:** Through user feedback from previous feasibility studies, I updated the Baby Steps system to have simple and straightforward flow that is *easy-to-use*.
- **Share child memories on the timeline:** Parents can easily share memories they added with their family and friends via web link and/or on social media of their choice with a simple click, which enhances *social benefits* of Baby Steps use.
- **Actionable, yet free-of-charge follow-ups:** If a child is not 'On Track' in any development categories, Baby Steps provides a toll-free hotline number operated by WithinReach, a non-profit organization who is Washington State's Help Me Grow affiliate. This way, parents can have contact points for questions or concerns they may have along the way and can be connected to early intervention services as needed. This contributes to *health & wellbeing benefit*.

7.1.1 Web Portal

The web portal has 7 tabs – Home, Milestones, Progress, Activities, Useful Links, Help and About – and a Settings page. (See Appendix A for screenshots of each page.)

- The Home tab has a timeline where memories can be added, prompts for memory keeping, three unanswered milestone questions and tips on vaccination and general parenting.
- The Milestone tab has age-appropriate milestone questionnaires. Each questionnaire consists of 30 questions covering developmental areas of communication, gross motor skills, fine motor skills, problem solving, and personal-social skills. The tab also has a print icon so that parents can print out the questionnaire in case they want to bring it to a doctors' visit.

- The Progress tab provides progress report generated based on the questionnaire responses from the Milestones tab. It also provides guidance on how to interpret results and possible follow-up actions parents can do such as activities to encourage development and a hotline number to call for further assistance, as needed.
- The Activities tab provides activity suggestions that parents can use to encourage developments. For example, for a 24-30 months old, a fine motor activity suggests to “make a necklace by stringing painted pasta on a string with a stiff tip, like a shoelace.” Activity suggestions are age-appropriate and categorized by developmental areas.
- The Useful Links tab is equipped with free resources parents can utilize. It covers a wide variety of topics, such as free resources on childhood development for parents, health information for children with special needs, low-cost clinics & immunizations and, online parenting communities and resources for families living in Washington State.
- The Help tab has a categorized list of questions parents might ask while using Baby Steps such as from 'what is Baby Steps?', and 'what is child development?' to questions like "I'm answering 'No' to too many questions, should I be worried?".
- The About tab introduces the Baby Steps project and the research team.
- The Settings page allows users to update contact information, add another caregiver and children.

The Control group had access to the Milestone tab, the Progress tab with plain text report, the Useful Links, the Help, the About tab, and Settings page. The Experimental group had access to all the web portal elements, such as a full progress report with the tree visualization, memory keeping with photos and videos, the timeline of a child's growth, and activity ideas parents can use to encourage development. Access differences between groups are summarized in Figure 19.

7.1.2 Text Messaging

Baby Steps Text (Chapter 4) also became a part of the Baby Steps Longitudinal study with a few additional features. As recommended from the Baby Steps Text study, I added memory keeping and activity features to the text messaging system. With the new text messaging system, participants could

receive activity suggestions by texting 'Activity' to Baby Steps, and Baby Steps would send an activity idea to do to encourage their child's development. They can also log memorable moments with a new feature, 'Remember'. Participants could send photos and text of their memories using a keyword 'Remember' and that memory would be added to a timeline on the Baby Steps web portal. And with memories stored in the timeline, participants could receive a random previous memory to reminisce about by texting a keyword 'Memory'. For other existing features and system implementation details, see Chapter 4.

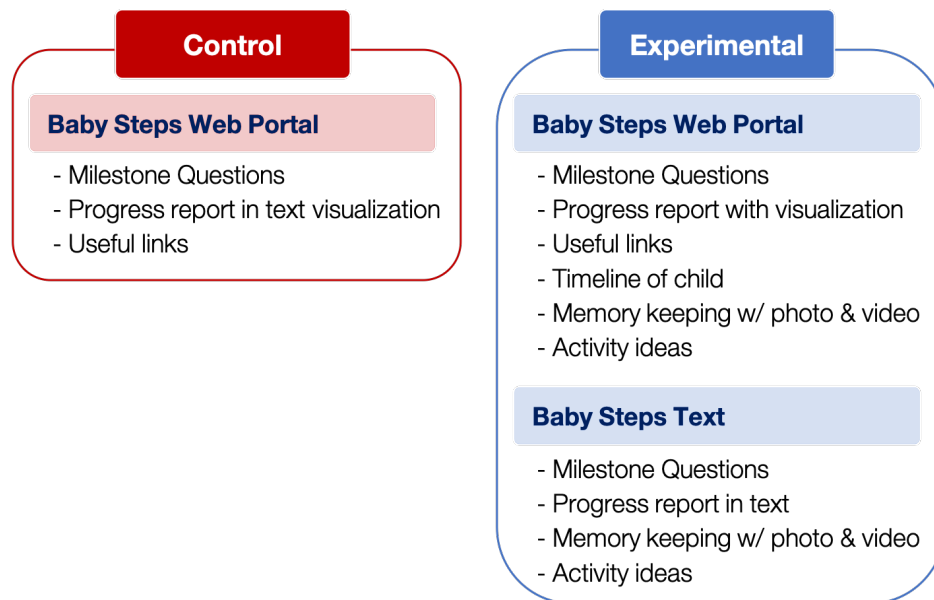


Figure 19. Summary of Baby Steps feature differences between control group and experimental group

7.2 METHOD

7.2.1 Participants

I recruited parent participants via mailing lists, flyers, advertisements on social network services, postings on moms list, and the UW Communication Studies Participant Pool. I aimed to recruit parents of a child between 7 and 12 months old, who live in Washington State, and who have English literacy to be able to read and respond to milestone questions. To ensure all participants met my criteria, I asked potential participants to take a screening survey. Because it was hard to recruit over 100 parents of 7-12

months old at once, recruitment happened on a rolling basis, and the entire recruitment took about 7 months. To examine how specific design elements affected self-monitoring behavior, I chose to have a between-subjects design. A total of 139 participants signed up for the study and I randomly assigned them into a control or an experimental group. Overall, I ended up with 67 participants assigned to a control group and 72 participants assigned to an experimental group.

7.2.2 Study Procedure

Participants who met all the criteria were invited to join the study, and participants who agreed to participate signed a web consent form and completed a pre-study survey on family demographics, child demographics, current parenting experience, and technology use. I then sent an invitation email to participants to create accounts and register on Baby Steps web portal (Appendix B). In previous Baby Steps feasibility studies, I registered participants in the system using information I gathered from pre-study surveys. But because I intended to design this study to resemble real-world usage cases as much as possible, participants had to sign themselves up from Baby Steps web portal. The invitation email contained a brief introduction of Baby Steps, information about child development, and details on study procedures, including the compensation plan. It also included invitation codes to use at the time of registration, which were different for a control group and an experimental group. This access code was used to control the access difference between groups. Once participants finished registration, they could freely use or not use Baby Steps. To study natural usage, I tried to minimize interactions initiated from the research team. The only interaction points triggered from me were 1) to send out a mid-study survey, 2) to send periodic compensation at the end of 4, 8, 12, and 16 months completion, and 3) to introduce a text messaging option when it became available (for an experimental group only).

A mid-study survey was sent out around mid-point of the study to solicit general feedback from participants on their experience of using Baby Steps. A total of 118 participants (control group: 54, experimental group: 64) completed the mid-study survey.

Due to a delay in system development, the text messaging option for the experimental group could not be launched at the start of the study and was added later in the study. Depending on when a

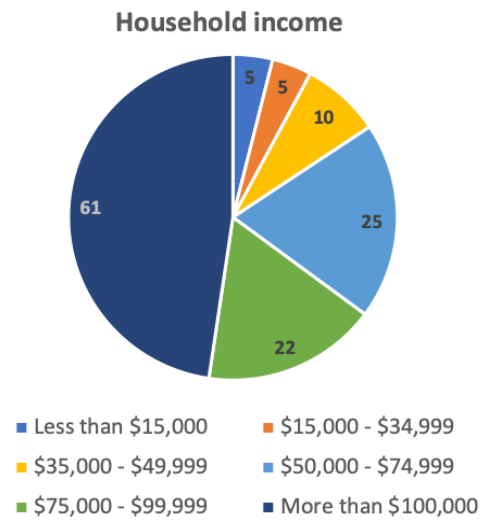
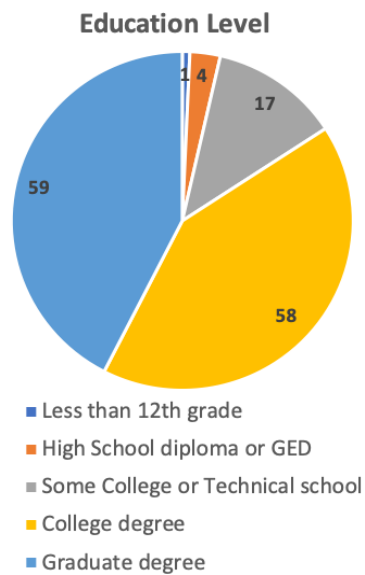
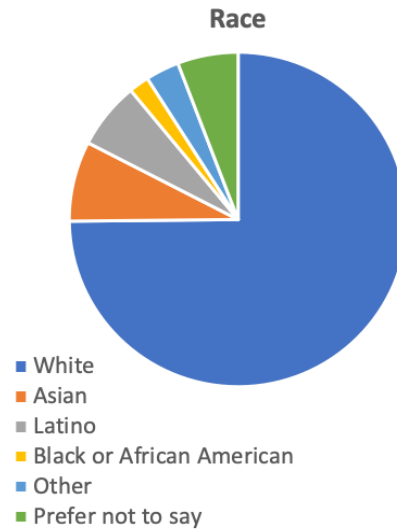
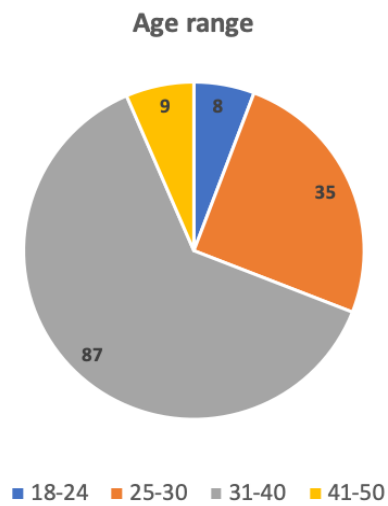
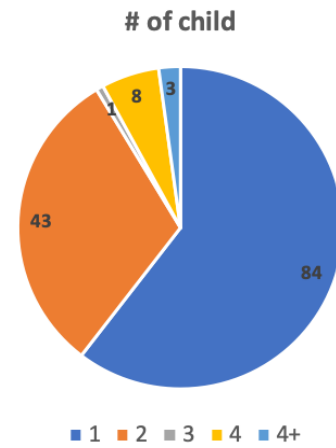
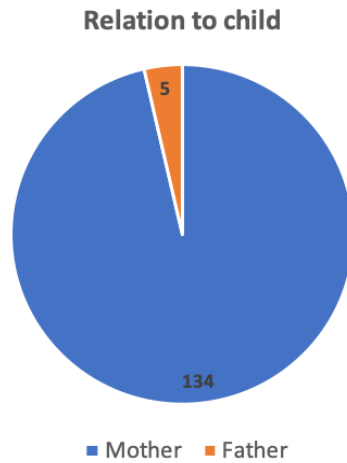


Figure 20. Participant demographics

participant started their participation, text messaging was introduced between 6 and 12 months into the study (see Figure 20). When text messaging became available, I sent out an email to introduce it to participants and provided an instruction sheet (Appendix C). Despite this change from the original plan, it enabled me to collect within-subjects data points from an experimental group and compare specific design characteristics of web portal and text messaging. I will refer to pre-text messaging as Phase 1 and post-text messaging as Phase 2.

At the end of the 20-month completion, I sent a post-study survey to all 139 participants, regardless of their study completion, and received 81 responses (control group 24, experimental group: 57) Participants received \$20 each on 4, 8, 12, 16 and 20 months completion and \$10 on each survey completion. In addition to this, participants who completed the entire 20-month participation received \$20 completion bonus.

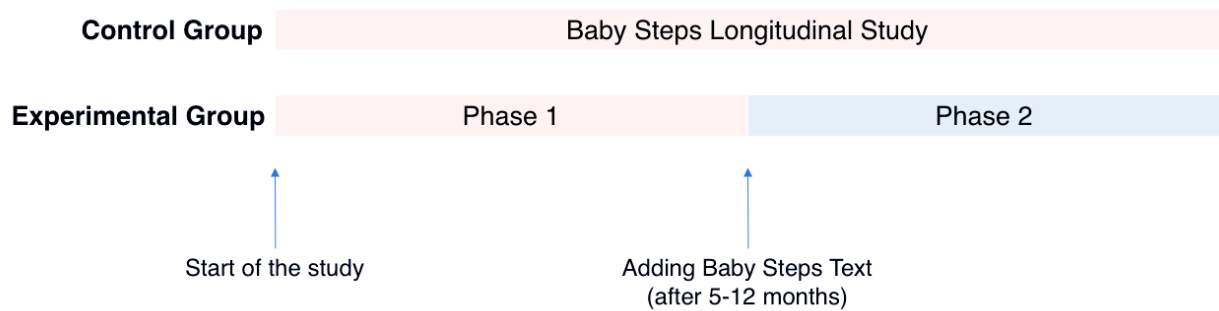


Figure 21. Timeline of Baby Steps longitudinal study by group

7.3 RESULTS

In this section, I report findings from the 20-month deployment study. I first report on between-group comparisons. As the goal of this investigation was to evaluate **T3**, my analysis is centered around them.

7.3.1 Overall Activity – Between Group Comparison

General Usage (Control Group vs. Experimental Group)

From the usage log, I was able to run descriptive statistics and compare general usage pattern between groups. Overall, the experimental group responded much more milestone questions, a total of 15619

responses (min: 4, Max: 449, average: 216.93, SD: 102.42, Median: 205) compared to the control group's 10075 responses (min: 6, Max: 509, average: 150.37, SD: 130.95, median: 90), resulting in 44% more responses per participant on average. What is more important to note is the number of questions answered within a valid time period. Each questionnaire is targeted for a specific age range, and they have a time period where the answer is valid. As a child grows up, they grow out of the valid questionnaire range and they will have different sets of milestones to achieve, i.e., the next questionnaire. Thus, responses to the 12-month questionnaire are only valid if they were answered while the child is between 11 months 0 days through 12 months 30 days old, for example. And it turned out the experimental group answered to more questions within valid time period than the control group did (72% of control group's responses vs. 83% of experimental group's responses) This may have resulted from the experimental group interacting with Baby Steps more frequently than the control group (4.7 times per participant for the control group vs. 8.4 times per participant for the experimental group). In addition, while no control group participant who updated their answers to questions (e.g. 'not yet' to 'sometimes' or 'yes') when their child were able to perform the milestone, twenty experimental group participants (28%) updated their answers to questions and 95% of those updates were made within the valid time range. The experimental group checked the web progress reports 1.4 times more than the control group did (6.6 times per participants for the control group vs. 9 times per participants in the experimental group.) Lastly, there was a big difference in study completion rates between the two groups. While only 38% of the control group finished the entire 20-month journey, 88% of the experimental group finished the study till the end. Figure 3 illustrates milestone response patterns for participants on each group over 20 months.

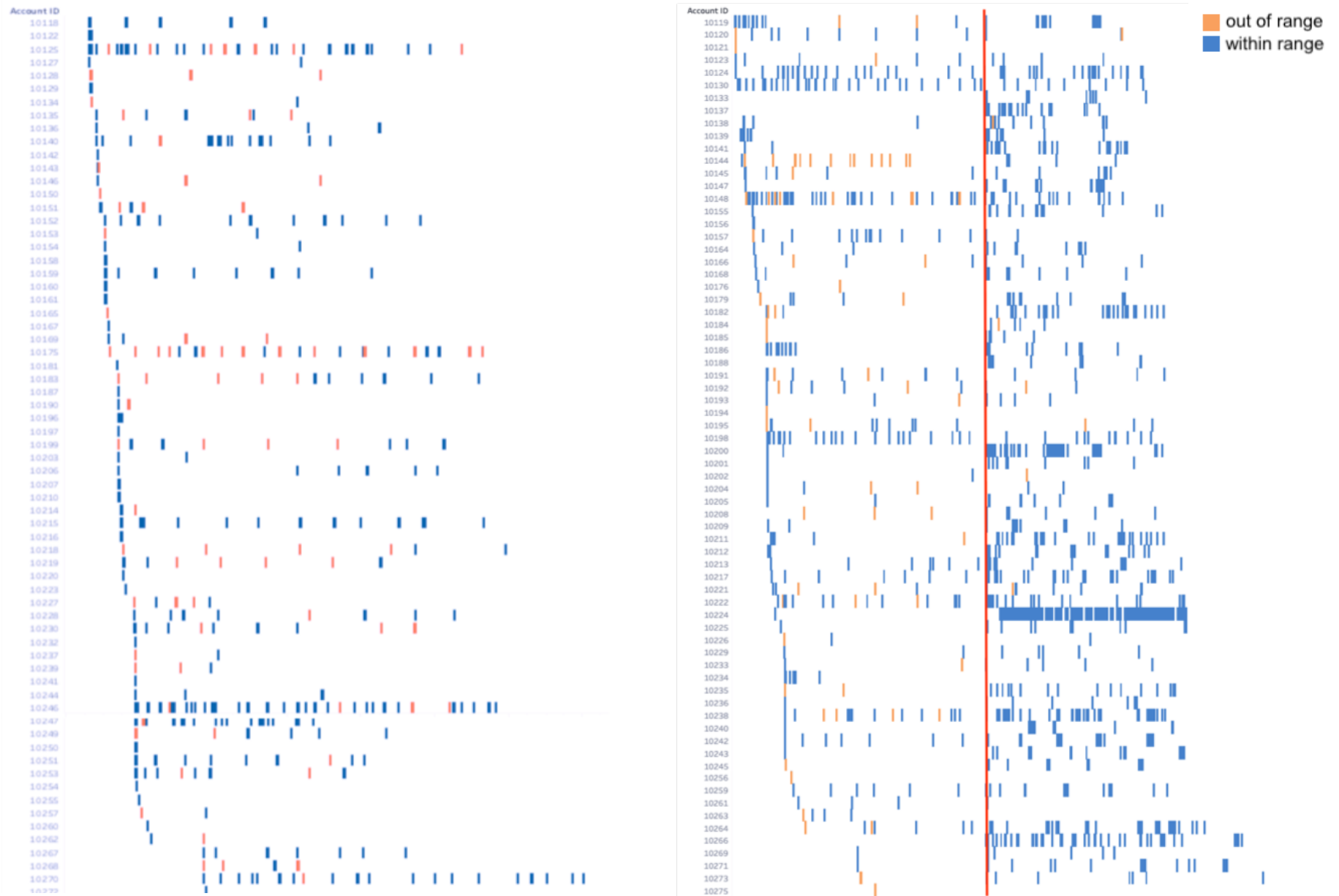


Figure 22. Milestone response patterns of participants in each group over 20 months. Each user is a row and time is a column. Milestone questions answered out of the valid range are highlighted in orange and the red line on the experimental group indicates when the text messaging was introduced.

Follow Up with Progress Report

One of the goals of this longitudinal study was to find out whether Baby Steps is effective in nudging parents to follow up with healthcare providers when progress reports shows signs of possible developmental delay. To investigate this, I partnered with WithinReach Washington, a local non-profit organization that provides a free statewide program for family health and serves many low-income families to provide a toll-free hotline number, which participants could contact if they had questions or concerns about child development. However, the result showed that WithinReach was not well-utilized among participants. With the help of WithinReach, I was able to find that a total of 6 participants (3 from the control group and 3 from the experimental group) who had contacted WithinReach. But their contact happened prior to the study, indicating there were no participants who contacted WithinReach for their concerns. Through survey data, I was able to find out that several participants in the experimental group contacted their healthcare providers based on their concerns about progress reports provided by Baby Steps.

"Used this data to discuss delays with pediatrician, got referrals to early intervention services for therapy and other developmental delays and child will be starting developmental preschool next month with an IEP." (E20)

"On one occasion, he was delayed in his development (based on the progress report) and at his doctor's well check appointment I brought it up and after some tests, it was discovered that my child had iron deficiency. He was prescribed a supplement and right after, my son's condition improved, and he eventually caught up to his development milestone. I would not have discovered or noticed the iron deficiency if not for the app" (E34)

Because there are a very small number of cases, it is too early to conclude that the differences in progress reports (tree visualization with details vs. plain text) made differences in the number of follow-ups, but it is promising to know the progress report for the experimental group was successful in triggering follow up actions from several parents.

Control Group's Request for Experimental Group Features

There were requests for features from the control group to improve their experience, which was already available to the experimental group, demonstrating the effectiveness of the Baby Steps design. For example, the control group was frustrated that they had never received reminders to complete the questionnaire.

"Busy parents need more reminders!!" (C41)

"I never remembered to use it or when the next milestone was. I need reminders or an app to use, rather than I website I had to go to on my own." (C58)

Feedback from participant C58 shows that participants think Baby Steps delivering timely information on developmental tracking would lower parents' mental burden, which Baby Steps Text is already capable of. Also, there were some participants who wished to have a way to keeping memories of milestone achievements.

"(I want a) place to save photos of them [a child] attempting the tasks" (C26)

"It would be nice to have a record of when my child was able to do some of the tasks at each milestone" (C3)

Overall, there are two very promising points that I can summarize here. i) Because regular and timely screening of child development is the key to early detection of many types of developmental disabilities, it is encouraging to see the experimental group interacted with Baby Steps more frequently, answered more milestone questions while they were in the valid range, and checked progress reports more often than the control group did. ii) As early child developmental tracking is not a one-time check but a long process that requires continuous monitoring over 5 years, it is promising that a large number of experimental group participants were able to complete the 20-month long study.

7.3.2 Overall Activity – Experimental Group

Phase 1 vs Phase 2

Depending on participants' date of enrollment, participants in the experimental group were introduced to another way of interacting with Baby Steps via text messaging after 6-12 months in the study. Out of 72 experimental group participants, 52 participants registered their mobile phone numbers in the system and received milestone questions via text messages. Because the text messaging option was introduced during the study, I was able to collect within-subject data points for the experimental group (Phase 1 vs. Phase 2). This comparison helped me compare the specific design characteristics of web portal and text messaging.

Because text messaging reduces access burden and delivers timely information to participants, I hypothesized it would increase the response rate and responses within valid time range. In accordance with my hypothesis, the introduction of a text messaging option increased the number of responses significantly. The number of responses before text messaging was added to the system was 5691 (min: 2, Max: 240, average 84.94, SD: 64.95, median: 60) but after text messaging was introduced, it increased by 34% to 7639 responses (min: 3, Max: 300, average: 119.36, SD: 53.30, median: 120) during a similar time period. What is more surprising were the 5 participants who answered fewer than 5 milestone questions via the website over the first 10 months of participation answered over 200 milestone questions respectively after text messaging was added in (over less than 10 months.)

In addition, more milestone question responses were answered within the valid range in Phase 2 (95%, 7279 out 7639) compared to Phase 1 (77%, 5691 out of 7980). And since the introduction of text messaging, 77% of all messages (5857 of 7639) were answered via text, indicating participants' strong preference (Figure 22).

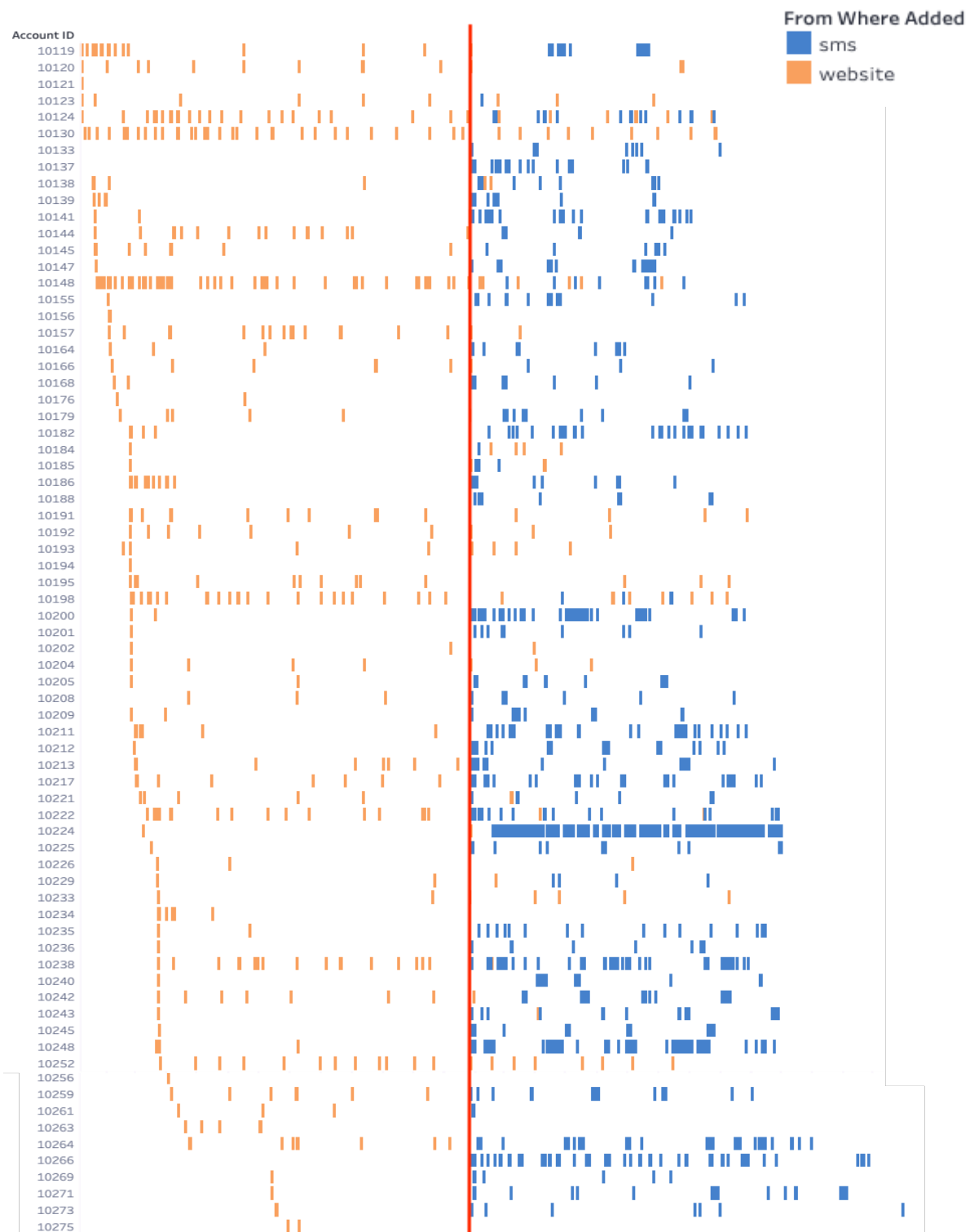


Figure 23. Usage of web portal and text messaging by participants over time. Each user is a row and time is a column. Introduction of text messaging is highlighted in the red line while milestone questions answered via web portal are in orange and text messaging are in blue.

7.3.3 Features Exclusive to Experimental Group

Text messaging

Many participants shared their satisfaction with the addition of the text messaging option as i) it delivers milestone questions to their everyday technology (increasing *ease of use* benefit), ii) enables them to answer on the go and saves time for them (increasing *time* benefit) and iii) serves as a reminder for them (reducing *mental* burden).

"LOVED the text portion. It's so easy to get sidetracked with kids and constantly forget. I needed text prompts and it made it easy to participate." (E23)

"I mainly used the text out of convenience - it appeared on my phone and I reacted vs. remembering or choosing to go to web portal" (E68)

"The text was far easier for me because without reminders I forgot to use the web portal. Being able to complete the questions via text and get development feedback immediately wherever I was very convenient. (...) sometimes I would get busy during the survey and the reminder would prompt me to finish later " (E41)

"Text was much, much better, so easy to use and good for on the go. Gave me fun ideas to try in the moment as well." (E45)

Also, the post-study survey showed that 78% of participants who had responded to milestone questions via text messaging at least once liked the current practice of receiving the next milestone questions immediately after they respond to a milestone question, rather than periodic messages. However, despite its popularity and strength in timely screening, the text messaging option still had room for improvement. While participants appreciated that text messages reminded them to answer the developmental questionnaire, they were not satisfied with the reminder message systems sent after 48-hour of no response. Many expressed it was more often than they'd like and they needed a snooze feature or an option to set a preferred time to be reminded.

"I did not like the reminder text messages; I would have wanted to set the date and time when I would receive questions; weekends would have been easier for me to complete questions" (E15)

"It might be nice to be able to put a snooze on the program and have it come back in a couple of hours or to say what time of day you would prefer to get messages" (E12)

But at the same time, some acknowledged that the annoying, yet constant reminder helped them get through all questions.

"This feature was okay. I would typically sit down and do all of the questions at once because I found it annoying to be repeatedly texted." (E12)

"It annoyed me sometimes, but in the end, I completed them all" (E57)

Some also acknowledged that text was easier but not as visual (e.g. missing visualizations in progress reports). One possible way to solve this issue could be including a link to the web portal progress report to supplement the text nature of the medium, providing a full experience for those who wish to have them.

Web Portal - Progress Report with Tree Visualization

The experimental group appreciated some web portal features that were only available to them. For example, when asked about their top favorite things about the web portal, the tree visualization for the progress report was one of the most popular things. Participants thought it was "cute" and "adorable" and matched well with developmental growth.

"I like the tree pictures as depictions of progress" (E42)

Also, parents appreciated that the progress report informed them about possible developmental delay and how to act on it.

"My child was behind schedule when I didn't expect it, but I appreciated the heads up." (E6)

"the app was actually quite useful for a first-time parent like me. It helped monitor my child's development and alerted me to some delays he experienced (with appropriate guide on what to do about it). I'm glad I participated and thankful for the app." (E7)

Web Portal – Age-Appropriate Activity Suggestions

Another popular feature was the activities tab, which has six or more age-appropriate activities per development category (e.g., Communication, Fine Motor, Gross Motor, Problem Solving, Personal & Social). Parents can try them with their children to encourage developments in specific areas. One example of fine motor activity for 16 to 20 months old is “After washing your hands, show Chloe how to tear lettuce and greens into a bowl. Tell your family who made the salad.” Participants appreciated that it gave them ideas of things they could do with their children in general as well as a first actionable solution when their child’s developmental progress is not on track.

“I loved the milestones and activities suggestions. I got many ideas of things to try out with my little man from your activities.” (E17)

“Because he was just slightly behind and there were tips, activities and resources to check out, I focused on those things first before going to a pediatrician. It ended up working itself out before our next appointment anyway, but it was good to keep in mind and start working on.”

Web Portal – Timeline for Memory Keeping

One of the biggest differences between the experimental group and the control group was sentimental memory keeping. When designing the system, I expected adding sentimental memory-keeping to developmental tracking would relieve possible emotional burdens on parental anxiety and strengthen positivity of the system by having a way to celebrate achievements of the child (e.g. first steps, first visit to zoo, etc.). However, unlike my expectation, memory-keeping was not utilized well from participants. Only 44 out of the 72 (61%) experimental group parents logged memories at least once and 23 of them (52%) used it 3 times or less. From the post-study survey, I was able to determine it was because participants want a way to keep their memories in a somewhat permanent way, if they can make an effort and put in time to keep them, and Baby Steps did not provide a way to support that.

"I would want to be able to print the pages out into a baby book. I would happily pay money to have a service complete an electronic baby book that I could then have as a hard copy. I stopped using baby steps to write down Milestones because I realize that I wouldn't use the portal in the long run and if I was going to take the time to write down Milestones I would rather do them in a physical Baby Book" (E12)

"I was not clear on how the information was going to be retained. At first, I was using the site every few days to track new developments, but then I realized that I probably wasn't going to be able to keep the data. There was no option for printing it or saving it, so I ended up transitioning to a paper baby book instead." (E27)

This was a result because I did not have a chance to evaluate memory keeping from previous studies and failed to gauge parents' expectations. Adding a feature that enables printing a physical photo book or export data from the timeline of memories logged in Baby Steps might be able to solve this problem.

7.3.4 Findings from Surveys and Interviews

To supplement this understanding and delve into participant's experience of use, I surveyed and interviewed participants. In addition to some quotes I used in previous sections, here I report on the findings and themes that emerged from the qualitative data.

Raising Awareness and Empowering Parents

As the core of the study was about answering developmental milestone questions, parents acknowledged that the use of the Baby Steps system informed them about child development and raised awareness around it, which lead them to think about it more regularly in their daily lives.

"milestone surveys were helpful - enjoyed learning about milestones in that manner - nice to know about activity ideas to help develop skills based on age" (E16)

"Made me think about development more intentionally on a regular basis"(E39)

"I loved being able to track milestones and recognize delays or potential problems early on and more often than the doctor's checkups. At the office you are distracted by everything else going on. With Baby Steps, you are able to properly think about and track developmental milestones and progression." (E10)

"Keeping up with children's development progress is useful and helps us keep track on how they grow. The continuation of this program would help us learn more and increase the bonds between parent and child. I learn a lot from this program and I hope it will continue to improve and grow."(E15)

Some parents also shared their stories of how Baby Steps facilitated their conversation with health professionals and advocated their child when needed.

"(...) the questions helped guide me in asking our family doctor the right questions, and the culture of the system/template you created equipped me with the understanding that I wasn't alone in the problems and challenges faced by me as a father/parent." (C10)

"I have twins and one twin was delayed in many areas compared to his brother. This helped us to be proactive about when talking with his pediatrician" (E10)

"I didn't mention the study by name, but the study raised awareness for me about my child's development and things needed to look out for. For example, I advocated strongly for her to get speech therapy. Her pediatrician wanted to wait before the referral, but after no progress for two months, I called again for a referral." (E6)

Previous studies have shown that families with a high level empowerment are more resilient to stressors and have lower feelings of stress, depression, and helplessness than families who are not empowered [89]. Also, other studies have shown a correlation between parent empowerment and parent advocacy, an internal locus of control, and self-efficacy [85, 119]. Hence, it is an encouraging result to see how the use of Baby Steps is empowering parents and has increased their advocacy.

Parental Anxiety

As mentioned earlier, tracking and screening child development can impose parental anxiety and scare them away in the fear of a possibility of learning about their child's developmental delay. To minimize this emotional burden, I aimed to carefully convey progress by working on the exact wording with experts from a local non-profit organization. As a result, participants who had children meeting developmental progress on time shared that Baby Steps was giving them peace of mind in the process of answering developmental questionnaires and checking progress reports.

"The questionnaires lowered my stress about milestones as I saw my child was developmentally appropriate"(E62)

"Quelled any fears about development"(E39)

"It helped me feel like less of a failure as a parent."(C33)

However, when the child's development is not on track, discomfort and parental anxiety was inevitable. Several participants shared that things were even worse because they were going through postpartum depression themselves and finding out about the possibility of child development was not helping.

"Sometimes I worry when I answer NO to a developmental question. I worry that my child is behind" (E56)

"I found it disheartening and uncomfortable when my child hadn't reached certain milestones. My child was a slow crawler and walker and I was concerned about it, so getting reinforcements that he was behind made me uncomfortable I don't think I understood the purpose of the app clearly when I signed up. I had thought it would be more of an online baby book and set up something to track progress. I have had postpartum depression so doing something that told me something might be wrong with my son didn't fit into my health plan" (E27)

"It would be nice to get more reassurance from the progress report. I had very bad postpartum anxiety, and I became anxious taking the test because I felt like it was telling me that something was wrong with my child" (E13)

The design and wording of Baby Steps needs to be refined further, especially for parents who receive 'not on track' or 'needs monitoring' progress reports, because the underlying goal of Baby Steps is to create awareness that leads to long-lasting change among parents, not to further discomfort or guilt among well-meaning parents.

7.4 DISCUSSION

As a result of my overall design process and 20-month longitudinal study, I have uncovered a number of insights for improving the usability and long-term engagement of Baby Steps.

7.4.1 Recommended Changes to Baby Steps

Expand child development monitoring to capture a bigger picture

Participants expressed their desire to track other aspects of child development, such as weights and heights, reminders on immunization, well-child visits through Baby Steps to have complete record.

"Couldn't track things I easily wanted to (like immunizations, weights, etc.)" (E19)

"Option to upload a picture of child for every month/development milestones and ability of the app to compile these and present in a scrolling, picture progress report" (E34)

And memory keeping in Baby Steps can also be more proactive and prompt parents to upload pictures to create picture progress reports of growth alongside of a developmental progress report.

A Need for Baby Steps Mobile App

When asked to choose other formats of Baby Steps participants might be interested in using in the mid-study survey, 'dedicated app' was the most popular choice (79%, 93 out of 118) followed

by text messaging (38%, 45 out of 118). It was also confirmed when many participants named a Baby Steps mobile app as a feature they would like to add in future versions of Baby Steps.

“I actually think that the system would work really well as an app. Notifications would remind me to add milestones and pictures. Adding pictures etc. would be easy since they’re all on my phone anyway. The survey could be taken via the app and questions sent via notifications if I forget to answer.” (E41)

“Make it into an app, be able to text pictures and timeline additions to the app, have it send reminders like when I should schedule a yearly well child visit.” (E13)

Also, there were many participants who referred to the Baby Steps web portal as an ‘app’ in surveys. Though web portal works on a mobile browser, it was not specifically designed to accommodate it and some participants complained that the ‘app’ (web portal) was clunky and needed improvement. According to the Pew Research Center [147], ownership of smartphones was not high enough and only 43% of adults in the United States owned one, compared to the 89% of cellphone ownership back in 2012 when I started this project. I did not choose to develop a smartphone application and focused on more universally accepted technologies because the goal of the project was making child development monitoring accessible to all, regardless of their backgrounds, and a smartphone application could marginalize people who cannot not afford an expensive smartphone. However, the smartphone ownership has grown rapidly in the U.S., up to 81%, and for people in the age between 18 and 35, it has gone up to 95%. As smartphones are already equipped with functionalities that parents want and appreciate from Baby Steps (e.g., reminder notifications, easily enables adding memory from camera/ gallery, fully supports visual presentations of progress reports, etc.) there exists a possibility that a Baby Steps app might yield many benefits to many people.

Provide More Details on the Process and Empower parents

Parents also expressed their desire to learn more details across the process and wanted to learn more than just answering the questions provided. Some wanted specifics details of milestone

questions and their importance, while others wanted more detailed explanations of activities or progress reports.

“Videos of child showing a skill (examples of ASQ questions)” (E18, E62)

“I would like to know why some of the tasks at each milestone were important or what they were actually reflecting.”

“Additional information- when my child was at the cusp of reaching a milestone, I would get very mild anxiety about why she had not achieved it completely... So, adding additional screens or information to just ease the potential worry” (C54)

Although each question and activity are categorized into five developmental categories, parents wanted to know more than ‘this question is related to communication skills,’ or ‘this activity will encourage development of gross motor skills.’ Therefore, preparing additional information and explanation would be helpful in educating parents about child development in depth.

For example, for a 6-month communication milestone question “when you call Chris when you are out of sight, does he look in the direction of your voice?”, it can include a brief explanation such as “turning his head in the direction of sound shows sign of Chris’s hearing ability. As hearing loss may lead to communication development delay, it’s still important to watch for signs that he’s hearing properly.”

Also, in addition to the information currently on the progress report (e.g., link to activities to encourage development, hotline number to contact for questions or concerns,) it can also provide more guidance on how to communicate with doctors. There were several participants who used progress reports to communicate their concerns with healthcare providers and advocated for their children (see section Parental Anxiety under 7.3.4 Findings from Survey and Interview for more stories).

"Used the milestone progress report to concisely describe current development and any areas of concern. Did not print and bring report in with me. However, the report helps you to step back and see your child's whole development rather than just fixate on only 1 or 2 concerns." (E9)

But often times, healthcare providers were not interested in results of developmental tracking parents did outside of the clinic, and it discouraged them from bringing up the issue again.

"We let his pediatrician know he was participating and brought a progress report to a visit. The Dr. didn't seem too interested in this though, so we haven't continued to bring it to visits (though I do try to complete the surveys and review the progress reports)" (E5)

Therefore, the next version of Baby Steps can include a guidance on how to communicate with healthcare providers, including how to introduce the program and how to initiate a conversation around signs they got from the Baby Steps report.

Better Connection with Healthcare Providers

There were several cases where participants completed the Ages and Stages Questionnaire, a questionnaire Baby Steps that was adopted from, in several other places such as pediatric clinics or schools. They felt it was unnecessary redundancy and wanted a way to share their answers with other entities, when needed. I do not envision Baby Steps replacing developmental screenings happening at doctors' offices, but rather complementing them by providing a way to monitor outside of the clinic for parents. Therefore, it is recommended to integrate the developmental questionnaire part of Baby Steps with medical systems so parents can share data directly with the doctor, and doctors can utilize what parents have been tracking outside of the clinic.

7.5 CHAPTER 7 SUMMARY

In this chapter, I presented the redesign of Baby Steps, based on my understandings of user burdens and benefits in using health monitoring systems, which I acquired through explorations from previous chapters. I then reported the results of a 20-month longitudinal study of Baby Steps

with 139 families. This work provided empirical data to evaluate **T3** by examining the question of “how can we design technology to support child development tracking in the long term?” The result showed the importance of enhancing benefits and reducing burdens in technology and their potential effect on technology adoption and sustained use. This work makes two contributions: 1) as long-term studies over 6 months are not common in the field of HCI, the result of the 20-month deployment study with 139 families makes an empirical research contribution to the HCI community, 2) the design of Baby Steps contributes specific design artifacts by showing how health tracking systems can be designed to keep motivation and sustain use.

Chapter 8. Discussion & Conclusion

Across the studies presented in this dissertation, I investigated how to design technologies to support regular health monitoring and ways to lower perceived user burdens and benefits for long-term engagement. Beginning with two empirical studies to investigate how technologies people already use can be used for regular health monitoring, I have developed two systematic ways to understand and measure user burden and benefits associated with technology use. Based on insights gained from these foundational studies, I conducted a 20-month longitudinal study to study long-term engagement. In this chapter, I summarize lessons learned by reflecting on the

theoretical work and deployment studies presented in previous chapters; discuss contributions I offer in this dissertation and areas of future exploration; and then close with concluding remarks.

8.1 LESSONS LEARNED

8.1.1 Active Health Monitoring and Patient Empowerment

Through three deployment studies in Chapters 3, 4 and 7, I was able to find that active health monitoring facilitated by the Baby Steps system showed potential for increased parental awareness around child development, which can lead to patient empowerment. Because passive tracking removes the burden of data entry, requiring almost no additional action from the user, it is widely accepted in the areas of health and wellness monitoring (e.g., step counters, sleep trackers, heart rate monitors, etc.). While automated, passive health monitoring through wearable devices and/or apps can reduce burdens of gathering health data, there still are many types of health monitoring, such as disease management for diabetes, asthma or developmental screenings, that need direct user input or action to capture the whole picture more accurately. Also, while automated health monitoring technologies may provide an easy method to collect data toward user's longer-term health goals, there are limitations on the effectiveness of such technologies for modifying health behaviors in the long term, unless collected data is curated and used in a meaningful way. A recent study suggests that simply tracking and gathering data does not benefit its users unless they get feedback from it, analyze it, receive some type of intervention, and observe how tracked data changes [71].

To maximize the benefits of active health monitoring that requires user input, the Baby Steps system utilized several prompts parents could use in addition to milestone questions, such as checking activity recommendations, progress reports, and providing possible follow up actions they could take. Parents reported that these active monitoring prompts from Baby Steps improved their awareness and engagement around child development, since it allowed parents to play a more active role in understanding child development and communicating it with healthcare providers using the knowledge they acquired in the process. And the studies

consistently showed that the benefits of regular monitoring were not limited to children with developmental and behavioral delays. Even for typically developing children, families benefited by increased awareness of appropriate developmental and behavioral expectations. This suggests that in some areas of health monitoring, such as screening, it might be beneficial to require some level of active monitoring from users, instead of fully automated passive monitoring, as it can give users an opportunity to play an active role in achieving their desired health goal.

8.1.2 Burdens and Benefits in Health Monitoring

In addition to building new knowledge about how health monitoring technologies can positively affect its users, this dissertation explored the importance of enhancing benefits and reducing burdens in technology and their potential effect on technology adoption and sustained use. I presented an exemplary work of applying my foundational understanding of user burden and benefit to iterative work of the Baby Steps system design. Although the User Burden Scale and User Benefit Scale were not specifically developed for the technologies in the domain of health, the overall concept and definition of each construct are to be applicable to a wide range of technologies, including health monitoring technologies. Therefore, even when some questionnaire items might not be specifically relevant, the general concept and each construct within UBS and UBeS can be used as guidelines for designers who wish to review their design in regard to burdens and benefits the technology imposes and provides.

Also, one burden type that did not become most apparent in Baby Steps Text and Baby Steps Longitudinal studies was a privacy burden. No participant raised concerns around privacy and the burdens of managing it. I believe this is because I tried to be transparent about the data management plan to participants – through the consent process and web portal, I communicated how Baby Steps managed data, including data storage, the scope of people with data access, and the timeline for data destruction. However, people who buy health monitoring technologies these days often do not have full capacity to control over their data nor a good understanding of how their tracked data is managed, which might have been “hidden” in the terms of service. Therefore,

more effort is needed to transparently communicate the privacy and security of personal data generated through health monitoring to lower this burden on its users.

Finally, the study showed that when it comes to designing for benefits of technology use, merely adding features to the design with good intention might not be sufficient. The studies showed that despite my effort to find ways to enrich the benefits of the technology in each applicable construct of user benefit from UBeS, many participants were not aware of the features that might have been beneficial to use. This result suggests that it is necessary to promote relevant features to the users for their actual utilization of them, which can be accomplished through onboarding and nudging reminders during use.

8.2 CONTRIBUTIONS

The first contribution of this work is an empirical, generative understanding on how to support families in tracking their child's development. Across the studies, I presented parents' perspectives about seamlessly blending child development monitoring into the technologies they already use, their patterns of use, as well as difficulties and concerns they had. By connecting results from both quantitative and qualitative analysis, I documented design opportunities for supporting long-term health monitoring by reducing burdens and enhancing benefits of use.

The second contribution of this work is the creation of design artifacts such as @BabySteps, Baby Steps Text and Baby Steps Web Portal. By integrating health interventions into everyday technologies people commonly use, my work contributed new knowledge about how parent-generated health data can be collected and used in a simple but thoughtful way, expanding the reach of its benefits.

Thirdly, my work makes a methodological contribution by conducting a 20-month long deployment. In the field of HCI, it is rare to find long-term studies that run over a month. Thus, the approaches and strategies I used in my longitudinal study can provide guidance to researchers who seek to execute long-term studies.

Finally, my work makes theoretical contributions and methodological contributions. I have identified a novel concept in constructs of user burden and benefits and laid the groundwork for new models of these two constructs. Building on these theoretical contributions, I also created User Burden Scale (UBS) and User Benefit Scale (UBeS), making methodological contributions. By creating two quick and easy measurements for assessing user burden and benefit in technology, I provided useful methods that could yield reliable, valid results to help HCI researchers and designers. I hope this will help understand the ways that technologies may have an impact on the users' lives beyond usability and enjoyment.

8.3 FUTURE WORK

In this dissertation, I addressed burdens and benefits of technology as well as challenges and opportunities of designing health monitoring technologies. While doing so, I opened up many areas of research that need to be examined further.

First, Baby Steps showed a way to engage parents in developmental tracking and how it can raise awareness around child development and empower them. The next big step would be to understand how to streamline connections between families and healthcare providers of their preference seamlessly, as well as finding ways to make more proactive connections between families and healthcare providers. Also, despite my efforts, I came short on involving more fathers in the process. Therefore, more work is needed to support involvement of different caregivers other than the mother of a child in the process, such as fathers and daycare workers.

There are also other aspects that can benefit from further research. Because developmental monitoring is a unique health monitoring area – unlike many other types of health self-tracking, in child development monitoring caregivers do it on behalf of their child. And the common problems self-trackers experience with collected data (e.g., what to do with tracked data, how to make use of it, etc.) become more complicated here. Thus, there are future research opportunities to investigate how to keep and handle tracked data – both developmental and sentimental - in a more meaningful way. For example, the uncertainty of how the sentimental record was going to

be managed was one of the biggest reasons participants were reluctant to use the feature. Studying how to mitigate and communicate concerns around data sustainability as well as examining design opportunities for data handover from parents to their children as they grow can be a good next step. Additionally, future research should explore how public health systems like Baby Steps can accommodate families with limited English proficiency.

Finally, my work introduced the perceived burdens and benefits people might feel in their technology use and how to understand them. Using this as a groundwork, future work can study how to adopt UBS and UBeS specifically for health technologies. Also, making sense of how UBS and UBeS are correlated, and possibly developing a short version of the combined survey to quickly measure and analyze, could be valuable for many designers and researchers. In addition, while current UBS and UBeS focuses on burden and benefit on a personal level, there are additional research opportunities to study burdens and benefits of technologies on collaborative use. Finally, investigating differences in perceived burdens and benefits over time or over different contexts (e.g., personal use vs. technology required for work) can give good guidance on how user burden and benefit changes and how to adapt the design accordingly.

8.4 CONCLUSION

Health monitoring to manage one's health or that of a loved one has been practiced for several decades using methods such as paper and pencil, or just memory [170]. The practice of regular health monitoring and the knowledge that the data tracks has the potential to bring awareness and support health behaviors, thus helping to improve quality of life [71]. However, despite these foreseeable benefits of use, people fail to adopt health monitoring practices, not to mention sustaining the practices in the long-term, because of perceived user burdens, such as time, effort, and lacking information and knowledge needed.

In this dissertation, by integrating and building on theories and methods from HCI, design, and child development, I presented how everyday technologies people already use can be used to support long-term health monitoring. I investigated the opportunities and challenges

of designing health monitoring technologies and documented how we can understand user burdens and benefits in technologies by developing two quickly administrable measures. This work shows that by reducing user burdens and enhancing user benefits associated with health monitoring, we can engage, inform, educate, and empower users to play a more active role in their health management or that of their loved ones. I hope that this work will shed new light for designers on how health monitoring technologies can be designed to support regular and long-term health monitoring.

References

1. Ahmad Rahmati, Chad Tossell, Clayton Shepard, Philip Kortum, and Lin Zhong. 2012. Exploring iPhone usage: the influence of socioeconomic differences on smartphone adoption, usage and usability. In Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services (Mobile HCI '12), 11-20.
2. Alaina Darby, M.W. Strum, E. Holmes and J. Gatwood. 2016. A review of nutritional tracking mobile applications for diabetes patient use. *Diabetes technology & therapeutics*. 18(3): 200-212.
3. Alexander J. Rothman and Peter Salovey. "Shaping perceptions to motivate healthy behavior: the role of message framing." *Psychological bulletin* 121, no. 1 (1997): 3.
4. Amanda Lazar, Christian Koehler, Joshua Tanenbaum, and David H. Nguyen. "Why we use and abandon smart devices." In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 635-646. ACM, 2015.
5. Anita Khokhar. 2009. Short text messages (SMS) as a reminder system for making working women from Delhi Breast Aware. *Asian Pac J Cancer Prev*. 10, 2: 319-22.
6. Anna M. Wichansky. 2000. Usability testing in 2000 and beyond. *Ergonomics* 43, 7: 998-1006.
7. Anna Vilella, J.M. Bayas, M.T. Diaz, C. Guinovart, C. Diez, D. Simó, A. Muñoz and J. Cerezo. 2004. The role of mobile phones in improving vaccination rates in travelers. *Preventive medicine*, 38(4): 503-509.
8. Anna Vilella, Jose-Maria Bayas, Maria-Teresa Diaz, Caterina Guinovart, Consolación Diez, Dulcis Simó, Amparo Muñoz, and Javier Cerezo. 2004. The role of mobile phones in improving vaccination rates in travelers. *Preventive medicine* 38, 4: 503-509.
9. Anthony Rodgers, Tim Corbett, Dale Bramley, Tania Riddell, Mary Wills, Ruey-Bin Lin, and Mark Jones. 2005. Do u smoke after txt? Results of a randomised trial of smoking cessation using mobile phone text messaging. *Tobacco control* 14, 4: 255-261.
10. Batya Friedman. 1996. Value-sensitive design. *interactions*, 3(6): 16-23.
11. Ben S. Gerber, Melinda R. Stolley, Allison L. Thompson, Lisa K. Sharp, and Marian L. Fitzgibbon. 2009. Mobile phone text messaging to promote healthy behaviors and weight loss maintenance: a feasibility study. *Health informatics journal* 15, 1: 17-25.

12. Ben Shneiderman. 1992. Designing the user interface: strategies for effective human-computer interaction. Vol. 3. Reading, MA: Addison-Wesley.
13. Ben Shneiderman. 2004. Designing for fun: how can we design user interfaces to be more fun?. *interactions*, 11(5): 48-50.
14. Ben-Sasson, A., Ben-Sasson, E., Jacobs, K., & Saig, E. (2017, May). Baby CROINC: an online, crowd-based, expert-curated system for monitoring child development. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare* (pp. 110-119). ACM.
15. Beom Suk Jin and Yong Gu Ji. 2010. Usability risk level evaluation for physical user interface of mobile phone. *Computers in Industry* 61, 4: 350-363.
16. Boyle, C. A., Boulet, S., Schieve, L. A., Cohen, R. A., Blumberg, S. J., Yeargin-Allsopp, M., ... & Kogan, M. D. (2011). Trends in the prevalence of developmental disabilities in US children, 1997–2008. *Pediatrics*, 127(6), 1034-1042.
17. Brian Shackel. 1984. The concept of usability. *Visual display terminals: Usability issues and health concerns*: 45-87.
18. Brianna S. Fjeldsoe, Yvette D. Miller, and Alison L. Marshall. 2010. MobileMums: a randomized controlled trial of an SMS-based physical activity intervention. *Annals of Behavioral Medicine* 39, 2: 101-111.
19. Bright Futures Steering Committee, & Medical Home Initiatives for Children with Special Needs Project Advisory Committee. (2006). Identifying infants and young children with developmental disorders in the medical home: An algorithm for developmental surveillance and screening. *Pediatrics*, 118(1), 405-420.
20. Caroline Free, Rosemary Knight, Steven Robertson, Robyn Whittaker, Phil Edwards, Weiwei Zhou, Anthony Rodgers, John Cairns, Michael G. Kenward, and Ian Roberts. 2011. Smoking cessation support delivered via mobile phone text messaging (txt2stop): a single-blind, randomised trial. *The Lancet* 378, 9785: 49-55.
21. Catalina M. Danis, Jason B. Ellis, Wendy A. Kellogg, Hajo van Beijma, Bas Hoefman, Steven D. Daniels, and Jan-Willem Loggers. 2010. Mobile phones for health education in the developing world: SMS as a user interface. In *Proceedings of the First ACM Symposium on Computing for Development*, 13.
22. Catherine M. Dang, Sylvia Estrada, Catherine Bresee, and Edward H. Phillips. 2013. Exploring potential use of internet, E-mail, and instant text messaging to promote breast

- health and mammogram use among immigrant Hispanic women in Los Angeles County. *The American Surgeon* 79, 10: 997-1000.
23. CDC – MMWR Surveillance Summary. 2012. 61, 3 (March 2012), 1-19.
 24. Child Health. 2009. http://www.who.int/topics/child_health/en/.
 25. Chris Bourne, V. Knight, R. Guy, H. Wand, H. Lu, and A. McNulty. 2011. Short message service reminder intervention doubles sexually transmitted infection/HIV re-testing rates among men who have sex with men. *Sexually transmitted infections* 87, 3: 229-231.
 26. Coleen A. Boyle, Sheree Boulet, Laura A. Schieve, Robin A. Cohen, Stephen J. Blumberg, Marshalyn Yeargin-Allsopp, Susanna Visser, and Michael D. Kogan. 2011. Trends in the prevalence of developmental disabilities in US children, 1997–2008. *Pediatrics* (2011), 1034-1042.
 27. Colin A. Depp, Brent Mausbach, Eric Granholm, Veronica Cardenas, Dror Ben-Zeev, Thomas L. Patterson, Barry D. Lebowitz, and Dilip V. Jeste. 2010. Mobile interventions for severe mental illness: design and preliminary data from three approaches. *The Journal of nervous and mental disease* 198, 10: 715.
 28. Da Costa, Thiago Martini, Paulo Lísias Salomão, Amilton Souza Martha, Ivan Torres Pisa, and Daniel Sigulem. "The impact of short message service text messages sent as appointment reminders to patients' cell phones at outpatient clinics in Sao Paulo, Brazil." *International journal of medical informatics* 79, no. 1 (2010): 65-70.
 29. Dan Ledger and Daniel McCaffrey. "Inside wearables: How the science of human behavior change offers the secret to long-term engagement." *Endeavour Partners* 200, no. 93 (2014): 1.
 30. Daniel A. Epstein, An Ping, James Fogarty, and Sean A. Munson. "A lived informatics model of personal informatics." In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 731-742. ACM, 2015.
 31. Daniel A. Epstein, Monica Caraway, Chuck Johnston, An Ping, James Fogarty, and Sean A. Munson. "Beyond abandonment to next steps: understanding and designing for life after personal informatics tool use." In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 1109-1113. ACM, 2016.
 32. David De Vaus. 2002. *Analyzing social science data: 50 key problems in data analysis*. Sage.
 33. David Wright. 2011. Should privacy impact assessments be mandatory?. *Communications of the ACM* 54, 8: 121-131.

34. Deb Levine. 2009. Using technology to promote youth sexual health. *Texting 4 Health* 57-65.
35. Deb Sledgianowski & Songpol Kulviwat. "Social network sites: antecedents of user adoption and usage." *Americas Conference on Information Systems (AMCIS) Proceedings*: 83.
36. Diane Bricker, Jane Squires, Linda Mounts, L. Potter, Robert Nickel, Elizabeth Twombly, and Jane Farrell. "Ages and stages questionnaire." Paul H. Brookes: Baltimore. 1999.
37. Dianne Cyr, Milena Head, and Alex Ivanov. "Design aesthetics leading to m-loyalty in mobile commerce." *Information & Management* 43, no. 8 (2006): 950-963.
38. Domenico Campolo, Eliseo Stefano Maini, Francesco Patane, Cecilia Laschi, Paolo Dario, Flavio Keller, and Eugenio Guglielmelli. "Design of a sensorized ball for ecological behavioral analysis of infants." In *Proceedings 2007 IEEE International Conference on Robotics and Automation*, pp. 1529-1534. IEEE, 2007.
39. Donald A. Norman. *Emotional design: Why we love (or hate) everyday things*. Basic Civitas Books, 2004.
40. Donna M. D'alessandro and Nienke P. Dosa. "Empowering children and families with information technology." *Archives of pediatrics & adolescent medicine* 155, no. 10 (2001): 1131-1136.
41. Elizabeth T. Jordan, Ellen M. Ray, Pamela Johnson, and W. Douglas Evans. 2011. *Text4Baby*." *Nursing for women's health* 15, 3: 206-212.
42. Eric Granholm, Dror Ben-Zeev, Peter C. Link, Kristen R. Bradshaw, and Jason L. Holden. 2011. *Mobile Assessment and Treatment for Schizophrenia (MATS): a pilot trial of an interactive text-messaging intervention for medication adherence, socialization, and auditory hallucinations*. *Schizophrenia bulletin*: sbr155.
43. Erik Grönvall & Nervo Verdezoto. "Beyond self-monitoring: understanding non-functional aspects of home-based healthcare technology." In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pp. 587-596. ACM, 2013.
44. Eun Kyoung Choe, Marisa E. Duarte, and Julie A. Kientz. 2010. *Understanding and Designing Computing Technologies that Convey Concerning Health News*. *Design & Emotion*.

45. Eun Kyoung Choe, Marisa E. Duarte, Hyewon Suh, Wanda Pratt, and Julie A. Kientz. "Communicating Bad News: Insights for the Design of Consumer Health Technologies." *JMIR human factors* 6, no. 2 (2019): e8885.
46. F. Cecchi, S. M. Serio, M. Del Maestro, C. Laschi, and P. Dario. "Design and development of sensorized toys for monitoring infants' grasping actions." In *2010 3rd IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics*, pp. 247-252. IEEE, 2010.
47. Fang Chen, Natalie Ruiz, Eric Choi, Julien Epps, M. Asif Khawaja, Ronnie Taib, Bo Yin, and Yang Wang. Multimodal behavior and interaction as indicators of cognitive load. *ACM Transactions on Interactive Intelligent Systems (TiiS '12)* 2, 4: 22.
48. Felicia Cordeiro, Daniel A. Epstein, Edison Thomaz, Elizabeth Bales, Arvind K. Jagannathan, Gregory D. Abowd, and James Fogarty. "Barriers and negative nudges: Exploring challenges in food journaling." In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 1159-1162. ACM, 2015.
49. Fidele Ngabo, Judith Nguimfack, Friday Nwaigwe, Catherine Mugeni, Denis Muhoza, David R. Wilson, John Kalach, Richard Gakuba, Corrine Karema, and Agnes Binagwaho. 2012. Designing and implementing an innovative SMS-based alert system (RapidSMS-MCH) to monitor pregnancy and reduce maternal and child deaths in Rwanda." *The Pan African Medical Journal* 13.
50. Fred D. Davis. 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 319-340.
51. Fred D. Davis. 1993. User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. *International journal of man-machine studies*, 38(3): 475-487.
52. Geraldine Dawson & Julie A. Osterling. "Early intervention in autism." *The effectiveness of early intervention*: 307-326.
53. Gillian R. Hayes, Donald J. Patterson, Mohan Singh, Dana Gravem, Julia Rich, and Dan Cooper. 2011. Supporting the transition from hospital to home for premature infants using integrated mobile computing and sensor support. *Personal Ubiquitous Comput.* 15, 8 (December 2011), 871-885.
54. Glenn Flores, Milagros Abreu, Mary Anne Olivar, and Beth Kastner. 1998. Access barriers to health care for Latino children. *Archives of Pediatrics & Adolescent Medicine* 152, 11: 1119-1125.

55. Godwin J. Udo. 2001. Privacy and security concerns as major barriers for e-commerce: a survey study. *Information Management & Computer Security* 9, 4: 165-174.
56. Gregory S. Liptak, Lauren B. Benzoni, Daniel W. Mruzek, Karen W. Nolan, Melissa A. Thingvoll, Christine M. Wade, and G. Edgar Fryer. "Disparities in diagnosis and access to health services for children with autism: data from the National Survey of Children's Health." *Journal of Developmental & Behavioral Pediatrics* 29, no. 3 (2008): 152-160.
57. Günther Gediga, Kai-Christoph Hamborg, and Ivo Düntsch. 1999. The IsoMetrics usability inventory: an operationalization of ISO 9241-10 supporting summative and formative evaluation of software systems. *Behaviour & Information Technology* 18, 3: 151-164.
58. Haewoon Kwak, Hyunwoo Chun, and Sue Moon. "Fragile online relationship: a first look at unfollow dynamics in twitter." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1091-1100. ACM, 2011.
59. Han X. Lin, Yee-Yin Choong, and Gavriel Salvendy. 1997. A proposed index of usability: a method for comparing the relative usability of different software systems. *Behaviour & information technology* 16, 4-5: 267-277.
60. Harriet Fell, Cynthia Cress, Joel MacAuslan, and Linda Ferrier. "visiBabble for reinforcement of early vocalization." In *ASSETS*, vol. 2004, pp. 161-168. 2004.
61. Harrison, D., Marshall, P., Bianchi-Berthouze, N., & Bird, J. (2015, September). Activity tracking: barriers, workarounds and customisation. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 617-621). ACM.
62. Heather Cole-Lewis, and Trace Kershaw. 2010. Text messaging as a tool for behavior change in disease prevention and management. *Epidemiologic reviews* 32, 1: 56-69.
63. Hee Young Jeong, Sun Young Park, and John Zimmerman. 2008. Opportunities to support parents in managing their children's health. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems (CHI EA '08)*, 225-3230.
64. Hee Yun Lee, Joseph S. Koopmeiners, Taeho Greg Rhee, Victoria H. Raveis, and Jasjit S. Ahluwalia. 2014. Mobile phone text messaging intervention for cervical cancer screening: changes in knowledge and behavior pre-post intervention. *Journal of medical Internet research* 16, 8.
65. Hee Yun Lee, Joseph S. Koopmeiners, Taeho Greg Rhee, Victoria H. Raveis, and Jasjit S. Ahluwalia. 2014. Mobile phone text messaging intervention for cervical cancer screening:

- changes in knowledge and behavior pre-post intervention. *Journal of medical Internet research* 16, 8.
66. Helen Petrie & Nigel Bevan. "The Evaluation of Accessibility, Usability, and User Experience." *The universal access handbook* 1 (2009): 1-16.
 67. Hyewon Suh, John R. Porter, Alexis Hiniker, and Julie A. Kientz. 2014. @ BabySteps: design and evaluation of a system for using twitter for tracking children's developmental milestones. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*, 2279-2288.
 68. Hyewon Suh, John R. Porter, Robert Racadio, Yi-Chen Sung, and Julie A. Kientz. "Baby Steps Text: Feasibility Study of an SMS-Based Tool for Tracking Children's Developmental Progress." In *AMIA Annual Symposium Proceedings*, vol. 2016, p. 1997. American Medical Informatics Association, 2016.
 69. Hyewon Suh, Nina Shahriaree, Eric B. Hekler, and Julie A. Kientz. "Developing and validating the user burden scale: A tool for assessing user burden in computing systems." In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pp. 3988-3999. ACM, 2016.
 70. Ian Janssen and Allana G. LeBlanc. "Systematic review of the health benefits of physical activity and fitness in school-aged children and youth." *International journal of behavioral nutrition and physical activity* 7, no. 1 (2010): 40.
 71. Ian Li, Anind Dey, and Jodi Forlizzi. "A stage-based model of personal informatics systems." *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 2010.
 72. Ilhan Aslan, Martin Murer, Verena Fuchsberger, Andrew Fugard, and Manfred Tscheligi. 2013. Workload on your fingertips: the influence of workload on touch-based drag and drop. In *Proceedings of the 2013 ACM international conference on Interactive tabletops and surfaces (ITS '13)*, 417-420.
 73. J. C. Herz. Wearables are totally failing the people who need them most. *Wired*.
 74. Jacob O. Wobbrock and Julie A. Kientz. "Research contributions in human-computer interaction." *interactions* 23, no. 3 (2016): 38-44.
 75. Jagdish N. Sheth, Bruce I. Newman, and Barbara L. Gross. *Consumption values and market choices: Theory and applications*. 1991.

76. Jakicic JM, Davis KK, Rogers RJ, King WC, Marcus MD, Helsel D, Rickman AD, Wahed AS, Belle SH. Effect of wearable technology combined with a lifestyle intervention on long-term weight loss: the IDEA randomized clinical trial. *Jama*. 2016 Sep 20;316(11):1161-71.
77. Jakob Nielsen. "Usability metrics: Tracking interface improvements." *IEEE Software* 6 (1996): 12-13.
78. Jakob Nielsen. 1991. Usability metrics and methodologies. *ACM SIGCHI Bulletin* 23, 2: 37-39.
79. James Clawson, Jessica A. Pater, Andrew D. Miller, Elizabeth D. Mynatt, and Lena Mamykina. "No longer wearing: investigating the abandonment of personal health-tracking technologies on craigslist." In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 647-658. ACM, 2015.
80. James E. Bailey and Sammy W. Pearson. 1983. Development of a tool for measuring and analyzing computer user satisfaction. *Management science* 29, 5: 530-545.
81. Jane Ogden and Jessica Lo. 2012. How meaningful are data from Likert scales? An evaluation of how ratings are made and the role of the response shift in the socially disadvantaged. *Journal of health psychology* 17, 3: 350-361.
82. Jason I. Hong, Jennifer D. Ng, Scott Lederer, and James A. Landay. 2004. Privacy risk models for designing privacy-sensitive ubiquitous computing systems. In *Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques (DIS '04)*, 91-100.
83. Jeffrey Rubin and Dana Chisnell. 2008. *Handbook of usability testing: how to plan, design and conduct effective tests*. John Wiley & Sons.
84. Jennifer D. Uhrig, Megan A. Lewis, Carla M. Bann, Jennie L. Harris, Robert D. Furberg, Curtis M. Coomes, and Lisa M. Kuhns. 2012. Addressing HIV knowledge, risk reduction, social support, and patient involvement using SMS: results of a proof-of-concept study. *Journal of health communication* 17, sup1: 128-145.
85. Jennifer S. Nachshen, "Empowerment and families: Building bridges between parents and professionals, theory and research." *Journal on Developmental Disabilities* 11, no. 1 (2005): 67-76.
86. Jenny Preece, Helen Sharp, and Yvonne Rogers. 2015. *Interaction Design - beyond human-computer interaction*. John Wiley & Sons.

87. Jeroen B. Guinée. 2002. Handbook on life cycle assessment operational guide to the ISO standards." *The international journal of life cycle assessment* 7, 5: 311-313.
88. Jill Fattor and Anne Friedlander. 2009. Texting 2 increase physical activity. *Texting 4 Health: A Simple, Powerful Way to Change Lives*, 67-74
89. Joan B. Simon, John J. Murphy, and Shelia M. Smith. "Understanding and fostering family resilience." *The Family Journal* 13, no. 4 (2005): 427-436.
90. Jocelyn Scheirer, Raul Fernandez, Jonathan Klein, and Rosalind W. Picard. 2002. Frustrating the user on purpose: a step toward building an affective computer. *Interacting with computers* 14, 2: 93-118.
91. John Brooke. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194: 4-7.
92. John D. Gould and Clayton Lewis. 1985. Designing for usability: key principles and what designers think. *Communications of the ACM* 28, 3: 300-311.
93. John L. Bennett 1984. Managing to meet usability requirements: establishing and meeting software development goals. *Visual display terminals*, 161-84.
94. John P. Chin, Virginia A. Diehl, and Kent L. Norman. 1988. Development of an instrument measuring user satisfaction of the human-computer interface. In *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '88)*, 213-218.
95. Jonathan Klein, Youngme Moon, and Rosalind W. Picard. 2002. This computer responds to user frustration: Theory, design, and results. *Interacting with computers* 14, no. 2 (2002): 119-140.
96. Joseph A. Gliem and Rosemary R. Gliem. 2003. Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales. *Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education*.
97. Joseph S. Dumas and Janice Redish. 1999. *A practical guide to usability testing*. Intellect Books.
98. Judith Donath. 2014. How social media design shapes society. In *Proceedings of the extended abstracts of the 32nd annual ACM conference on Human factors in computing systems (CHI '14)*, 1057-1058.
99. Judy Jeng. "Usability assessment of academic digital libraries: effectiveness, efficiency, satisfaction, and learnability." *Libri* 55, no. 2-3 (2005): 96-121.

100. Julia Merryman, Andrea Tartaro, Miri Arie, and Justine Cassell. 2008. Designing Virtual Peers for Assessment and Intervention for Children with Autism. In Proceedings of the 7th International Conference on Interaction Design and Children (IDC '08). ACM, New York, NY, USA, 81–84.
101. Julie A. Gazmararian, Baiyu Yang, Lisa Elon, Megan Graham, and Ruth Parker. 2012. Successful enrollment in Text4Baby more likely with higher health literacy. *Journal of health communication* 17, sup3: 303-311.
102. Julie A. Kientz, Rosa I. Arriaga, and Gregory D. Abowd. 2009. Baby steps: evaluation of a system to support record-keeping for parents of young children. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09), 1713-1722.
103. Julie A. Kientz, Rosa I. Arriaga, Marshini Chetty, Gillian R. Hayes, Jahmeilah Richardson, Shwetak N. Patel, and Gregory D. Abowd. 2007. Grow and know: understanding record-keeping needs for tracking the development of young children. In Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '07), 1351-1360.
104. Julie A. Kientz. 2010. Understanding parent-pediatrician interactions for the design of health technologies. In Proceedings of the 1st ACM International Health Informatics Symposium (IHI '10), 230-239.
105. Jum C. Nunnally, Ira H. Bernstein, and Jos MF ten Berge. 1967. *Psychometric theory*. Vol. 226. New York: McGraw-Hill.
106. Jurek Kirakowski and A. Dillon. 1988. *The computer user satisfaction inventory (CUSI): Manual and scoring key*. Cork, Ireland: Human Factors Research Group, University College of Cork.
107. Jurek Kirakowski and Mary Corbett. 1993. SUMI: The software usability measurement inventory. *British journal of educational technology* 24, 3: 210-212.
108. K.M. Cheng, V.J. Lin, K. Nijhawan, A. Westhem and M.S. Bernstein. 2017. Apps with Benefits: Using Benefits and Burdens to Predict Mobile App Usage. In Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems. ACM, 2452-2458.
109. Karen P. Tang, Sen H. Hirano, Karen G. Cheng, and Gillian R. Hayes. "Designing a mobile health tool for preterm infant wellness." In 2012 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops, pp. 232-235. IEEE, 2012.

110. Kari Mäkelä, Teemu Paavola, and Markku Stenman. 2010. Development of short message service application for patient-provider communication in clinical psychiatry. *Telemedicine and e-Health* 16, 7: 827-829.
111. Karin Källander, James K. Tibenderana, Onome J. Akpogheneta, Daniel L. Strachan, Zelee Hill, Augustinus HA ten Asbroek, Lesong Conteh, Betty R. Kirkwood, and Sylvia R. Meek. 2013. Mobile health (mHealth) approaches and lessons for increased performance and retention of community health workers in low-and middle-income countries: a review. *Journal of medical Internet research* 15, 1.
112. Kate Starbird and Leysia Palen. "Voluntweeters: Self-organizing by digital volunteers in times of crisis." In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 1071-1080. ACM, 2011.
113. Kevin Patrick, Fred Raab, Marc A. Adams, Lindsay Dillon, Marian Zabinski, Cheryl L. Rock, William G. Griswold, and Gregory J. Norman. 2009. A text message-based intervention for weight loss: randomized controlled trial. *Journal of medical Internet research* 11, 1.
114. Kevin Patrick, Fred Raab, Marc Adams, and L. Dillon. 2009. mDiet: A personalized approach to weight management using text messaging. *Texting 4 Health: A Simple, Powerful Way to Change Lives*, 35-47.
115. Kun-Ho Yoon and Hee-Seung Kim. 2008. A short message service by cellular phone in type 2 diabetic patients for 12 months. *Diabetes research and clinical practice* 79, 2: 256-261.
116. Kwok Chi Leong, Wei Seng Chen, Kok Weng Leong, Ismail Mastura, Omar Mimi, Mohd Amin Sheikh, Abu Hassan Zailinawati, Chirk Jenn Ng, Kai Lit Phua, and Cheong Lieng Teng. 2006. The use of text messaging to improve attendance in primary care: a randomized controlled trial. *Family practice* 23, 6: 699-705.
117. Lathy Prabhakaran, Wai Yan Chee, Kia Chong Chua, John Abisheganaden, and Wai Mun Wong. 2010. The use of text messaging to improve asthma control: a pilot study using the mobile phone short messaging service (SMS). *Journal of telemedicine and telecare* 16, 5: 286-290.
118. Laura Boccanfuso, Erin Barney, Claire Foster, Yeojin Amy Ahn, Katarzyna Chawarska, Brian Scassellati, and Frederick Shic. "Emotional robot to examine different play patterns and affective responses of children with and without ASD." In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 19-26. IEEE, 2016.

119. Lauren Brookman-Frazee, and R. L. Koegel. "Using parent/clinician partnerships in parent education programs for children with autism." *Journal of positive behavior interventions* 6, no. 4 (2004): 195-213.
120. Lawrence Mbuagbaw, Lahana Thabane, Pierre Ongolo-Zogo, Richard T. Lester, Edward Mills, Jimmy Volmink, David Yondo et al. 2011. The Cameroon mobile phone SMS (CAMPS) trial: a protocol for a randomized controlled trial of mobile phone text messaging versus usual care for improving adherence to highly active anti-retroviral therapy. *Trials* 2011, 12: 5.
121. Leanne M. Hirshfield, Erin Treacy Solovey, Audrey Girouard, James Kebinger, Robert JK Jacob, Angelo Sassaroli, and Sergio Fantini. 2009. Brain measurement for usability testing and adaptive interfaces: an example of uncovering syntactic workload with functional near infrared spectroscopy. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*, 2185-2194.
122. Learn the Signs. Act Early. 2009; <http://www.cdc.gov/ncbddd/autism/actearly/>.
123. Lee J. Cronbach and R. L. Thorndike. 1971. Educational measurement. Test validation, 443-507.
124. LENA. Retrieved Sept. 18, 2013 from <http://www.lenababy.com>
125. Linda J. Sax, Shannon K. Gilmartin, and Alyssa N. Bryant. "Assessing response rates and nonresponse bias in web and paper surveys." *Research in higher education* 44, no. 4 (2003): 409-432.
126. Louise Barkhuus and Anind K. Dey. 2003. Location-Based Services for Mobile Telephony: a Study of Users' Privacy Concerns. In *INTERACT*, 3: 702-712.
127. Louise Demers, Rhoda Weiss-Lambrou and Bernadette Ska. 2002. The Quebec User Evaluation of Satisfaction with Assistive Technology (QUEST 2.0): an overview and recent progress. *Technology and Disability* 14, 3: 101-105.
128. Louise Demers, Rhoda Weiss-Lambrou and Bernadette Ska. 2002. The Quebec User Evaluation of Satisfaction with Assistive Technology (QUEST 2.0): an overview and recent progress. *Technology and Disability* 14, 3: 101-105.
129. Maddalena Fiordelli, Nicola Diviani, and Peter J. Schulz. 2013. Mapping mHealth research: a decade of evolution. *Journal of Medical Internet Research (JMIR)*, 15, 5.
130. Maeve Duggan. 2013. Cell phone activities 2012. Pew Research Center.

131. Marios Koufaris. "Applying the technology acceptance model and flow theory to online consumer behavior." *Information systems research* 13, no. 2 (2002): 205-223.
132. Mark Bilandzic. 2010. The embodied hybrid space: designing ubiquitous computing towards an amplification of situated real world experiences. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction (OzCHI '10)*, 422-427.
133. Marsha White, and Steve M. Dorman. "Receiving social support online: implications for health education." *Health education research* 16, no. 6 (2001): 693-707.
134. Matthew Price, Deborah Williamson, Romina McCandless, Martina Mueller, Mathew Gregoski, Brenda Brunner-Jackson, Eveline Treiber, Lydia Davidson, and Frank Treiber. 2013. Hispanic migrant farm workers' attitudes toward mobile phone-based telehealth for management of chronic health conditions. *Journal of medical Internet research* 15, 4.
135. Meredith Ringel Morris. "Social networking site use by mothers of young children." In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pp. 1272-1282. ACM, 2014.
136. Michael Buhrmester, Tracy Kwang, and Samuel D. Gosling. 2011. Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data?. *Perspectives on psychological science* 6, 1: 3-5.
137. Miguel M. Serrano, Yu-Ping Chen, Ayanna Howard, and Patricio A. Vela. "Lower limb pose estimation for monitoring the kicking patterns of infants." In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2157-2160. IEEE, 2016.
138. Misha Kay, Jonathan Santos, and Marina Takane. 2011. mHealth: New horizons for health through mobile technologies. *World Health Organization*, 66-71.
139. Morris B. Holbrook. *Consumer value: a framework for analysis and research*. Psychology Press, 1999.
140. Nadia Dowshen, Lisa M. Kuhns, Amy Johnson, Brian James Holoyda, and Robert Garofalo. 2012. Improving adherence to antiretroviral therapy for youth living with HIV/AIDS: a pilot study using personalized, interactive, daily text message reminders. *Journal of medical Internet research* 14, 2.

141. Nicolas Ducheneaut, Nicholas Yee, Eric Nickell, and Robert J. Moore. "Alone together?: exploring the social dynamics of massively multiplayer online games." In Proceedings of the SIGCHI conference on Human Factors in computing systems, pp. 407-416. ACM, 2006.
142. Nicole B. Ellison, Charles Steinfield, and Cliff Lampe. "The benefits of Facebook "friends": Social capital and college students' use of online social network sites." *Journal of computer-mediated communication* 12, no. 4 (2007): 1143-1168.
143. Nigel Bevan. "Measuring usability as quality of use." *Software Quality Journal* 4, no. 2 (1995): 115-130.
144. Paul Legris, John Ingham, and Pierre Colletette. "Why do people use information technology? A critical review of the technology acceptance model." *Information & management* 40, no. 3 (2003): 191-204.
145. Penelope Sweetser and Peta Wyeth. "GameFlow: a model for evaluating player enjoyment in games." *Computers in Entertainment (CIE)* 3, no. 3 (2005): 3-3.
146. Per Egil Kummervold and Halgeir Holthe. 2008. Communicating textual health information to the mobile phones of visually impaired users. *Journal of telemedicine and telecare* 14, 4: 186-189.
147. Pew Research Center "Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally" 2019.
148. Predrag Klasnja, Sunny Consolvo, and Wanda Pratt. "How to evaluate technologies for health behavior change in HCI research." In Proceedings of the SIGCHI conference on human factors in computing systems, pp. 3063-3072. ACM, 2011.
149. Quantified Self. <http://quantifiedself.com>.
150. R.N. Bolton and K.N. Lemon. 1999. A dynamic model of customers' usage of services: Usage as an antecedent and consequence of satisfaction. *Journal of marketing research*, 171-186.
151. Rebecca Guy, Jane Hocking, Handan Wand, Sam Stott, Hammad Ali, and John Kaldor. 2012. How effective are short message service reminders at increasing clinic attendance? A meta-analysis and systematic review. *Health services research* 47, 2: 614-632.
152. Richard E. Boyatzis. 1998. Transforming qualitative information: Thematic analysis and code development. Sage.
153. Richard J. Holden and Ben-Tzion Karsh. "The technology acceptance model: its past and its future in health care." *Journal of biomedical informatics* 43, no. 1 (2010): 159-172.

154. Richard T. Lester, Paul Ritvo, Edward J. Mills, Antony Kariri, Sarah Karanja, Michael H. Chung, William Jack et al. 2010. Effects of a mobile phone short message service on antiretroviral treatment adherence in Kenya (WelTel Kenya1): a randomised trial. *The Lancet* 376: 1838-1845.
155. Robert Harmon, David Raffo, and Stuart Faulk. 2003. Incorporating price sensitivity measurement into the software engineering process. In *Management of Engineering and Technology (PICMET'03), Technology Management for Reshaping the World*. Portland International Conference on, 316-323.
156. Robert S. Weiss. 1995. *Learning from strangers: The art and method of qualitative interview studies*. Simon and Schuster.
157. Ron Neville, Alexandra Greene, John McLeod, Andrew Tracy, and John Surie. 2002. Mobile phone text messaging can help young people manage asthma. *British Medical Journal (BMJ)* 325, 7364: 600.
158. Rudy Den Buurman. 1997. User-centred design of smart products. *Ergonomics* 40, 10: 1159-1169.
159. Ruth M. Parker, Elena Dmitrieva, Sergei Frolov, and Julie A. Gazmararian. 2012. Text4baby in the United States and Russia: an opportunity for understanding how mHealth affects maternal and child health. *Journal of health communication* 17, sup1: 30-36.
160. S. Bauer, R. Percevic, E. Okon, R. U. Meermann, and H. Kordy. "Use of text messaging in the aftercare of patients with bulimia nervosa." *European Eating Disorders Review: The Professional Journal of the Eating Disorders Association* 11, no. 3 (2003): 279-290.
161. Sajay Arthanat, Stephen M. Bauer, James A. Lenker, Susan M. Nochajski, and Yow Wu B. Wu. 2007. Conceptualization and measurement of assistive technology usability. *Disability & Rehabilitation: Assistive Technology* 2, 4: 235-248.
162. Salar Jahedi and Fabio Méndez. 2014. On the advantages and disadvantages of subjective measures. *Journal of Economic Behavior & Organization* 98: 97-114.
163. Samantha A. Adams. "Revisiting the online health information reliability debate in the wake of "web 2.0": an inter-disciplinary literature and website review." *International journal of medical informatics* 79, no. 6 (2010): 391-400.
164. Sandra G. Hart and Lowell E. Staveland. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*, 52: 139-183..

165. Santosh Krishna and Suzanne Austin Boren. 2008. Diabetes self-management care via cell phone: a systematic review. *Journal of diabetes science and technology* 2, 3: 509-517.
166. Sasikanth Avancha, Amit Baxi, and David Kotz. 2012. Privacy in mobile technology for personal healthcare. *ACM Computing Surveys (CSUR)* 45, 1: 3.
167. Sean A. Munson, Debra Lauterbach, Mark W. Newman, and Paul Resnick. "Happier together: integrating a wellness application into a social network site." In *International Conference on Persuasive Technology*, pp. 27-39. Springer, Berlin, Heidelberg, 2010.
168. Seokwoo Song, Juho Kim, Bumsoo Kang, Wonjeong Park, and John Kim. "BebeCODE: Collaborative Child Development Tracking System." In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 540. ACM, 2018.
169. Shelah Leader, Phillip Jacobson, James Marcin, Ralph Vardis, Mark Sorrentino, and Dennis Murray. 2002. A method for identifying the financial burden of hospitalized infants on families. *Value in Health* 5, 1: 55-59.
170. Simon Walford, Edwin A.M. Gale, Simon Philip Allison, Robert B. Tattersall "Self-monitoring of blood-glucose: improvement of diabetic control." *The Lancet* 311.8067 (1978): 732-735.
171. Simona Rosati and Anna Saba. "The perception of risks associated with food-related hazards and the perceived reliability of sources of information." *International journal of food science & technology* 39, no. 5 (2004): 491-500.
172. Suk-Il Kim and Hee-Seung Kim. 2008. Effectiveness of mobile and internet intervention in patients with obese type 2 diabetes. *International journal of medical informatics* 77, 6: 399-404.
173. Sunny Consolvo, Predrag Klasnja, David W. McDonald, Daniel Avrahami, Jon Froehlich, Louis LeGrand, Ryan Libby, Keith Mosher, and James A. Landay. "Flowers or a robot army?: encouraging awareness & activity with personal, mobile displays." In *Proceedings of the 10th international conference on Ubiquitous computing*, pp. 54-63. ACM, 2008.
174. Sunyoung Park, Heeyong Jeong, and John Zimmerman. "ENSURE: Support for parents in managing their children's health." *the Proceedings of Design and Emotion* (2008).
175. Svetlana Yarosh, Panos Markopoulos, and Gregory D. Abowd. "Towards a questionnaire for measuring affective benefits and costs of communication technologies." In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pp. 84-96. ACM, 2014.

176. Tae-Jung Yun, Hee Young Jeong, Tanisha D. Hill, Burt Lesnick, Randall Brown, Gregory D. Abowd, and Rosa I. Arriaga. 2012. Using SMS to provide continuous assessment and improve health outcomes for children with asthma. In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium (IHI '12)*, 621-630.
177. Tamir Miloh, Rachel Annunziato, Ronen Arnon, Jill Warshaw, Sanobar Parkar, Frederick J. Suchy, Kishore Iyer, and Nanda Kerkar. 2009. Improved adherence and outcomes for pediatric liver transplant recipients by using text messaging. *Pediatrics* 124, 5: e844-e850.
178. Teresa E. Seeman. "Social ties and health: The benefits of social integration." *Annals of epidemiology* 6, no. 5 (1996): 442-451.
179. Tero Pikkarainen, Kari Pikkarainen, Heikki Karjaluo, and Seppo Pahnla. "Consumer acceptance of online banking: an extension of the technology acceptance model." *Internet research* 14, no. 3 (2004): 224-235.
180. Thiago Martini Da Costa, Bárbara Jaqueline Peres Barbosa, Durval Alex Gomes e Costa, Daniel Sigulem, Heimar de Fátima Marin, Aduino Castelo Filho, and Ivan Torres Pisa. 2012. Results of a randomized controlled trial to assess the effects of a mobile SMS-based intervention on treatment adherence in HIV/AIDS-infected Brazilian women and impressions and satisfaction with respect to incoming messages. *International journal of medical informatics* 81, 4: 257-269.
181. Thiago Martini Da Costa, Paulo Lísias Salomão, Amilton Souza Martha, Ivan Torres Pisa, and Daniel Sigulem. 2010. The impact of short message service text messages sent as appointment reminders to patients' cell phones at outpatient clinics in Sao Paulo, Brazil. *International Journal of Medical Informatics* 79, 1: 65-70.
182. Thomas Beauvisage. 2009. Computer usage in daily life. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems (CHI '09)*, 575-584.
183. Thomas J. Johnson and Barbara K. Kaye. "Using is believing: The influence of reliance on the credibility of online political information among politically interested Internet users." *Journalism & Mass Communication Quarterly* 77, no. 4 (2000): 865-879.
184. Tom Page. "Touchscreen mobile devices and older adults: a usability study." (2014)..
185. Tracy L. Westeyn, Gregory D. Abowd, Thad E. Starner, Jeremy M. Johnson, Peter W. Presti, and Kimberly A. Weaver. 2012. Monitoring Children's Developmental Progress Using Augmented Toys and Activity Recognition. *Personal Ubiquitous Comput.* 16, 2 (Feb. 2012), 169-191

186. Trevor Perrier, Nicola Dell, Brian DeRenzi, Richard Anderson, John Kinuthia, Jennifer Unger, and Grace John-Stewart. 2015. Engaging Pregnant Women in Kenya with a Hybrid Computer-Human SMS Communication System. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15) 1429-1438.
187. Trixie Tracker Software. Retrieved Sept. 18, 2013 from <http://www.trixietracker.com>
188. Ulla Strandbygaard, Simon Francis Thomsen, and Vibeke Backer. 2010. A daily SMS reminder increases adherence to asthma treatment: a three-month follow-up study. *Respiratory medicine* 104, 2: 166-171.
189. Vedran Ostojic, Branimir Cvorisec, Sanja Barsic Ostojic, Dimitry Reznikoff, Asja Stipic-Markovic, and Zdenko Tudjman. 2005. Improving asthma control through telemedicine: a study of short-message service. *Telemedicine Journal & e-Health* 11, 1: 28-35.
190. Vicki L. Hanson. "Age and web access: the next generation." In Proceedings of the 2009 International Cross-Disciplinary Conference on Web Accessibility (W4A), pp. 7-15. ACM, 2009.
191. Victoria L. Franklin, Annalu Waller, Claudia Pagliari, and Stephen A. Greene. 2006. A randomized controlled trial of Sweet Talk, a text-messaging system to support young people with diabetes. *Diabetic Medicine* 23, 12: 1332-1338.
192. William Albert and Thomas Tullis. *Measuring the user experience: collecting, analyzing, and presenting usability metrics*. Newnes, 2013.
193. William Douglas Evans, Jasmine L. Wallace, and Jeremy Snider. 2012. Pilot evaluation of the text4baby mobile health program. *BMC public health* 12, 1: 1031.
194. World Bank. 2012. *Information, Communication Technologies, and infoDev (Program)*. Information and Communications for Development 2012: Maximizing Mobile. World Bank Publications.
195. Xiaowen Feng, Susy Chan, Jacek Brzezinski, and Chitra Nair. "Measuring enjoyment of computer game play." *AMCIS 2008 Proceedings* (2008): 306.
196. Yung-Ming Li and Yung-Shao Yeh. "Increasing trust in mobile commerce through design aesthetics." *Computers in Human Behavior* 26, no. 4 (2010): 673-684.

Appendix A.

Baby Steps Web Portal Design

A-1. Landing Page

BABY STEPS

User Name

Password

Sign In

[Forgot your username or password?](#)

Home

Useful Links

Help

About

Welcome to Baby Steps!

A tool for parents of young children



Baby Steps can help you...

Track developmental milestones

Your first time? Sign up here.

Username

Invitation code

Register

[Baby Steps](#) is a research project at the [University of Washington](#) and is funded by the [National Science Foundation](#).

A-2. Home Page

BABY STEPS

Hello Mom!

Logout | Settings

Home

Milestones

Progress

Activities

Useful Links

Help

About

Emma

Christopher

Clare

Beyonce

Jack

Without holding on to anything for support, Emma kicks a ball by swinging her leg forward.



Emma Took Frank to the park



Favorite color



When looking at a picture book, Emma tells you what is happening in the picture. (e.g. 'barking,' 'running,' 'eating' or 'crying')



Emma's interactions with siblings



Emma tries to cut paper with child-safe scissors.



First time away from Emma



Let's Go for a Walk



Love Notes



Oh what a baby



While standing, Emma can throw a ball forward overhand by raising her arm to shoulder height.



Add a Memory

Memory

Date

Text

Photo

Video

Choose File

No file chosen

Save

Cancel

What did Emma do recently?

What is the funniest thing Emma did recently?

Submit

Does Emma climb the rungs of a ladder of a playground slide and slide down without help?

Not Yet

2019-03-04

Edit

Can Emma catch a large ball with both hands? Stand about 5 feet away and give 2-3 tries.

Not Yet

2019-03-04

Edit

Activity idea:

Allow Emma to step up on a footstool in the bathroom to see herself in the mirror. Teach them to wash their face with soap.

Tip of the Day

MMR vaccine. Offered 2nd dose between age 4-6 yrs. 2nd dose may be given before age 4 if 28 days have passed since 1st dose.

163

A-3. Milestone questions tab

BABY STEPS

Hello Mom!

Logout | Settings

Home

Milestones

Progress

Activities

Useful Links

Help

About

Emma

Christopher

Clare

Beyonce

Jack

2 Months

11 / 30 Complete

4 Months

5 / 30 Complete

6 Months

30 / 30 Complete

8 Months

30 / 30 Complete

9 Months

29 / 30 Complete

10 Months

0 / 30 Complete

12 Months

30 / 30 Complete

14 Months

3 / 30 Complete

16 Months

18 / 30 Complete

18 Months

21 / 30 Complete

20 Months

6 / 30 Complete

22 Months

48 Months

Milestones for 45 months to 51 months

30/30 Responses

Communication

When asked "What do you do when you are _____" (e.g. hungry, tired, etc.)?", does Emma answer properly?

Sometimes

2019-03-04

EditAdd to Timeline

Does Emma name at least 3 items from a common category such as animals, food? (e.g. 'Tell me some animals')

Sometimes

2019-03-14

Edit

Can Emma tell you at least 2 things about common objects? If you ask about a ball, she may reply, 'It's round. It's big.'

Yes

2019-03-04

EditAdd to Timeline

Does Emma use words such as 'a,' 'the,' 'am,' 'is,' and 'are' to make complete sentences, like 'I am going to the park'?

Does Emma use word endings, like 's,' 'ed,' and 'ing'? Does she say things like 'I am playing,' or 'I kicked the ball'?

Without help, can Emma follow 3 directions that are unrelated? For example: 'Clap hands, walk to the door, sit down.'

A-3a. Progress report (Control Group)

BABY STEPS 

Hello hyewon!
Logout | Settings

HomeProgressUseful LinksHelpAbout

Jason

Jason's Milestone Progress

39 Months to 45 Months26/30 Responses

- Communication: Behind schedule
- Fine Motor: On schedule
- Gross Motor: Close to on schedule
- Personal-Social: Incomplete
- Problem Solving: On schedule

Understanding Jason's Progress

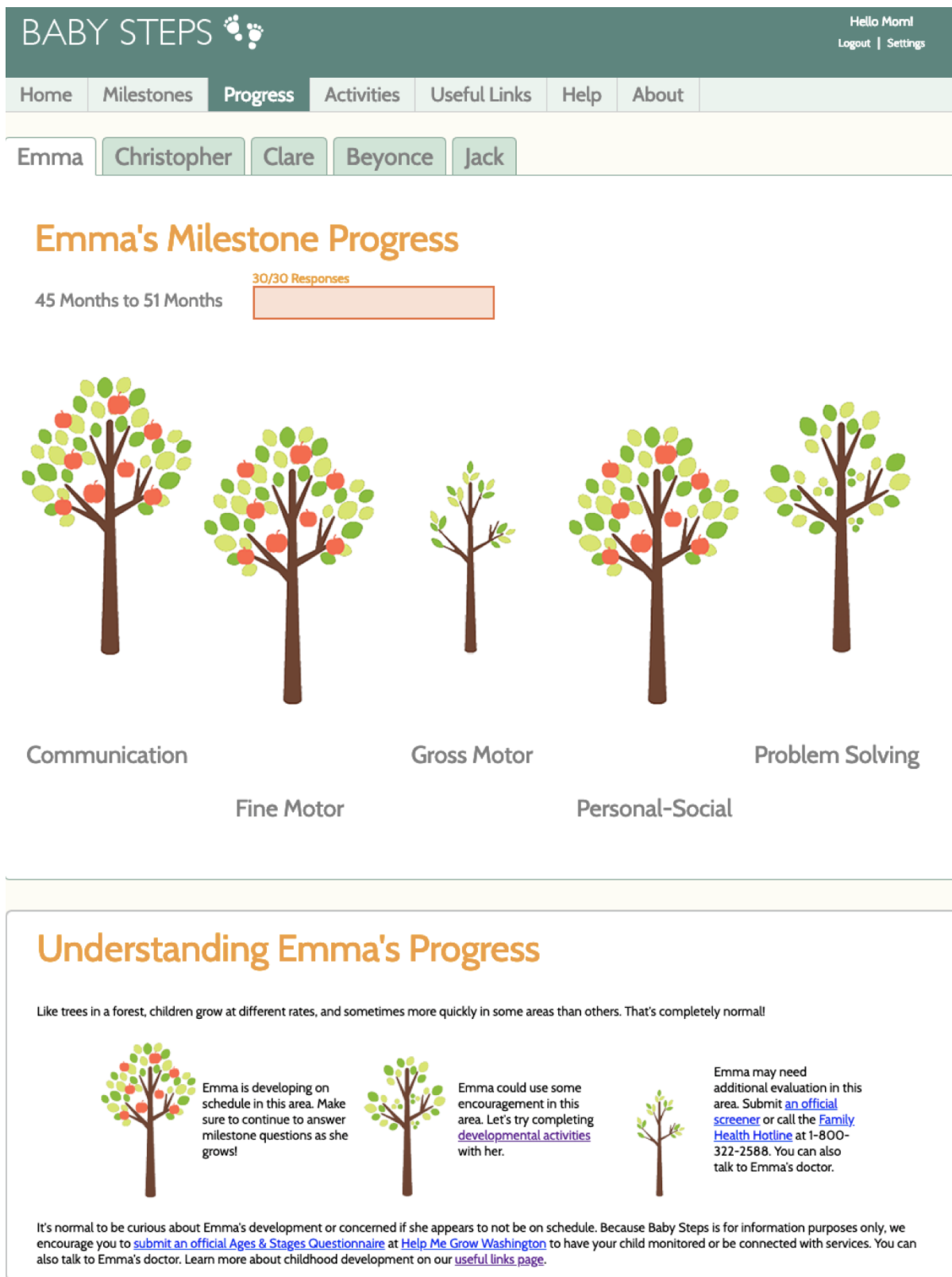
On schedule: Jason is developing on schedule in this area. Make sure to continue to answer milestone questions as he grows!

Close to on schedule: Jason could use some encouragement in this area. Let's try completing [developmental activities](#) with him.

Behind schedule: Jason may need additional evaluation in this area. Submit [an official screener](#) or call the [Family Health Hotline](#) at 1-800-322-2588. You can also talk to Jason's doctor.

It's normal to be curious about Jason's development or concerned if he appears to not be on schedule. Because Baby Steps is for information purposes only, we encourage you to [submit an official Ages & Stages Questionnaire](#) at [Help Me Grow Washington](#) to have your child monitored or be connected with services. You can also talk to Jason's doctor. Learn more about childhood development on our [useful links page](#).

A-3b. Progress report (Experimental Group)



A-4. Activity Ideas

BABY STEPS

Hello Mom!
Logout | Settings

Home

Milestones

Progress

Activities

Useful Links

Help

About

Emma

Christopher

Clare

Beyonce

Jack

1 - 4 Months

4 - 8 Months

8 - 12 Months

12 - 16 Months

16 - 20 Months

20 - 24 Months

24 - 30 Months

30 - 36 Months

36 - 42 Months

42 - 48 Months

48 - 54 Months

54 - 60 Months

Activities for Children 48-54 Months Old

Communication

Make puppets. Put on a puppet show of a story. Have a conversation with puppets by taking turns asking and answering questions.

We did this!

Take Emma on a trip to someplace new. Plan activities with Emma and ask questions about the trip when you return home.

☒

Edit

Add to Timeline

Place some household items in a paper bag. Let Emma pick an item without looking, and guess what it is.

☒

Edit

Help Emma make a book about herself by drawing/gluing pictures on paper. Have her "read" you her story and explain pictures.

☒

Edit

Give Emma simple directions to help around the house. Let her know what a big helper she is.

☒

Edit

Add to Timeline

Lay on your back outside and point out different cloud shapes and patterns together. Ask Emma what the clouds look like.

We did this!

A-5. Useful Links

BABY STEPS

Hello Mom!
[Logout](#) | [Settings](#)

Home

Milestones

Progress

Activities

Useful Links

Help

About

Useful Links

Child Development Information

Free resources on childhood development for parents

- [Developmental Screening \(FRC\)](#)
- [CDC Act Early campaign](#)
- [First Signs](#)
- [Baby Center Milestones](#)
- [CDC "Milestone Moments" booklet](#)
- [Zero to Three](#)

Children with Special Needs

Health information, resources, and family support for children with special needs

- [Children with Special Health Care Needs](#)
- [Special Needs Parent Support](#)
- [Special Needs Information Resources](#)
- [Family Voices of Washington](#)

Low-Cost Clinics & Immunizations

Community clinics that serve resource-constrained families

- [CDC Immunization Schedules for Your Children](#)
- [Community Clinics](#)
- [Immunization Clinics](#)
- [Free Clinics](#)

Parent Support

Early learning, safety, and parent support

- [Early Learning & Education Resources](#)
- [Parent & Family Support](#)
- [Parenting Classes](#)
- [Consumer Product Safety & Recalls](#)
- [Pregnant & Parenting Teen High School Program \(GRADS\)](#)

Online Parenting Communities

Talk about parenting issues and more with peers online

- [Baby Center Community Forums](#)
- [Reddit Parenting subreddit](#)
- [Mothering Forums](#)
- [CafeMom](#)
- [What to Expect Forums](#)

Washington State Families

Resources for families living in Washington State

- [Reach Out and Read](#)
- [Child Profile Health Information Mailings](#)
- [Program for Early Parental Support \(PEPS\)](#)
- [ParentTrust](#)
- [Puget Sound Parenting](#)

A-6. Help

BABY STEPS



Hello Mom!

[Logout](#) | [Settings](#)

Home

Milestones

Progress

Activities

Useful Links

Help

About

What is Baby Steps?

[Baby Steps and Child Development](#)

Milestone Questions

Activities

Baby Steps Web Portal

Baby Steps Text

Baby Steps Research Study

Baby Steps and Child Development

What is child development?

Babies grow not only physically but also psychological and emotionally. The early years of a child's life are very important for his or her health and development. As a parent, you notice when your baby smiles for the first time, rolls over, sits, walks and begins to talk. These moments mean that your child is learning and developing. Baby Steps helps you monitor your child's development in the following five developmental areas:

Communication	Child's language skills, this includes what they can say and what they understand
Gross Motor	Child's use and coordination of their arms and legs when they move and play
Fine Motor	Child's movement and coordination of their hands and fingers
Problem Solving	Child's problem solving skills and how they play with toys
Personal-Social	Child's self-help skills and their interactions with others

Why is developmental screening important?

According to new research from the U.S. Centers for Disease Control and Prevention (CDC), 17 percent of children have a developmental or behavioral disability such as autism or an intellectual disability (ADHD) in the United States. However, many children with developmental delays are not identified until after they start school. To ensure the best outcomes through early intervention, early detection is the key. Regular childhood screening through first 5 years of life can contribute to early detection.

What is developmental delay?

Developmental delay is when your child does not reach developmental milestones at the same time as other children who are within a similar age range. Children develop at their own pace, so it's impossible to tell exactly when children will learn a given skill. However, the developmental milestones give a general idea of the changes to expect, as a child gets older. As a parent, you know your child best. If your child is not meeting the milestones for his or her age, or if you think there could be a problem with your child's development, talk with your child's healthcare provider and share your concerns or contact the [Help Me Grow program](#) if you live in Washington State.

What do I do if I think my child's development progress is not on schedule?

The milestone questions asked by Baby Steps were inspired by the [Ages & Stages Questionnaire \(ASQ\)](#), which is used for screening only and not diagnosis of developmental delay. While the ASQ is well-validated, the questions in Baby Steps are different from those used on this questionnaire due to technology limitations (e.g., 140 character

169

A-7. About

Hi hyewon!

Logout | Settings

Home

Progress

Useful Links

Help

About

Research

Contact Us

Research

We are interested in designing technology to help detect, record, and track important developmental milestones that occur in children during their first 5 years of life. By tracking milestones, we can help parents and healthcare providers detect developmental delays such as autism or deafness earlier, which can improve the effects of interventions and lead to better outcomes for children.

We have developed design guidelines for developing technology to support new parents in record-keeping and implementing novel technologies to support better record-keeping and decision-making about developmental progress. We have used these guidelines to develop a set of tools, called Baby Steps, to help study how parents track information about their children and use it to make decisions about care.

Baby Steps is funded by the National Science Foundation.

What is Baby Steps?

A set of **interactive tools for parents to track developmental progress** in young children using a variety of technologies and connect to resources as needed.



Why Baby Steps?

Early detection is key to ensuring the best outcomes for these children & regular childhood screening through first 5 years of life is essential to early detection



Goal: Every child through age 5 is regularly screened and connected to early intervention services regardless of income, residency, or background

170

A-8. Settings

BABY STEPS

Hello Mom!
[Logout](#) | [Settings](#)

[Home](#) | [Milestones](#) | [Progress](#) | [Activities](#) | [Useful Links](#) | [Help](#) | [About](#)

My Account Settings

Account Information

Username: babysteps

Password: *****

Edit

Contacts

Name: Mom Baby Steps

Email: babystepsuw@gmail.com

Twitter: @babystepsuw

Mobile: 2099207553

Receive Texts: Yes

Receive MMS: Yes

Messages:

Name: Uncle

Email: russellhuebn

Twitter:

Mobile:

Receive Texts: Yes

Receive MMS: Yes

Messages:

Appendix B.

Invitation email to register on Baby Steps web portal (experimental group)



Hyewon Suh <fly2theskyhw@gmail.com>

[UW Baby Steps] Create your account on Baby Steps web portal and start using the system! (Invitation code included)

Hyewon Suh <hyewon25@uw.edu>

Tue, Mar 22, 2016 at 10:08 PM

To: Hyewon Suh <hyewon25@uw.edu>

Bcc: Elise.vanwest@gmail.com, Lauren Offenbecher <lauren.offenbecher@gmail.com>

Hello!

Thank you for completing the pre-study survey! We will send \$10 Amazon gift card to your email soon!

Now that you have completed everything to begin the study, we would like to invite you to Baby Steps web portal - babystepsuw.org. Create your account with an invitation code provided below. (Your 15-month long study participation will start once you create an account on Baby Steps web portal.)

Invitation code: 2241

Baby Steps web portal (Twitter and Text messaging option will become available soon!) has age-appropriate milestone questionnaires for your child. As a parent, you notice when your baby smiles for the first time, rolls over, sits, walks and begins to talk. These moments mean your child is learning and developing. Baby Steps helps you monitor your child's development in the following five developmental areas:

- **Communication** Child's language skills, this includes what they can say and what they understand
- **Gross Motor** Child's use and coordination of their arms and legs when they move and play
- **Fine Motor** Child's movement and coordination of their hands and fingers
- **Problem Solving** Child's problem solving skills and how they play with toys
- **Personal-Social** Child's self-help skills and their interactions with others

Regular childhood screening through first 5 years of life can contribute to early detection of developmental delays, which affects about 13% of children 3 to 17 years of age in the United States. Throughout 15 months of your study participation, try to access Baby Steps regularly and complete the questionnaires as your child develops.

From this point, Baby Steps team will try to remain behind the scene with a few exceptions:

- when Twitter and Text messaging option becomes available (Stay tuned!)
- mid-study survey (no more than three times)
- periodic compensation (\$20 on 4, 8, 12, and 15 months completion and \$10 on each survey completion)

But if you have any questions about any parts of study participation, please don't hesitate to contact Hyewon Suh at (206) 395-8050 or by email at babystep@uw.edu. I also attached a copy of the consent form for you to keep.

Once again, our team is very happy to have you on our study and hope you find this experience to be both enjoyable and useful.

Thank you!

Baby Steps Research Team

 **Baby Steps (L) Consent Form (approved).pdf**
1213K

Appendix C.

Instruction sheet on Text Messaging

How to start using Baby Steps via text messages:



Starting tomorrow, you'll receive developmental milestone questions via text messages (from 425-374-0735).

If you do not wish to receive text messages or do not want to receive messages with pictures, you can choose to opt out from Baby Steps web portal (<http://babystepsuw.org/>) Settings -> edit Contacts -> uncheck boxes.

Respond to milestone questions

After watching your child, respond to questions by sending **Y / yes**, **S / sometimes**, or **N / not yet**. If you have multiple children registered with Baby Steps, also include child's name (e.g. Diana S). After you reply, you will receive the next question until you answer all 30 questions for a given period.

Can Jose throw a ball using both hands?

sometimes

If you do not want to answer a specific question at the moment, you may choose to come back to it later by replying **L or later** and you will soon receive the next question. If you want to take your time, simply do not respond to questions and the system will automatically remind you again in 48 hours.

Check your child's progress

Text **Report** to see how your child is developing based on your responses. Have any questions or concerns about your child's development? Call the Family Health Hotline at 1-800-322-2588 for resources.

report

- Communication: On Track
- Gross Motor: Let's Watch
- Fine Motor: More Evaluation Needed
- Problem Solving: On Track
- Personal-Social: Let's Watch

Receive activity suggestions

Want to try some activities to encourage your child's development? Text **Activity** to Baby Steps and receive an activity idea. Want more? Text again!

Record memorable moments

Is there a moment you want to add to your child's timeline? Text **Remember** followed by the memory. You can also include a picture that goes with the memory!

Remember Alison's first birthday!



Receive a random memory

Do you want to see what's on your timeline? Text **Memory** to receive a previous memory and reminisce!

To request help

If you don't remember commands or need more information, text **Help** to Baby Steps for instructions.

To stop receiving messages

If you want to stop receiving milestone questions from Baby Steps, text **STOP**.

To start receiving messages again

If you want to start receiving messages after you stopped, text **START**.

*Baby Steps Text is **NOT** case sensitive