Assessing the Impact of Usability Design Features of an mHealth App on Clinical Protocol Compliance Using a Mixed Methods Approach

by

Mandar Patwardhan

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved June 2016 by the
Graduate Supervisory Committee:

Kevin Gary, Chair
Armando Pina
Ashish Amresh

ARIZONA STATE UNIVERSITY

August 2016

ABSTRACT

In the last decade the number of people who own a mobile phone or portable electronic communication device has grown exponentially. Recent advances in smartphone technology have enabled mobile devices to provide applications ("mHealth apps") to support delivering interventions, tracking health treatments, or involving a healthcare team into the treatment process and symptom monitoring. Although the popularity of mHealth apps is increasing, few lessons have been shared regarding user experience design and evaluation for such innovations as they relate to clinical outcomes. Studies assessing usability for mobile apps primarily rely on survey instruments. Though surveys are effective in determining user perception of usability and positive attitudes towards an app, they do not directly assess app feature usage, and whether feature usage and related aspects of app design are indicative of whether intended tasks are completed by users. This is significant in the area of mHealth apps, as proper utilization of the app determines compliance to a clinical study protocol. Therefore it is important to understand how design directly impacts compliance, specifically what design factors are prevalent in non-compliant users. This research studies the impact of usability features on clinical protocol compliance by applying a mixed methods approach to usability assessment, combining traditional surveys, log analysis, and clickstream analysis to determine the connection of design to outcomes. This research is novel in its construction of the mixed methods approach and in its attempt to tie usability results to impacts on clinical protocol compliance. The validation is a case study approach, applying the methods to an mHealth app developed for early prevention of anxiety in middle school students. The results of three empirical studies are shared that support the construction of the mixed methods approach.

To my mother and in memory of my father.

ACKNOWLEDGMENTS

I would like to specially acknowledge Dr. Kevin Gary for his continuous support and guidance during the last two years. He constantly pushed me to my limits to bring out the best in me. I am grateful for getting an opportunity to work with him.

I would like to thank Dr. Armando Pina, Ryan Stoll and the entire psychology research team for their continuous support during last two years. This thesis work wouldn't have been possible without their help and invaluable feedback. I would like to acknowledge Dr. Ashish Amresh for his continuous encouragement and inputs during this period.

There are several names that I would like to mention here; Priyanka Vats and Rahul Parekh for being a family away from family for last two years; Soham Abhyankar for helping me through every phase of my thesis work and being someone I could always count on; and last but not the least my friends at ASU (Akshay, Japa, Aveesha, Avani, Chinmay, Priyansha, Kartik and Sampada) who always stood by me and supported me.

I thank my parents for giving me this opportunity to pursue my career in the field of my interest. It would not have been possible without their support and love.

TABLE OF CONTENTS

Page

CHAPTER Page

CHAPTER Page

CHAPTER Page

APPENDIX Page

LIST OF TABLES

Table Page

Table Page

LIST OF FIGURES

Figure Page

1 INTRODUCTION

In the last two years the number of people who own a mobile phone or portable electronic communication device has grown exponentially (Poushter, 2016). Recent advances in mobile technology have enabled mobile devices to perform functions previously not possible with handheld devices, such as delivering interventions, tracking health treatments, involving a healthcare team into the treatment process, symptom monitoring and more (Klasnja & Pratt, 2012). These innovative applications (or "apps") have evolved into a new field known as mobile health (mHealth) (Fiordelli et al., 2013; Free et al., 2010). mHealth technologies include mobile phones, PDAs, smartphones; handheld and ultra-portable computers such as tablet PCs. The biggest advantages of using mobile devices, and in particular mobile phones, for health are that these devices are personal, intelligent, connected, and always with people (Fogg, BJ, 2009; Whittaker, 2012).

Although the popularity of mHealth apps is increasing, few lessons have been shared regarding the user experience design and evaluation for such innovations as they relate to clinical outcomes. This research studies the impact of usability features on clinical protocol compliance and success, going beyond painting technology with a broad brush, but instead acknowledging that the design details matter. On the smartphone platform, there is a lot of competition for user attention, and the research community is emphasizing usability studies of mHealth applications (Jaspers, 2009). These usability studies primarily rely on survey instruments to assess efficacy. Surveys are effective in determining user perception of usability and positive attitudes towards an app. However surveys do not tell the entire story. They miss out on the details of the app feature usage, and whether feature usage and related aspects of app design indicate whether intended tasks or completed by users or not.

The main contribution of this thesis is a mixed methods approach to assess the usability of mHealth applications. This mixed method approach extends the use of surveys by adding user interaction log analysis to determine compliance, and clickstream analysis to attempt to determine patterns of where users get “off track” (become non-compliant to a design intent). Surveys are used in the traditional way, primarily to measure a user’s perception of the usability of the mHealth app. This thesis does explore customizations of existing survey instruments tailored to get user feedback about design features in an app (Chapter 6, section 6.2.7). Log analysis consists of log data; a form of data representing interactions of the user with the mobile application. Log analysis in this research is used for measuring user task completion. Clickstream analysis is a method popular in web analytics that deals with identifying usage patterns in web pages. An application of this method from web analytics is used in this research to identify usage patterns from log data. These patterns are used for task identification and interaction sequence mining. Together, these three methods give a better understanding of the impact of the mobile app design on clinical outcomes, most importantly compliance. To my knowledge this is the first study using a mixed method approach for usability validation. This study can be used as a causal connection between design of the mHealth application and its impact on clinical outcomes.

The mixed method approach for usability validation of mHealth apps can be applied to a wide variety of healthcare domains. For the purpose of this thesis, the healthcare domain in consideration is chronic disorders. U.S. National Center for Health Statistics defines chronic disease as a long-lasting condition that can be controlled but not cured (CDC, 2016). Examples of well-known chronic diseases are sickle cell disease, asthma, cardiovascular disease, arthritis and cancer. There is a consistent increase in chronic disorders and the number of people suffering from chronic illnesses (Wang et al., 2014). Rapid technological advances, increasing adoption rates, and the ubiquitous nature of modern smartphones make them a promising option for chronic illness diagnoses and management.

This research focuses on anxiety as a chronic disorder, specifically targeting child anxiety. This disorder is among the most prevalent psychiatric problems in children with rates ranging from 5% to 10% and as high as 25% in adolescents (Patwardhan, Stoll, Hamel, Amresh, et al., 2015). Anxiety disorders may cause significant impairments, that typically fail to remit spontaneously, and are prospectively linked to clinical depression and problematic substance use for some children (Silverman, Pina, & Viswesvaran, 2008). Because of these reasons, targeting anxiety disorders is a significant public health concern.

Clinicians are often interested in measuring compliance of patients to the clinical protocol. According to the study conducted by Cramer, compliance is defined as "*degree or extent of conformity to the recommendations about day-to-day treatment by the provider with respect to the timing, dosage, and frequency*" (Cramer et al., 2008, Volume 2, Number 1, Page 44). Matthews (Matthews et al., 2015) suggests that compliance to a clinical protocol can be increased by emphasizing design while developing the app. He suggests that a more specific and user centric mHealth app yields better compliance. One objective of this research focuses on the design of an mHealth app specific to childhood anxiety disorders by involving subject matter experts and care providers directly in a participatory design process with an intent of increasing protocol compliance.

Pina and colleagues (Silverman, Kurtines, Jaccard, & Pina, 2009) present a prevention and early intervention protocol named REACH for child anxiety disorder treatment. Prevention involves building protective factors and skills, increasing support, and reducing risk factors or stressors. Early intervention deals with addressing a condition early in its stages of manifestations. This research is based on REACH. The details of the REACH protocol are explained in chapter 3 section 3.3.

This thesis' contributions are a case study in participatory design of an mHealth app for a pediatric chronic disease, and a novel method of usability validation that attempts to tie design outcomes to clinical outcomes (namely compliance). Although these contributions are limited to a single domain, protocol, and app, the outcomes are of interest due to the chronic disorder domain (anxiety), the nature of the intervention (preventative-early intervention), the use of an app to increase protocol compliance, and the integration of concepts from innovative design technology (gaming, notifications, user experience design) resulting in improved clinical outcomes. Further, I hope this research contributes to a growing multidisciplinary need to connect clinical research methods with (software) engineering processes.

The rest of this thesis work is divided into 6 chapters. Chapter 2 presents a review of related multidisciplinary literature. Chapter 3 sets the context for this research by presenting the problem in detail and providing an overview of the research approach. Chapter 4 describes the multidisciplinary team based design process and the implementation steps for developing an mHealth app, followed by a preliminary study conducted with the mHealth app. Chapter 5 explains mixed methods of usability validation, used in this research. Chapter 6 gives details about the experimental studies and discusses the results. Chapter 7 concludes with lessons learnt and future scope in this area.

2 BACKGROUND LITERATURE

This chapter discusses literature related to this research. As this research deals with the design of an mHealth application for addressing child anxiety issues, the first section will briefly discuss the evolution of mHealth. The second section reviews relevant literature about childhood anxiety and the clinical methods that act as a foundation to solve these problems. Subsequently, existing mHealth applications for child anxiety issues will be discussed. Finally the literature that emphasizes on the design of the mHealth application is discussed to define the context for this research work.

2.1 Evolution of mHealth

Bashshur and colleagues (Bashshur et al., 2011) present a taxonomy of telemedicine. According to the authors, telemedicine originated in 1969 that led to the origination of eHealth in late 1990s and mHealth was introduced by Istepanian and colleagues (Istepanian, Laxminarayan, & Pattichis, 2006) in 2003. Kay in her research says that mHealth is "*medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants (PDAs), and other wireless devices*" (Kay, 2011, Page 6). For the purposes of this research, the term 'mHealth' is used to mean clinical and public health activities involving mobile devices.

2.2 Domain of Childhood Anxiety

A significant problem faced by patients suffering from mental health and especially anxiety related issues is the attitude towards the treatment. Mobile apps are designed in a way to handle the stigma associated with receiving mental health care. They also have the potential to reduce health disparities and improve the engagement of the patients in this type of health care delivery process (Price et al., 2014). Patients tend to report a preference for mobile devices in the completion of research studies because a mobile device gives a sense of privacy to the patient. Owning and using a phone in public is something that is more socially accepted than completing a paper form. Adherence to real-time self-monitoring may be enhanced when conducted via mobile phone or a similar handheld device (Price et al., 2014).

Anxiety disorders are among the most prevalent psychiatric problems in children with rates ranging from 5% to 10% and as high as 25% in adolescents. Anxiety disorders also cause significant impairment, typically fail to spontaneously remit, and are prospectively linked to clinical depression and problematic substance use for some children. This group of negative outcomes has led to the development of evidence-based interventions aimed at pre-empting anxiety disorders (Pina et al., 2012).

Aaron Beck in early 1960's developed and introduced Cognitive Behavioral Therapy (CBT) (Michelle et al., 2014). CBT is psychotherapy advised for mental disorders such as Social Anxiety Disorder (SAD) and depression. Its basic principle is understanding the beliefs and behavioral patterns of the patient, using a variety of goal-oriented strategies to change their underlying flawed beliefs (Michelle et al., 2014). Pramana (Pramana et al., 2014) says CBT has been recognized by the American Psychological Association Taskforce as an effective treatment for childhood anxiety disorders and it is considered first-line choice by most children and families. But the authors also acknowledge that the classic design of the evidence-based prevention programs consists of 16-20 weekly therapy session that may discourage families to commit to the entire length of the treatment because of schedule and transportation difficulties. Pramana's research emphasizes a more feasible and affordable version of CBT named as Brief CBT or BCBT; wherein the 16-20 week CBT session is replaced by 8 sessions. This research is based on a BCBT protocol named REACH (Silverman et al., 2009). REACH for Success (REACH hereafter) is a school-based cognitive-behavioral protocol designed for 4th and 5th graders for the indicated prevention and early intervention of childhood anxiety and related problems. The details of the REACH protocol are explained in chapter 3, section 3.3.

Homework, or self-practice at home, is a central concept to CBT. In CBT, after a face-to-face session with therapists, patients are typically asked to carry out homework. Patients fill in worksheets or diaries using pen-and-paper forms between two consecutive face-to-face sessions with the therapist. Richard LeBeau (LeBeau et al., 2013) suggests that improvement of homework compliance has the potential to be a highly practical and effective way to improve clinical outcomes in CBT targeting anxiety disorders. Though homework provides crucial information about patients to the therapist (Michelle, Jarzabek, & Wadhwa, 2014), the way in which homework diary activities are carried out also matters. Stone (Stone et al., 2002) in his paper talks about the comparison between a paper based homework activity vs an electronic based homework activity. He observed that patients are highly non-compliant with a paper-based version of a homework diary, whereas they are highly compliant with an electronic version of the same diary.

mHealth apps can play central roles in evidence-based therapies. Apps designed for mobile devices present an opportunity to extend the reach of the therapist beyond the face to face sessions. This research takes this idea of using an mHealth app based on REACH to provide out of session support and to provide a means to complete homework in the form of activities in the app. The next subsection discusses existing mHealth apps that have been developed for childhood anxiety disorders.

2.3 Existing Child Anxiety Based Studies

The pilot study based on Brief CBT for child anxiety treatment that is relevant to this research, uses an mHealth platform entitled SmartCAT (Smartphone-enhanced Child Anxiety Treatment) (Pramana et al., 2014). SmartCAT consists of an mHealth app for patients to practice out of session tasks, an online portal for therapists to monitor skill practice, communicate with patients, and manage rewards, and a communication protocol to manage communication between the app and the portal. In a pilot study with nine children (9 to 14 years old) involving a BCBT, patients reported high usability for the app and were compliant with the BCBT protocol. The limitation to this study was that the sample size was small to assess the feasibility, utility and acceptability of the SmartCAT platform. The results of this research suggest that an mHealth app like SmartCAT can be successfully integrated into CBT for children with anxiety disorders. The author acknowledges that including goal setting through rewards has the potential to increase a patient's participation in the treatment.

The second study that is relevant to this research is *'FRIENDS for Life'* (Fisak Jr, Richard, & Mann, 2011; Rodgers & Dunsmuir, 2015) (FRIENDS hereafter), a school-based CBT program developed by Paula Barrette at the Pathways Institute in Australia, that develops children's skills to enhance emotional regulation, coping mechanisms and thinking styles. There are several studies that demonstrate the effectiveness of FRIENDS in reducing anxiety (Barrett, 2000; Rodgers & Dunsmuir, 2015; Stallard et al., 2014), both immediately after program implementation and at longer term follow-up and when facilitated by trained teachers, nurses and mental health professionals. The website (FunFriends, 2016) for FRIENDS program does mention a mobile app called FunFriends. This app is designed for the consolidation and maintenance of the skills taught in the FRIENDS programs. It uses a game based approach. It requires parents to play this game with children to practice and refine resilience techniques. Though the website gives a brief overview of the FunFriends app, no literature was found that demonstrates its effectiveness in FRIENDS prevention process. The classic design of FRIENDS program is simply not feasible or sustainable in schools (e.g., there are too many sessions, sessions are too long, manuals are too cumbersome and not organized for real world implementation, too much training is required, and preparation is too time consuming) (Patwardhan et al., 2015).

In contrast, REACH was created from evidence-based exposure-based cognitive-behavioral protocols as a practical intervention that can build a foundation for sustainable large-scale diffusion. That is, REACH was streamlined into 6 sessions (instead of the typical 12-15), each 20-30 minutes in length (rather than the typical 60 to 90 minutes), and uses an easy- to-follow manual (each session is condensed into one page front and back while FRIENDS, for example, has an 89 page manual). One concern with REACH, however, is that such a streamlined protocol may result in a lower dosage of the active change ingredients and fewer opportunities for children to practice coping skills because there are fewer sessions and less practitioner feedback time. This drawback can be addressed by filling the gap between sessions by providing means to practice active change ingredients by using an mHealth app.

These two programs based on CBT were most relevant to the domain of childhood anxiety. However there are other apps worth mentioning that play a key role delivering mental disorder treatment. Rizvi (et al., 2011) describes an app called 'DBT Field Coach' that is used as a means to facilitate instructions, exercises, reminders and other components to help borderline patients cope with emotional crises. Examples of components used in the app are video and audio messages from the therapist, games designed to distract from intense emotions, and motivational images. Results showed the patients did use the app when needed, there was reduction in intense emotions and substance use cravings, and the app assisted in improving symptoms of depression and distress during the treatment period. Another such application called CBT MobilWork was mentioned in a research paper by Price (et al., 2014) that was an application for adults with severe depression. This app prompts users to complete basic homework assignments and coaches them through the process in real time.

As seen in these two apps, mobile applications offer several methods to complete activities like homework exercises, promoting adherence, collecting real-time data through a prompt for assessment and even providing helpful feedback when the patient engages in an activity or completes an assessment.

2.4 mHealth Application Usability

There has been a recent increase in the use of mobile technologies that address various mental health issues, but there is also an acknowledgement that only deploying these technologies may not be enough, and that the design of the patient-facing element (in this case, the mHealth app) is crucial to patient engagement (Patel, Asch, & Volpp, 2015). Different techniques like personas or role-play have been used to gain an understanding of context and clinical settings (Matthews et al., 2015), but it is also necessary to involve individuals with direct experience of a mental illness when developing patient-facing systems. These efforts, though valuable, focus on only high lever user needs and does not take into consideration the context of people living with mental illness (Matthews et al., 2015).

Matthews (et al., 2015) in his research presents a novel method called *"in situ design"* of designing an mHealth application for patients suffering from bipolar disorder. By "in situ design" they put forth an idea of involving patients and clinicians in the design process of the application. They take into consideration the distinctive characteristics of bipolar disorder and related illnesses in order to identify more innovative and effective designs. Matthews also talks about introducing *"Rewarding Interactions"* in the form of leader boards and badges to make challenging or unappealing tasks more engaging. In situ design showed the importance of involving patients during the design phase of the application. On similar lines, this thesis involves psychology subject matter experts (SMEs) as proxies in the design process and uses a prototypical iterative process based on the feedback received from SMEs. The details of design process will be explained in chapter 4.

Usability is important to the design, development, and delivery of mHealth applications. The ISO standard (ISO 9241, 1988) defines *usability* as the extent to which a user can use a product to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use. Formal usability testing involves recording and analyzing user interactions with the application to figure out common usability issues. Usability issues include something that prevents task completion, digresses someone from intended goal or creates confusion. Poor usability is considered to be a primary cause for failed adoption of health technologies. Price (Price et al., 2014)in his research says that patients will not use technology that is difficult to use or perceived as irrelevant to their needs. Thus, usability validation is an essential component while evaluating any mHealth strategy.

Jaspers (Jaspers, 2009) discusses four usability testing methods to evaluate a design against its requirements. The first method is *expert-based versus user-based evaluation*. In an expert-based technique, the user interface is evaluated based on heuristics or questions derived from a general knowledge of how humans process through tasks. A user-based approach includes performance measurement, keystroke analysis, log-file data, and satisfaction questionnaires. The second method is *heuristic based usability testing* (Nielsen, 1994). It is used to evaluate a user interface based on recognized usability principles (or heuristics). It does a twofold evaluation of the interface, one to get the basic idea, second to know the detailed flow of the app. Flaws are recorded against heuristics and are reported accordingly. The third technique is *cognitive walkthrough*. This technique deals with evaluating a system design based on learnability and exploration. It is highly structured and explicitly guided by user tasks. In this technique, the user is supposed to explore the task based on guidelines, such as: 1) User sets a goal to be accomplished, 2) User inspects available options on the screen. 3) User selects the option that is the best option to reach the goal that was set 4) User performs action and gives feedback. The last technique for usability testing discussed in this paper was '*Think Aloud*'. In this technique, usability evaluation consists of collecting think aloud protocols and then analyzing these protocols to obtain a model of the cognitive process. These protocols are collected by instructing subjects to solve a problem while 'thinking aloud'; that is, stating directly what they think.

In conclusion Jaspers suggests that the use of a particular technique in testing the usability of the system purely depends on the context and the availability of the subjects and having an idea about their background. Concurrent use of more than one method seems to show better results in finding usability issues. User-centered design based evaluations (cognitive walkthrough) seem to find a lot of usability issues in healthcare applications.

Currently, the usability studies conducted by the psychologists primarily rely on survey instruments to assess efficacy. Surveys are effective in determining user perception of usability and positive attitudes towards an app. However surveys do not tell the entire story. They miss out on the details of the app feature usage, and whether feature usage and related aspects of app design indicate whether intended tasks or completed by users or not. This thesis was motivated from the Jaspers paper and it incorporates mixed method approach in evaluating usability of the application. This mixed method approach extends the use of surveys by adding user interaction log analysis to determine compliance, and clickstream analysis to attempt to determine patterns of where users get "off track" (become non-compliant to a design intent). The details of this mixed method approach will be explained in chapter 3, section 3.2.

A new area of research in the usability of mHealth applications is interaction analysis. This is an emerging area that focuses on identifying intentions of users when they launch an application on their smartphone, and understanding which tasks they actually execute. Lettner (Lettner et al., 2014) describes using user sequences (clickstreams) in mobile apps to understand the actual completion of the task by users against intended tasks. In this paper the author talks about a novel approach of automatically extracting and grouping user sequences against predefined tasks and presenting them visually. This visual representation helps to find out if the designer's intention of how users should perform designed tasks, and how they actually execute them, matches, and where it differs. This paper was a motivation for this thesis because one of the mixed method approach used in this thesis for usability testing is similar to Lettner's approach of identifying expected versus actual usage patterns of the users.

The literature discussed in this chapter introduced concepts like mHealth, CBT and usability that play an important role in this thesis. The papers discussed in the literature explain the importance of mHealth as a platform to assist clinicians in solving anxiety disorders. The research by Matthews is most relevant because it aligns perfectly with the motivation of this thesis. However, unlike the papers above, this research emphasizes on the design and usability of the mhealth application that impacts the patient's adherence to the clinical protocol. As seen in the research by LeBeau (et al., 2013) , compliance to the clinical protocol (compliance to the homework activity prescribed by the clinician in this case) plays a significant role in getting the desired clinical outcomes. This research takes this idea and uses an mHealth application that assists the patients to practice homework in between the therapy sessions thereby making them compliant with the clinical protocol.

The next chapter sets the context for this research. It outlines the research questions that drives this thesis work and presents the method in brief. It also gives a brief overview of the solution and the validation process.

3 RESEARCH CONTEXT

This research studies the impact of user experience design features on clinical protocol compliance and success, going beyond painting technology with a broad brush, but instead acknowledging that the design details matter. The main contribution of this thesis is a mixed methods approach to assess the usability and the impact of design features on clinical objectives of mHealth applications. This chapter sets the context of my research by presenting the research questions, followed by a description of the research methodology consisting of mixed methods, and an overview of the case study conducted for validation.

3.1 Research Questions

This thesis' contributions are a case study in the participatory design of an mHealth app for a pediatric chronic disorder (child anxiety disorder), and a novel method of usability validation (mixed methods approach) that attempts to tie design outcomes to clinical outcomes (namely compliance). My research questions are:

RQ 1: Does the introduction of an mHealth app in the anxiety prevention process increase compliance of patients to the clinical protocol?

RQ 2: Do specific user experience design features of an mHealth app significantly affect the rate of patient compliance to the clinical protocol?

The first research question addresses whether the introduction of an mHealth app increases patient compliance to a clinical protocol. In a clinical protocol targeted towards patients suffering from anxiety disorders, the patients have to complete a set of activities prescribed by the clinician. The patients are said to be compliant with the protocol if they complete these activities as prescribed by the clinician between the protocol sessions. These activities are considered as homework assignments. There are several methods to facilitate homework assignments in a clinical trial. Our hypothesis is that using an mHealth app during the prevention process as a means to complete homework assignments may increase patient compliance to the clinical protocol.

The first research question simply asks if patient compliance increases due to the introduction of mHealth technology. Answering this question positively validates the application of the technology in this domain. However it does not provide insight into the causal connection between the technology and the desired clinical outcome (compliance). The second research question dives deeper into the details of the app usage to identify particular user experience design features of the mHealth app that may impact compliance. There are specific design features in the app, such as age appropriate theming, notifications, gamification, and ease of navigation that may affect mHealth app usage in a positive or negative way. This knowledge of key design features affecting patient compliance helps to inform better design of the app. This combination of multiple methods (surveys, log analysis, interaction mining) to determine this connection of design decisions to clinical outcomes (compliance) is a novel contribution of this work. The next section will discuss the mixed method approach to assess the usability of the mHealth app.

3.2 Research Methodology

On the smartphone platform, there is a lot of competition for user's attention and the research community is emphasizing usability studies of mHealth applications (see Chapter 2 section 2.4). These usability studies primarily rely on survey instruments to assess efficacy. This thesis uses a mixed methods approach to assess the usability of mHealth application. This mixed method approach extends the use of surveys by adding user interaction log analysis to determine compliance, and clickstream analysis to attempt to determine where users get “off track” (become non-compliant to a design intent). Surveys are used in the traditional way, primarily to measure a user’s perception of the usability of the mHealth app. I explore customizations of existing survey instruments tailored to get user feedback about design features in an app. Log analysis consists of log data; a form of data representing interactions of the user with the mobile application. Log analysis in this research is used for measuring user task completion. Clickstream analysis is a method popular in web analytics that deals with identifying usage patterns in web pages. An application of this method from web analytics is used in this research to identify usage patterns from log data. These patterns are used for task identification and interaction sequence mining. Together, these three methods give a better understanding of the impact of the mobile app design on clinical outcomes, most importantly compliance. The next subsections will discuss the mixed methods in detail. The discussion will start with surveys, followed by log analysis and finally clickstream analysis.

3.2.1 Surveys

Surveys are effective in determining user perception of usability and positive attitudes towards an app. Survey data in this research consists of the user feedback in the form of answers to the survey questionnaire. The surveys used in the study were based on the USE questionnaire that measures usability, satisfaction, and ease of use (Lund, 2001). Table 1 presents an example of a survey used in this research that collects responses of patients for the specific statements about the app to assess its usability. A ten-point scale is provided to give responses for each statement.

Surveys are used primarily to measure user perception of mHealth app usability. This research customizes existing survey instruments to get tailored user feedback about design features in an app. For example, statements 7 to 9 are tailored to get patient’s feedback about a specific screen in the app. The responses to these statements are gathered and are referred to as survey data in this research. Statistical analysis of this data yields one component of the usability measures.

Not at all		Somewhat								Very much
I am happy with this app	1	2	3	4	5	6	7	8	9	10
I would tell a friend about this app	1	2	3	4	5	6	7	8	9	10
This app is fun to use	1	2	3	4	5	6	7	8	9	10
This app works the way I would want it to work	1	2	3	4	5	6	7	8	9	10
I was able to use the app on my own without any help	1	2	3	4	5	6	7	8	9	10
I would want to continue working with the app	1	2	3	4	5	6	7	8	9	10
The text message in the screen was easy to read	1	2	3	4	5	6	7	8	9	10
Buttons on this screen made it easy to navigate	1	2	3	4	5	6	7	8	9	10
I liked the color scheme used on this screen	1	2	3	4	5	6	7	8	9	10

Table 1 Sample Survey Questionnaire

3.2.2 Log Analysis

While the patient is following steps depicted in the clinical protocol using an mHealth application, the patient's interactions on the phone are tracked via a logging system running on the mobile device. Each swipe and tap on the screen is logged along with the timestamp. This type of data is referred to as log data and the analysis of this data is referred as log analysis in this research. The log data gives the exact representation of the interaction of the user with the app and it gives a clear indication of usage. Table 2 shows an example of log data representing interaction of an actual user with the app. The first column represents the timestamp of the tap on the screen, the second column indicates the activity done by the patient and the last column indicates the location in the app where the said activity was carried out. Log data is used in this thesis to measure compliance of patients to the clinical protocol. Based on the log data, discrete interactions of the user representing entrance to an activity and completion of the activity are measured using counting measures. For example, the Worryheads in Table 2 (shown in bold) represents a scenario where user starts the activity and completes the steps inside an activity. This marks as a completion of the activity.

Timestamp	Activity Step	Place
11:08:52.524	RELAXATION_INTRO	RELAXATION
11:09:56.137	**WORRY_HEADS**	**LANDING_PAGE**
11:10:07.518	**WORRY_HEADS_NEXT_CLICKED**	**WORRY_HEADS_ACTIVITY**
11:10:12.060	**WORRY_HEADS_NEXT_CLICKED**	**WORRY_HEADS_ACTIVITY**
11:10:51.062	**WORRY_HEADS_O_RIGHT**	**WORRY_HEADS_ACTIVITY**
11:10:51.111	**WORRY_HEADS_COMPLETED**	**WORRY_HEADS_ACTIVITY**
11:14:39.274	DAILY_DIARY	LANDING_PAGE
11:14:44.732	DAILY_DIARY_STATE_ONE_NEXT_CLICKED	DAILY_DIARY
11:14:46.999	DAILY_DIARY_RESPOND_CLICKED	DAILY_DIARY
11:14:56.177	DAILY_DIARY_CANCEL_CLICKED	DAILY_DIARY
11:15:00.652	DAILY_DIARY_STATE_ONE_BACK_CLICKED	DAILY_DIARY
11:15:02.809	STOP_STARTED	LANDING_PAGE
11:15:12.375	STOP_RESPOND_BUTTON_CLICKED	STOP_ACTIVITY
11:16:01.776	STOP_DONE_BUTTON_CLICKED	STOP_ACTIVITY
11:16:04.646	STOP_S_NEXT_CLICKED	STOP_ACTIVITY
11:16:07.453	STOP_RESPOND_BUTTON_CLICKED	STOP_ACTIVITY
11:16:44.714	STOP_DONE_BUTTON_CLICKED	STOP_ACTIVITY

Table 2 Sample Log Data

3.2.3 Clickstream Analysis

On a website, clickstream analysis (sometimes called clickstream analytics) is the process of collecting, analyzing, and reporting aggregate data about webpages visited by users and the order in which they were visited. To do this type of analysis, researchers use the succession of mouse clicks made by each visitor called the *clickstream*. (Surya & Sharma, 2013). This idea of clickstream analysis is quite popular in web analytics but limited literature is available for its use in mobile applications. The use of clickstream analysis in this research deals with analyzing the application usage patterns of patients based on interactions observed in the log data. These usage patterns during clinical protocol sessions provide insights on the specific design features within the application responsible for the compliance measures observed during log analysis.

Figure 1 shows an example of a clickstream graph. Colored circles indicate states in an activity. The black circles indicate the start and end states of the activity. Let us take a hypothetical example with 19 children participating in study, wherein they have to complete a certain activity on the mobile app that consists of 4 states. Out of 19 children who started the activity, 12 completed all four steps and 7 backtracked to start step without completing all the four states. This type of behavior is referred to as a pattern. The black arrows indicate a pattern of compliance and the blue arrows indicate patterns of non-compliance.

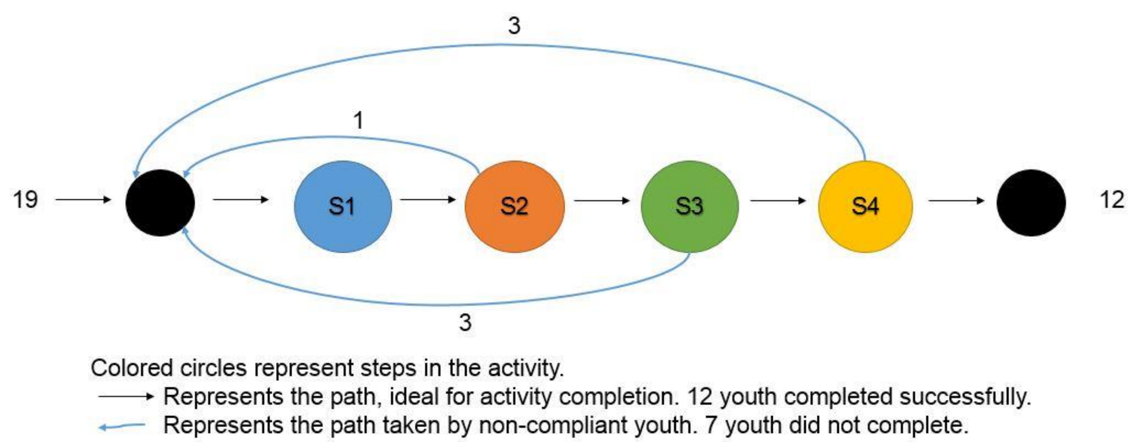


Figure 1 Clickstream Analysis Example

The previous two sections explained the research questions and the research method of this thesis. To understand the domain of the case study, it is important to discuss the child anxiety protocol in detail. The next section will provide details about the REACH protocol for childhood anxiety prevention. Subsequent section gives an overview of the implementation of the mHealth app used to conduct research studies for validation in this thesis.

3.3 REACH Protocol

As mentioned in the second chapter, REACH is an indicated prevention and early intervention targeting anxiety in children (Pina et al., 2012; Silverman et al., 2009). It is an exposure-based cognitive-behavioral protocol delivered in 6 sessions, each 20-30 minutes in length, and administered in a group format. REACH uses the core exposure-based cognitive and behavioral procedures common to the protocols typically evaluated via randomized controlled trials (RCTs) (Silverman et al., 2009; Silverman & Kurtines 1996). Each session (S) in the manual is organized in terms of Overview, Content (didactic, games), Review/Closing, and After the Session (homework). The protocol focuses on broad-based exposure and problem-solving skills that have a wide reach for the range of anxiety disorders targeted. Unique session content is presented in Table 3.

Core skill acquisition and practice tools include the use of Daily Diaries, Guided Relaxation, S.T.O.P. (Scared, Thoughts, Other-thoughts, Praise), and S.T.I.C. (Show That I Can). Daily Diaries are used to facilitate self-evaluation of emotion expressiveness. Subjects self-monitor and describe in writing the anxiety or fear provoking situations that occurred during the week. Subject also rates the severity of anxiety/fear associated with the situation using a 0-8 feelings thermometer. Lastly, subjects describe in writing thoughts that occurred before/during/after the situation (e.g., worries) and actions that resulted (e.g., avoidance behaviors).

Session Number	Session Content
1.	Introduction (group name, rules, and confidentiality), Learn about emotions, and Relaxation.
2.	Define worries, Learn cognitive self-control, and Practice cognitive self-control (Worryheads game).
3.	
4.	Define social skills and Learn about conversation skills (starting and managing conversations). Practice conversations (make-believe game).
5.	Learn about assertiveness and Practice assertiveness (stand-up! game).
6.	Learn to face situations and Engage in behavioral exposures to mild-moderate anxiety-provoking situations.

Table 3 REACH Session Details

In terms of guided Relaxation, children are provided with pre-recorded standardized step-by-step procedures designed to improve self-regulation of anxiety related physiological hyperarousal via breathing exercises, muscle tension/release exercises, and imagery. When it comes to cognitive self-control, a four-step coping plan is introduced via the "S.T.O.P." acronym where S = Scared? T = Thoughts, O= Other [thoughts], P = Praise. S.T.O.P. is first practiced via the Worryheads game by using pre-written emotionally ambiguous and anxiety provoking scenarios along with an accompanying "worry thought". Children are then asked to change the "worry thought" for a more realistic and alternative solution to the scenario provided. In the game, successful resolution of the worry thought results in advances toward a common goal for each player (reaching the end to win the game). Subsequently, with basic knowledge of S.T.O.P., children engage in prospectively applying the technique to situations that emerge as anxiety or fear provoking for them during the course of each week. Lastly, behavioral exposures are introduced via S.T.I.C. jobs (S.T.I.C. = Show That I Can). S.T.I.C.s are provided in the form of a pre-written or prepopulated Fear Hierarchies based on modules from the Anxiety Disorders Interview Schedule for Children where each avoidance behavior has been pre-populated for the child as individual exposures.

This section explained in detail the REACH protocol. Next section will give a brief overview of the mHealth app that was used for validation purpose in this research.

3.4 REACH mHealth App

This first generation of the REACH app was designed to provide support for the out-of-session practice of intervention skills rather than act as a stand-alone platform, as some have suggested that implementation of child anxiety interventions probably requires interventionist involvement (Pramana et al., 2014). It is important to note here that the design process of the REACH app was a combined effort of two teams; one consisting of software engineers and second consisting of psychology subject matter experts (SMEs). The efforts in developing the REACH app were guided by a User and Subject Matter Expert Centered Design that utilized personas, prototyping with an iterative process, and expert feedback from an advisory board comprised of practicing social workers, school psychologists, and counselors (Patwardhan et al., 2015). The details of the design process of the REACH app and the implementation details are explained in detail in chapter 4.

The Android app was self-contained; it did not rely on communication services to offload data storage or real-time processing. Instead, the focus was on leveraging the device as a dosage vehicle for intervention and data collection. In terms of technology features, we included speech capture, thematic and age-appropriate media, gaming (e.g. progressive reward incentives), notifications presented to the target user in fixed (daily time-based) and adaptive (based on user interactions) schedules, password-based authentication for adults (e.g., interventionists, parent, teacher), on-device database to store user responses and actions (e.g., to estimate alarm fatigue, motivation, clinical content such as ratings of distress associated with an anxiety provoking situation), and a data export feature (comma-separated files). The REACH app was used in the case study to provide out of session support to REACH protocol. The implementation details of the REACH app along with the details of the features are explained in chapter 4 section 4.3. The next section will talk about the evolution of the user studies conducted for this research.

3.5 Evolution of User Studies Based on the Mixed Methods

This case study consisted of three user studies. These user studies were conducted in public schools with parental consent and assent from the child. The first study was a preliminary user study wherein instructions were given to the children to complete a set of activities in the REACH app in a school based REACH protocol session. A survey was given to the children at the end of the clinical trial that consisted of questions based on identified design features of the app. The questions in this survey were framed based on Usefulness, Satisfaction and Ease of Use questionnaire (USE; Lund, 2001). The idea behind these surveys was to measure user's perception of the REACH app.

A second user study was conducted with a larger number of participants compared to the first study. Along with usability, compliance was measured for the activities using the log analysis method. In this study, clickstream analysis was done to identify the reasons for non-compliance in log analysis. Though this study was better than the first user study with respect to number of participants, it was still a single session based study and it did not provide insight as to what design features affected aspects of compliance and clinical outcomes

To overcome the limitations of the previous two studies, a third user study was conducted with children from public schools for six-week duration of the REACH protocol. Based on the log analysis in this study, a compliance measure for the REACH protocol was calculated. This study also identified the design features responsible for compliance.

In summary, this chapter sets the context for this research by providing details about the research questions, research method composed of mixed methods and finally giving details about the case study consisting of REACH protocol and the REACH mHealth app. The next chapter discusses the iterative design process and the implementation of the REACH mHealth app, followed by a full presentation of the validation studies done with the mixed methods approach.

4 DESIGN AND IMPLEMENTATION PROCESS

This thesis is based on the premise that emphasis on the design of an mHealth application will impact patient compliance to a clinical protocol, thereby resulting in desired clinical outcomes. To that end, this chapter discusses in detail about the design process and the implementation of the REACH mHealth application. The first section presents a gap analysis between REACH protocol components. Subsequently, the next section presents a patient centered design process. It is important to note here that the design process of the REACH app was combined effort of two teams; one consisting of software engineers and second consisting of psychology subject matter experts (SMEs). This chapter is largely from the paper that was published in Wireless Health 2015 conference (Patwardhan, Stoll, Hamel, Amresh, et al., 2015).

I was one of the developers in the software engineering team. I was involved throughout the design and implementation process. My role was, designing and implementing a database to manage the data for the REACH app. I was also responsible for implementing a background service for the app that was used to manage the REACH protocol schedule.

4.1 REACH Protocol Components

REACH is a pre-existing protocol, so the first design activity was to review program materials and workflow, seeking opportunities to effectively translate existing steps, and later innovating on smartphone-specific solutions to achieve the domain objectives for increased dosage, engagement, and feedback. To better understand the domain of the app, the SMEs shared the provider manual of the REACH protocol to the designers and the materials for delivering the protocol (board games, handouts, MP3s). The manual describes how the sessions, each conducted consecutively over the course of six weeks, employ specific practice worksheets, information gathering forms, and interactive exercises designed to train children in the preventive and coping skills. The main activities defined in the manual were Daily Diary, Relaxation, S.T.O.P, Worryheads board game, and S.T.I.C.

Table 4 summarizes the protocol component steps and highlights challenges in porting these steps to the mobile environment.

REACH	*Component Description / Design Challenges*
Daily Diary	Self-monitoring
	engagement; daily compliance; rich data entry
Relaxation	Pre-recorded audio exercises
	media porting and translation
S.T.I.C.	Behavioral exposures with adult feedback
	preserving steps; rewards; feedback
S.T.O.P.	Self-application of cognitive self-control plan
	encouraging tool engagement through positive UX
Worryheads	Learn and practice cognitive self-control plan with provided scenarios
	detailed alternatives; increasing dosage; feedback

Table 4 Reach Protocol Components and Gap Analysis

A round of stakeholder interviews involving the SMEs followed the domain research of the REACH protocol. These included working sessions between the design team leads and the SMEs, visits by the SMEs to the design team's lab, and synchronous question-answer sessions over email and videoconferencing. This step of the process addressed difficulties relating to understanding the protocol and assumptions on both sides regarding implementation objectives. This step took longer than expected, with a result of inconsistent understanding of implementation outcomes. The design team conducted an internal review to identify root causes and come up with design process alternatives. The causes identified included:

1. New terminology.

2. Gaps in understanding by the design team with respect to the protocol.

3. Assumptions of the designers based on past development experience.

4. Ad hoc communications patterns between SMEs and the design team, and within the design team itself.

5. A lack of understanding of the end user context.

Together, these issues are not uncommon in design processes, and some were addressed (1, 3, 4) through simple awareness of the issue in the team review. For example, improving ad hoc communication patterns was improved through more frequent design team meetings, clarifying the lines of communication with SMEs, and reiterating design team understanding of requirements back to the SMEs for validation. Issues #2 and #5 were more significant. Issue #2 represents a "blind spot" in design, due to factors such as missing information implicitly understood by the SMEs but not apparent to the design team. Issue #5 was a recognition that the design team did not understand who would be using the app and in what context. At this point the design team realized a more patient-centric approach was required to overcome these design obstacles.

4.2 A Patient Centered Design Process

The design process described in the previous section focused on translating a field manual; it is not surprising that the translation had gaps derived from implicit knowledge assumed by the manual authors and not understood by the designers. The software engineering researchers suggested a more user-centric approach, where the needs of the end user, in this case the patients, is the focus of the design process. The gold standard for such a design process is User-Centered Design (UCD). UCD assumes a participatory design process with end users, but for this research we prefer the more inclusive definition of UCD as "*the active involvement of users for a clear understanding of user and task requirements, iterative design and evaluation, and a multi-disciplinary approach*." (Vredenburg et al., 2002, Page 472, Volume No 4, Issue No 1). ISO 9241-210 identifies 6 principles to UCD (quote):

1. The design is based upon an explicit understanding of users, tasks and environments.
2. Users are involved throughout design and development.
3. The design is driven and refined by user-centered evaluation.
4. The process is iterative.
5. The design addresses the whole user experience.
6. The design team includes multidisciplinary skills and perspectives.

These principles were especially attractive to the design team due to the uniqueness of the domain and protocol, and identified issues understanding the end user context. The team realized the app would not be a direct translation of the paper-based REACH protocol, and needed to focus on context and end user experience.

There is a wide range of practices supporting UCD; the design team utilized personas, prototyping with iterative feedback, participatory design, and end user validation. The SMEs served as participatory designers, eliminating the back-and-forth ad hoc aspects of the initial process. They also served as proxies for the end users during design as gaining access to children (4th-5th grade users for an extended time for intense design activities was not possible). Access to end users would have certainly been preferable during the design process but was not possible at the time. However end user validation was emphasized before approving the app for protocol trial; these results are reported in chapter 6. Fortunately, prior domain research and SME interviews from the gap analysis proved useful in the context of the UCD.

4.2.1 Personas

The design team started the UCD process by developing personas, or proxies for categories of end users, and inviting the SMEs to review them. The SMEs were not familiar with personas, and after overcoming initial confusion about the technique, gained enthusiasm and effectively provided useful feedback. The personas shared with the SMEs are presented in Table 5.

Persona 1	Jacob is 10 years old, and is currently being raised by his single mother. He was held back for behavior problems as he tends to lash out when stressed. When confronted with even minor change he shuts down, and becomes irritable. His goal is to do as little as possible, or just enough so he doesn't get in trouble.
Persona 2	Jessie is 9 years old and very shy. In larger groups of 10 or more people she panics, and is dangerously on edge. She has a strong recognition of her symptoms, and works very hard at overcoming them. Her goal is to be free from required effort as soon as possible.
Persona 3	Mike is 12 years old. He finds it difficult interact in groups. He thinks that everyone has prying eyes on him and judging his every move. He loves to read books and is distracted by day dreaming. He gets very anxious and nervous in social situations.
Persona 4	Elizabeth is 10 years old. She is relatively overweight and is embarrassed in evaluative situations. When her classmates tease her, she cries and withdraws from interacting with peers. This typically happens during physical education and school games.

Table 5 Personas

Iterating over these personas led to several design insights that were previously not understood by the design team. For example, the design team came to understand subjects in this domain have a higher need for re-assurance; respond well to attention and approval, and are highly compliant (persona 2). Discussion of the personas with the SMEs further revealed that in community samples girls are more likely identified as “anxious” than boys, and anxious children fear the evaluative nature of social situations (personas 3 and 4). After capturing a clearer idea about end user context through discussing the personas created with the SME, the design team started a phase of rapid prototyping to ensure the SMEs provided frequent feedback on each design decision.

4.2.2 Rapid Prototyping

Rapid prototyping is an iterative design technique refining the details of interaction models and overall user experience. Early prototypes, or storyboards, focus on task sequences, or the mapping of task workflows to interface screens. This leads to user interaction modeling; the identification of user input actions effecting transitions between screens or for the capture of critical information. Later iterations refine these models and also layer in thematic elements, until a final design is converged upon. Iterations are meant to be short, frequent, and focused on answering specific questions regarding the user experience.

4.2.2.1 Storyboarding and Clickthrough Prototypes

The freely available Pencil prototyping tool (Pencil Tool, 2015) was used to construct screen and clickthrough mockups. Clickthroughs take simple screen mockups and overlay “hot regions” that advance the mock to a new screen, simulating a user interaction. One drawback is the tool runs its simulations in a web browser so tap and swipe gestures are not supported; however, the tool does support mobile UI “skins” to promote a look-and-feel consistent with the mobile user experience. Figure 2 shows an example of an early mockup created for S.T.O.P. activity.

The team created mockups of different scenarios in the app. Each mockup was peer-reviewed within the design team, validated against the documented protocol, and then presented to the SMEs for feedback. The design was iteratively refined until the scenario interactions were adequately captured, and the design team felt comfortable moving to implementation on the Android platform.

Figure 2 S.T.O.P. Mock-up in Pencil Tool

4.2.2.2 Translating Protocol Components

As identified in the gap analysis (section 4.1), some protocol components are a fairly straightforward translation, or port, to the mobile app, while others are not. For example, the Relaxation audio components were a straightforward port of the media to the device wrapped with a simple consistent interaction metaphor. Of course this component also requires the least user interaction of any of the components. On the contrary, the Worryheads game is a multiplayer board game involving cards. The app required limiting the game experience to a single user compared to the multiplayer board game. The design team replaced the physical cards in the board game with preset

"Situations" and "Thoughts" screens. The user was then presented with a choice of four of "Other Thoughts" options to choose from. Once the user selects a choice from possible options a praise message was showed on the screen to appreciate the correct answer. Screens depicting Worryheads are shown in chapter 5.

A design concern in translating the protocol was the significant amount of text a child is asked to input during activities such as the Daily Diary and S.T.O.P. The mobile device is not suited for textual input that goes beyond instant messaging or social media apps, and further the end users are at an age where they are often mobile-aware, but not proficient mobile typists. The fear was that textual input would be skipped or significantly limited, or in the worst-case cause frustration of the app to the extent children would abandon it. The design team identified speech capture input as a means to facilitate better information capture.

4.2.3 Injecting Innovations in the Mobile Experience

A challenge in applying mHealth concepts to existing clinical protocols is the desire to innovate versus leveraging validated protocol steps. For this research, the mobile platform provided the means for increasing dosage by virtue of the device being ever- present. However, ubiquity is not enough, end users must be motivated to practice the protocol. Engagement was addressed through innovative design features introduced in the mobile platform including thematic and age-appropriate media, game strategies (e.g. progressive reward incentives), and mobile notifications.

4.2.3.1 Designing an Appropriate Theme

A user interface theme refers to the consistent application of stylistic elements such as images, fonts, audio or video media, and user interface widgets (buttons, menus, taps, etc.). To gain acceptance of the app amongst users familiar with the paper protocol, the design team used the same theme used in the paper protocol. The team ensured that color codes and the fonts used in paper based protocol and the fonts used in the app are same. To design the features of the app, the team studied the paper-based versions of the activities to be performed by children to get a better idea of how to replicate the activities in the application. The team followed the same nomenclature of the existing activities in the screen designs reduce confusion and gain rapid acceptance.

The user experience required a gender-neutral, age-appropriate proxy for the human guide who assists in the existing REACH protocol. This proxy personifies the guide, providing instruction and feedback to the end user through the mobile interface. Initial ideas focused on themes such as “feed your pet” or “grow your plant” but were rejected as being either too “babyish” for the target age range or gender-biased.

The design team came up with the idea of an animated motivational character in the form of a blob. The design team referred to the character as “Blob” (Figure 3), but the male name is never used in the app itself. Based on game design concepts, “Blob” presents an age-appropriate, gender-neutral proxy for protocol guidance and feedback (Murray et al., 2013; Norman et al., 1986).

4.2.3.2 Progressive Reward Incentives

While one of the goals of the REACH protocol is to empower children to be intrinsically motivated to enact the protocol, at the training stage it is imperative to repeat the dosage faithfully in order to attain this intrinsic motivation. A common gamification technique is to employ leveled rewards as an extrinsic motivator for performing a targeted behavior (Ryan, R. M. & Deci, E. L., 2000). Therefore a simple progressive (leveled) set of rewards for extrinsic motivation was included in the app design. When an end user completes a task from the REACH protocol they get a reward in the form of the Blob's tricks. This way the user is motivated to follow the protocol and completing the tasks (dosage) so s/he can unlock more complicated tricks for the Blob.

One concern SMEs raised during the design process was the potential to inadvertently punish the child for not performing a task. Given the domain, a design invariant was specified to keep all interactions with the child positive; therefore, all language and emotive expressions of Blob throughout the app were scrubbed to ensure there were no negative connotations. For progressive rewards, a setting in the app was designed to unlock new tricks twice every week. The presence of these tricks also served as extrinsic motivation for engagement.

4.2.3.3 Smartphone Notifications

Mobile platforms offer an "always on" communications channel between service providers and end users. Most categories of mHealth apps emphasize the communications channel between clinicians and patients, or between patients and automated big data platforms on the cloud. The REACH app is unique in that it does not leverage the mobile device as a communications channel. In this generation of the app, the focus is on leveraging the device as an information collector and dosage vehicle for the protocol. In this sense the device serves more as a Personal Digital Assistant (PDA) than as a connected mobile phone.

In this modality it is still important to present to the end user a feeling of connectedness. The personification of Blob as a proxy guide is one way the design provides this connectedness. As a second design concept, the design team wanted to make use of mobile notifications, but without relying on cloud-based push notifications as these would require a persistent network connection. Therefore the design supports local notifications presented to the end user in both fixed and adaptive schedules.

Fixed schedules are daily time-based notifications, such as for the Daily Diary, to complete a regular interval task. Adaptive notifications require tracking end user interactions with the app and dynamically determining whether to issue a notification to engage with Blob again. The designers were concerned with the notion of alarm fatigue through over-notification, though the mobile device was given to the end users as a locked down tool for practicing the protocol, and not as a general- purpose smartphone for personal use.

4.2.3.4 Security and Privacy

Any mHealth app needs to be concerned with how user data is stored, transmitted, and identified. These concerns can become overbearing nonfunctional requirements on the app and down to the underlying mobile operating system providing the communication and storage services. At this stage of the app's development, it made more sense to de-identify data and work in a locked-down, disconnected mode. There were several simplifying assumptions the design team was able to make:

1. The emphasis on increased dosage over remote monitoring of compliance or personal health measurements puts this research in a different class of mHealth apps. Such apps push data to remote providers (often via a cloud-based service) and support human or automated communication reminders.

2. The relatively small number of participants in planned early studies meant the devices, with a specific chosen version of the mobile operating system, could be purchased and distributed to end users. The design team selected a Motorola phone running Android API version 19 (KitKat).

3. The relatively small number of participants makes it easier to de-identify the data and manage it external to the app. A secret user interaction combined

with a password protects access to functionality that supports exporting user interaction and task completion data (see above).

Of course these assumptions will have to change in future generations of the platform to facilitate broader adoption. But as a dosage augmentation platform, the design team leveraged the weekly visits with the psychologists combined with the computational sophistication of modern smartphone platforms to provide a self-contained solution.

4.2.3.5 Customized Navigation

The designer's intent of how users should perform designed task, and how they actually execute them in the field, matters. For the activities in the REACH app, navigation buttons, feedback messages and screens are used to customize the navigation of users within the app in such a way that, they complete the intended task in least possible steps. Feedback messages are embedded in activities like S.T.O.P. and Daily Diary to help the user to navigate between the steps of these activities. The *Home* button is provided in every activity in the app to give an option for user to directly go to the landing page.

4.2.3.6 Input Methods

Activities like S.T.O.P., Daily Diary and S.T.I.C. require a user to provide an input. There are two methods to provide a user response in the app. The first method is the default keyboard available on the android phone. This keyboard can be used to provide a text input that gets recorded by the app. The second, more innovative design feature used in the app for providing user response is speech recognition. Google Application Programming Interface (API) based speech recognition option is provided in the aforementioned activities. Users can tap on the mic icon on the keyboard and speak loudly and clearly to provide input in these activities. The Google API converts speech to text and provides a textual representation of the speech and displays it on the screen. These subsections discussed various innovations in the app in the form of design features. The usability validation mixed methods will examine, how the design features like *Appropriate Theming, Notifications, Customized Navigation, Rewards and Input Methods,* affect protocol compliance. The next section will focus on implementation details of the REACH app.

4.3 REACH App Implementation

The Android platform was selected to support the app. The openness of the Android platform, the availability of low-cost devices, the ease of the Google Speech API, and the ability to deploy the app without the involvement of an app store were the deciding factors for the first generation of the app. The first section will discuss the features of the app that are derived from the REACH paper based protocol activities. The second section will talk about the features of the REACH app that are not part of the REACH protocol but are provided as means to facilitate the patients, clinicians and researchers in this study. The last section will discuss the external validation of the design of the app based on the feedback received from the advisory board.

4.3.1 Features Specific to REACH Protocol

This section discusses the features of the app derived from the REACH protocol activities mentioned in chapter 3, section 3.3. When the user selects the app from the Android home screen, a landing page is shown allowing the user to select from 5 available activities (see Figure 3a).

The S.T.I.C. activity is shown in the Figure 3b. In this activity end users are encouraged to do a task they would normally avoid due to their anxiety. In the paper protocol, once a child completes the activity s/he receives a physical stamp from an adult (usually a teacher or parent). In the app this was implemented as a secret code entered by the adult, who could then provide an electronic stamp of approval.

a. Landing Page
b. S.T.I.C.
c. S.T.O.P.
d. Worryheads
e. Daily Diary
f. Relaxation

Figure 3 REACH Protocol App Features

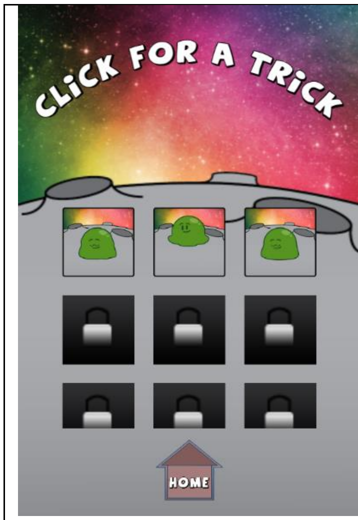
The S.T.O.P. activity (Figure 3c) asks the child to provide responses to a set of questions. Each response is stored in a SQLite database on the device. The figure 3c shows the “O” (Other Thoughts) step of the Worryheads game. This is basically a variant of the S.T.O.P. activity with preselected “S” and “T”s. The child has to consider the given “S” and “T” and select an appropriate “O” and “P” to complete the simulation.

The Daily Diary (Figure 3e) is a scheduled activity available to the child each day. The activity is available during school hours but notifications (Figure 4c) are only given after school hours. As shown in Figure 3e, the Daily Diary asks the child to reflect on potentially anxiety-provoking events from her/his day, and inquires about thoughts that came to mind in that situation. Children also rate how s/he handled and felt about the situation. This embedded diary is part of the organizational framework of REACH emphasizing the need to identify and confront anxiety provoking situations that are threatening but manageable.

The last activity on the landing page is Relaxation (Figure 3f). It consists of 5 relaxation audios that can be played by tapping on the buttons. Media player controls like play, pause and fast forward or reverse are provided while listening to an audio clip. The next section will talk about additional features of REACH app that are not part of the REACH protocol but are provided in the app for facilitating patients, clinicians and researchers.

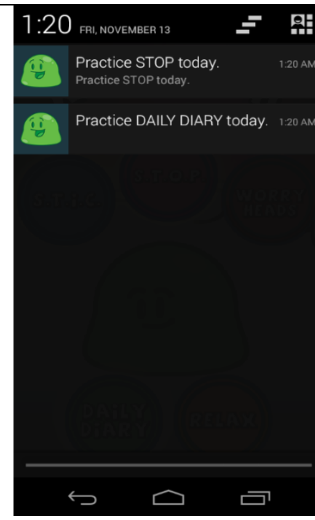
4.3.2 Additional Functions of the REACH App

In addition to the 5 protocol activities available from the landing page, there are additional features of the app that need to be explicitly mentioned (Figure 4). The end user can tap directly on the Blob and be taken to a table-oriented layout of “tricks” that Blob can perform (Figure 4a). The tricks (animations) available at any time are based on the protocol schedule. Further, activities that are overdue are highlighted by a soft gold pulsing glow around the button (Figure 4f) to provide a further visual cue to the end user to perform an activity.



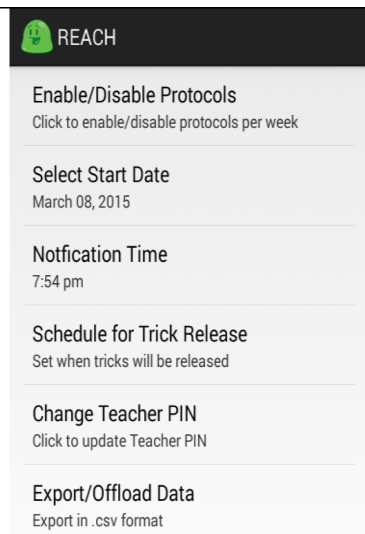
a. Blob Tricks

b. Praise



c. Notifications

d. Progress bars



e. Admin Preferences

f. Glowing Buttons

Figure 4 Reach App Additional Features

Additional features were provided by the app to control the app functions. An on-device database stores all end user responses, and tracks each user action. The REACH protocol schedule is stored in the database by default. There is an android service running in the background that checks the activities done by the patient on a particular day and prompts the patient via notifications to remind them if they have not completed the scheduled activities. Figure 4c shows an example of notifications reminding the patients to practice S.T.O.P. and Daily Diary activities.

Patients can track their progress using a swipe on the landing page. A progress bar screen appears as shown in Figure 4d. As you can see, the progress of all 5 activities is shown in the screen for that particular week. This feature helps the patients to keep a track of their progress during the week and also gives them an option to show their progress to either teachers, clinicians or parents.

Finally, in the face-to-face protocol, interventionists can personalize dosage schedules or tailor training activities during weekly visits. To support this in the app, a hidden admin feature was embedded only for the interventionist role. A specific multi-tap sequence combined with a secret PIN unlocks this feature so interventionists can decide if a protocol component should be enabled/disabled or otherwise modify the planned dosage for that week (Figure 4e). Additional settings include selecting the start date of the protocol, notification time windows and frequency, the schedule trick release, changing the teacher PIN, and exporting data.

The next section will discuss the external validation of the design of the REACH app based on a detailed feedback from an advisory board.

4.3.3 External Validation of REACH App Design

The highly iterative participatory design process described in section 4.2 enabled continuous feedback during app evolution. After completing the initial candidate release version, the design team and psychologists conducted two types of external validation. The first was two feedback sessions with external SMEs from a school advisory board

(SAB). The second was a usability study conducted with actual children as end users in the schools.

The SAB consisted of two school psychologists with experience delivering REACH, and two school district administrators who oversee student services and prevention efforts for 47 K-8 schools. Based on their experience with children, the SAB considered the developmental appropriateness of the design and program tools included (e.g., during the face to face sessions, children wanted to utilize Relaxation and play Worryheads on demand, so those activities were selected for inclusion in the app).

From the SAB feedback, three issues emerged:

1. Safety and security - would participants have access to texting and Internet on the devices?
2. Cost: would parents be responsible for the devices, if lost?
3. Flexibility - would versions of the app be available for the iPhone, smartboards, and tablets?

The first issue was addressed by adding security software SureLock to every device. The second was addressed by applying procedures used by the school relevant to laptop computers where parents are financially responsible. For flexibility, it was determined that preliminary data is necessary prior to investing in additional versions of the technology for different devices.

4.3.4 Preliminary User Study

This section will discuss results of the preliminary user study of the REACH app conducted with target end users of the application.

4.3.4.1 Participants

With parental consent (and assent from child), 22 youth (Mean age = 9.67 years, 12 girls, 12 Hispanic/Latino, 5 White, 1 Black, 1 Asian, 3 "other") from public schools participated in the 'system usefulness, satisfaction, and ease' aspect of this research. In addition, 77% reported knowing how to use an Android smartphone and 54.5% reported playing games using a smartphone "all the time".

4.3.4.2 Measures

System usefulness, satisfaction, and ease were assessed via 22- items from the Usefulness, Satisfaction, and Ease of Use Questionnaire (Lund, 2001) modified for children and adolescents. Youth responded to each item using a 10-point rating scale (1= "not at all" to 10 = "very much"). System ease of use (SYSUSE) was measured via 11 items (e.g., it is easy to use; it is simple to use), quality of support information (INFOQUAL) was measured via 3 items (e.g., instructions and messages are easy to understand; messages to fix problems are clear), system ease of learning (SYSEASE) was measured via 4 items (e.g., I easily remember how to use it; I quickly became good at it), and system satisfaction (SYSSATIS) was measured via 4 items (e.g., I am happy with this app; I would tell a friend about this app). Consistent with the original measure, alpha reliabilities were excellent: system ease of use ($\alpha = 0.92$), quality of support information ($\alpha = 0.83$), system ease of learning ($\alpha = 0.92$), system satisfaction ($\alpha = 0.88$), and stigma ($\alpha = 0.81$) scale scores, and overall usability score ($\alpha = 0.95$).

4.3.4.3 Procedures

Parents (primary caregivers, legal guardians) received a letter from the research team describing the nature of the study and the timeframe for participation (within the next 7 to 10 days). From those contacted, 26% provided child consent and every child provided assent (n=22). Youth with consent/assent provided data at a university laboratory or at their school. At the beginning of the study, each youth was provided with an envelope that contained a device and a questionnaire. After receiving the study materials, three phases (1-Listen to the Relaxation; play Worryheads game; 2-Write a daily-dairy or S.T.O.P. entry; 3-Play with the Blob) were implemented by trained research assistants. For a phase, each prescribed interactions with the app was 2-minutes and responding to the survey lasted about 5 minutes. At the end, youth were thanked for their participation in the study that lasted a total of 20 to 30 minutes.

4.3.4.4 Results

Descriptive statistics and correlations for the focal variables are given in Table 6. There were no missing data and some variables exceeded conventional cutoffs of |2| for skewness and |7| for kurtosis [16]: System Ease of Use (-3.04 skewness, 10.39 kurtosis), System Ease of Learning (-2.15 skewness; 3.9 kurtosis), and System Satisfaction (-2.23 skewness; 4.53 kurtosis). Moreover, statistically significant Shapiro-Wilks test values were found for these indicators and thus subsequent tests were conducted via non-parametric approaches. Specifically, Wilcoxon-Mann-Whitney tests were conducted to estimate any sex (boys vs. girls) or ethnicity/race (Hispanic/Latino vs. Non-Hispanic/Latino) variations in terms of: system ease of use, quality of support information, system ease of learning, and system satisfaction. No statistically significant mean differences were found suggesting robustness across sex and ethnicity/race.

Given these findings, mean estimates for the total sample were calculated and results showed that the REACH app system was highly and positively rated, for the most part, along the four dimensions of interest: system ease of use, quality of support information, system ease of learning, and system satisfaction with means ranging from 8.72 to 9.13. Also, as shown in Table 6, statistically significant correlations were found among the four dimensions with correlation coefficients ranging from .47 to .80 (p <

.05). Lastly, transforming SUSE-Y overall total scores into a traditional "grade" scale, analyses showed that the REACH app system earned an "A" grade from 55% of youth, "A-" from 14%, "B+" from 9%, "B" from 9%, and failing grades of "C-" or less from 13% (or 3 youth). Focusing those youth who rated the system with a "C-" grade or less, data showed that all three youth reported no knowledge of Android operating system. One of the three youth did not know how to connect the earbuds to the phone, had trouble placing earbuds in his ears, asked what he is supposed to press during the Worryheads, asked what the word "respond" means, and did not know what to press during the S.T.O.P. task. Another seemed "lost" during Worryheads and the third youth was distracted by SureLock pop-ups during testing.

	Mean	sd	Median	1	2	3	4
Overall Usability	35.69	19.84	38.23				
1.SYSUSE	8.94	1.48	9.24	--	.61**	.92**	.47*
2.INFOQUAL	9.13	1.28	9.67		--	.80**	.53*
3.SYSEASE	8.72	2.03	9.41			--	.48*
4.SYSSATIS	8.90	1.70	9.75				--

Table 6 Usability Study Results

Note: Ranges from 0 to 40 for Overall Usability, 0 to 10 for other variables; SYSUSE = system ease of use; INFOQUAL = quality of support information; SYSEASE = system ease of learning; SYSSATIS = system satisfaction; *p< .05; **p< .01

In summary, this chapter started off with discussing the multidisciplinary team based design process, followed by implementation details of the REACH app and finally showed the preliminary results of the user study conducted with the REACH app. Even though the study gave positive results, it was limited in scope and it did not provide insight as to what design features affected aspects of compliance and clinical outcomes.

The next chapter will discuss the mixed methods in detail followed by application of these mixed methods in further user studies.

5 MIXED METHODS

The main contribution of this thesis is a mixed methods approach to assess the usability of mHealth applications. This mixed methods approach extends the use of surveys by adding user interaction log analysis to determine compliance, and clickstream analysis to attempt to determine patterns of where users get “off track” (become non-compliant to a design intent). This chapter discusses the mixed methods approach in detail. The first section will cover survey analysis, followed by log analysis and finally clickstream analysis.

5.1 Survey Analysis

Usability studies primarily rely on survey instruments to assess efficacy. Surveys are effective in determining user perception of usability and positive attitudes towards an app. Surveys are used in the traditional way, primarily to measure a user’s perception of the usability of the mHealth app. This research also explores the customizations of existing survey instruments tailored to get user feedback about design features in an app.

Participants of the user study conducted in this research were provided with a survey questionnaire during the REACH sessions to provide feedback. The statements were framed using Lund's 2001 USE framework (Lund, 2001). Answers given by the youth during these surveys were measured using a scale of 1 to 10, 1 indicating that the youth strongly disagrees with the statement, 10 indicating that youth strongly agrees with the statement. This data is referred to as survey data in this research.

An example of survey data is shown in Table 7. The first column shows the statements which are tailored towards satisfaction, ease of use and usefulness aspect of the app. The columns on the right give a range of options to choose from as an answer to the statement.

	Not at all	Somewhat								Very much
1. I like the color of the buttons.	1	2	3	4	5	6	7	8	9	10
2. The space themed background looks good.	1	2	3	4	5	6	7	8	9	10
3. I like the smiling green blob.	1	2	3	4	5	6	7	8	9	10
4. I was able to navigate easily in the app.	1	2	3	4	5	6	7	8	9	10
5. In a few steps I can do what I want.	1	2	3	4	5	6	7	8	9	10

Table 7 Sample Survey and Answer Format

This is the traditional use of surveys to assess whether user thinks positive about the app or not. In chapter 4 section 4.3.4, the use of surveys for assessing the usability of the REACH app was explained.

The second aspect of using survey data in this research is tailoring survey statements towards particular design features of the application. This approach helps to get a detailed feedback from the user and to identify which of the design features affected the app usage. An example of tailored survey questionnaire with respect to specific design features is given in Table 8. The statements given in this table are tailored towards Blob tricks which is a design feature in the REACH app (chapter 4, section 4.2.3.2).

	Not at all	Somewhat								Very much
1. I liked playing the Blob's tricks.	1	2	3	4	5	6	7	8	**9**	10
2. The Blob doing tricks made me work harder to get more tricks	1	2	3	4	5	6	7	**8**	9	10
3. I liked the sounds the Blob makes	1	2	3	4	5	6	**7**	8	9	10
4. I liked the way Blob moves around.	1	2	3	4	5	6	7	8	**9**	10
5. I liked the Blob saying "Good Job!".	1	2	3	4	5	6	7	**8**	9	10
6. The Blob telling me that I did a good job made me try it again	1	2	3	4	5	6	7	**8**	9	10
7. Getting a Thumbs Up from the Blob made me work harder.	1	2	3	4	5	6	**7**	8	9	10
8. The Blob telling me that I did a good job made me try it again	1	2	3	4	5	6	**7**	8	9	10
9. Getting a Thumbs Up from the Blob made me work harder.	1	2	3	4	5	**6**	7	8	9	10

Table 8 Specific Survey Questions Based on Design Features

There are 9 statements in Table 8. The highlighted responses denote a score given by user for each statement. The mean of highlighted scores is 7.66. Thus for Blob tricks, the mean usability score is 7.66. The higher usability score indicates that the user feels positive about the particular design feature. In this way, surveys are tailored to get user's opinion about specific design features in the REACH app. The survey data consisting of answers given by users is considered for calculating a mean score per category of the design feature.

5.2 Log Analysis

Log analysis is based on the concept of log data. Log data is a form of data representing interactions of the user with the mobile application. Log analysis in this research is used for measuring user task completion. Compliance is measured using log analysis based on the notion of actual work versus expected work done by the user. For example, if a youth was asked to complete the Daily Diary activity that consists of four steps, log data will give a clear indication if the user actually completed all the four steps of Daily Diary (shows compliance) or the youth did not complete any of the steps in the activity or the youth partially completed the activity (completed less than four steps and was non-compliant). *Activity Attempted* is considered as user started the activity, whereas *Activity Completed* means user completed all the steps inside the activity. Based on this criteria, the user is either compliant or non-compliant. The model consists of a set of activities that are used as a reference (ideal case) for the analysis of the actual data (Table 9).

Activity	**Expected work in the activity**	**Completion criteria**	**Measure of compliance**
Relaxation	Play Intro file	0 (did not play) /1 (did play)	*Nominal Measure*
WorryHeads	S, T, O, P steps one after the other in order.	4 sequential steps to complete S.T.O.P. activity once.	*Ratio Measure*: # of attempts vs # of completions
Daily Diary (DD) OR S.T.O.P.	Either DD (Once) , or S.T.O.P. (Once)	4 sequential steps to complete activity once.	*Ratio Measure*: # of attempts vs # of completions
Blob Tricks	Open blob tricks, complete zero or more blob tricks.	12 tricks.	*Interval Measure*: # of tricks opened

Table 9 Ideal Set of Activities

5.2.1 Measures

There are three levels of measurement used in this analysis (De Land et al., 1990). First is nominal measurement, wherein one simply names or categorizes responses. In this case, Relaxation log data is straightforward. The user either completed listening to the entire audio or he/she did not complete listening. This can be measured using a binary 0/1 nominal measure, where 1 represents user completing the activity and 0 represents user not completing the activity. If they get a binary score of 1, then he/she is said to be compliant with the Relaxation activity.

The second type of measurement is a ratio measure. The ratio scale of measurement is the most informative scale. It is a scale with the additional property that its zero position indicates the absence of the quantity being measured. Ratio measure can be used to make proportional comparisons. For example, if someone is 25% compliant and someone else is 50% compliant then the latter is “twice” as compliant – which we can only say if it is a ratio measure. The log data for Worryheads, S.T.O.P. and Daily Diary is measured using this measure. Number of activity attempts vs the number of activities completed is calculated for every user. For example, if the user attempts

Daily Diary two times and completes only once, then the percent completion for that user for Daily Diary is 50%. Similarly, S.T.O.P. and Worryheads compliance are measure by checking the number of attempts versus the number of completions.

The third type of measurement is interval measure. Interval scales are numerical scales in which intervals have the same interpretation throughout. There are 12 blob tricks in the app. Based on the log data each user gets a score that represents the number of tricks played by the user.

5.3 Clickstream Analysis

Clickstream is a concept from web analytics that deals with analyzing how users are using a particular website (Taniguchi, D., 2004). Typically in a client-server architecture, web loggers are used to track and store the data of mouse clicks on the webpages. This log data is analyzed to find specific usage patterns of the users to make future marketing decisions and attract them to the websites. This concept is called clickstream analysis (sometimes called web analytics). Similar to that of web analytics, clickstream analysis of log data is performed in this research to identify usage patterns of the users of the app. These patterns provide insights about compliance and non-compliance measured during log analysis.

The method used in clickstream analysis in this thesis follows a recent paper by Lettner and colleagues (Lettner et al., 2014) that studies design intent versus user task completion. This is an emerging area focusing on identifying intentions of users when they launch an application on their smartphone, and understanding the tasks they actually execute. The paper describes using user interaction sequences (clickstreams) in mobile apps to understand the actual completion of a task by users against reference (designed) tasks. The authors discuss a novel approach of automatically extracting and grouping user sequences against predefined tasks and presenting them visually. They also present alternatives for user interaction metrics based on algorithms from bioinformatics research. This chapter summarizes my approach for adapting these techniques for the purposes of connecting design features to compliance results.

5.3.1 Method

The concepts discussed in the paper (Lettner et al., 2014) such as data collection, pattern generation, reference string pattern versus actual string pattern, and interaction counting are applied in this thesis. This section summarizes the adaptation of the approach by these authors to my research context. In the following paragraphs "the paper" refers to (Lettner et al., 2014).

The paper defines a *state* as a certain visual representation in the mobile app (e.g. page or screen), that is valid under a defined context. Each state is augmented with time spent by users in the state and the number of interactions that take place on the associated screen. Based on the interaction sequences (clickstream), string patterns are generated for each user interaction session. Preprocessing consists of removing cycles in the patterns (based on the assumption that cycles in the string pattern represent the same contextual meaning) and merging similar patterns together to fit into one category.

The clickstream analysis method in this thesis follows a similar approach for data collection and pattern generation from this paper. Log data for each session of the REACH protocol (i.e. between each start and stop of an application) is collected for every user by recording each swipe and tap on the screen. This log data is taken as an input by a parser written in Java. The parser parses this data based on the concept of states. A state is defined as an activity in the REACH app and it is augmented by number of interactions that take place inside a particular activity. The parser generates string patterns of the log data for every user. Though there are similarities in data collection and pattern generation with respect to Lettner's method, there are considerable differences with respect to data preprocessing and sorting of patterns that are described in the next subsections.

5.3.2 Data Collection

Log data for each session of the REACH protocol is collected for every user by recording each swipe and tap on the screen. This data is parsed using a parser written in Java. The parser parses this data based on the concept of states. A state is defined as a unique screen in an activity of the REACH app. For activities like *Relaxation* and *Blob Tricks*, there is only one screen inside the activity and hence can be represented as a single letter (as shown in the Table 10). There are 12 Blob tricks in the REACH app but they are still considered as one state because each state corresponds to the same action of "*playing a trick*". Similarly, there are 5 audio files in Relaxation activity but they are still represented as a single state because the activity corresponds to a single action of "*playing relaxation audio*". For activities involving multiple screens, each screen corresponds to a state. This is because each state in the activity has its own significance in the activity. For example, Daily Diary activity has 4 screens that are represented as four different states (as shown in Table 10).

The parser is written in Java. It takes log data in CSV (comma separated value) format as an input. It uses the states given in Table 10 as a reference to parse the log data and generates a string pattern for each user. Following (Table 11) is an example of the log data and corresponding step by step pattern generated by the parser for this data.

State Symbol	State (activity in the app)
R	Relaxation
D1	Daily Diary Step 1
D2	Daily Diary Step 2
D3	Daily Diary Step 3
D4	Daily Diary Step 4
S1	S.T.O.P. Step 1
S2	S.T.O.P. Step 2
S3	S.T.O.P. Step 3
S4	S.T.O.P. Step 4
W1	Worryheads Step 1
W2	Worryheads Step 2
W3	Worryheads Step 3
W4	Worryheads Step 4
B	Blob Trick
STIC1	STIC started
STIC2	STIC ended

Table 10 State Symbol and Corresponding State in the REACH App

Log Data	Corresponding state generated by parser
RELAXATION_INTRO	R
WORRY_HEADS	
WORRY_HEADS_NEXT_CLICKED	W1
WORRY_HEADS_NEXT_CLICKED	W2
WORRY_HEADS_O_RIGHT	W3
WORRY_HEADS_COMPLETED	W4
DAILY_DIARY	
DAILY_DIARY_STATE_ONE_NEXT_CLICKED	D1
DAILY_DIARY_STATE_TWO_NEXT_CLICKED	D2
DAILY_DIARY_STATE_THREE_NEXT_CLICKED	D3
DAILY_DIARY_COMPLETED	D4
BLOB_TRICK_TWELVE	
BLOB_TRICK_COMPLETE	B

Table 11 Sample Log Data and Generated Pattern

The corresponding pattern generated by the parser for the log data is "$RW_1W_2W_3W_4D_1D_2D_3D_4B$". In this way user's interaction for a particular session can be represented using a pattern.

5.3.3 Pattern Grouping

After converting user clickstreams into respective patterns, grouping can be performed and categories of these patterns can be formed. The patterns generated by the parser can be categorized based on the concept of similarity estimation. This approach is similar to that of Lettner's approach of similarity estimation (Lettner et al., 2014). It compares reference string patterns with the actual string patterns and calculates the similarity score for each generated string with respect to the reference string. Lettner mentioned three methods of calculating similarity measures. The first method of similarity score calculation was by Levensthein (LV) that calculates the number of editing steps (insert, delete or replace) required in passing from one string to the other. The common drawback of this method is that this method considers all edit operations as equally expensive, and does not consider bulk operations (combinations of the same subsequences). This assumption, according to Lettner, does not work for longer sequence patterns. Lettner talks about two more algorithms for calculating similarity scores. Algorithms by Needlemen-Wunsch (Needleman et al., 1970) (NW) and Smith-Waterman (Smith et al., 1981) (SW). Both propose a dynamic programming approach for calculating similarity scores. These algorithms do not consider edit operations as equally expensive, rather they consider each group deletion as one edit operation. Needlemen-Wunsch method is used for global alignments of strings. Global similarities deal with comparing two string in their entirety and calculating similarity score for the entire length of the reference string. On the contrary Smith-Waterman method is used for calculating local similarities between two strings. Local string similarity means, instead of considering entire strings, it compares segments of all possible lengths of the reference and actual string. The choice of which method to choose largely depends on the nature of string patterns and the analysis being done on the strings.

The method to calculate similarity score is based on the Needlemen-Wunsch (NW) algorithm. The reason of choosing NW is due to the nature of the string comparison. As the NW method calculates global similarity scores, this is most relevant in comparing reference versus actual string patterns. The method used in this thesis to calculate the similarity score is explained in detail in Appendix B.

$$sim_{NW}(a, b) = 1 - (N / (max(len\ a, len\ b)) \quad\ldots\ldots(1)$$

Equation 1 written above calculates the similarity score between the strings *a* and *b* where *a* is a reference string and *b* is the actual string. *N* is the number of operations performed on the string *b*, to convert it to string *a*. The calculation of *N* is explained in detail in Appendix B. len *a* and len *b* give the length of each string. Based on this equation to calculate similarity score, there can be different scenarios where we get very high similarity scores (close to 1 or 1), average or low similarity scores (0.2 to 0.7) and negative or zero similarity scores. These three ranges of similarity scores are used to judge the usage of the app. High similarity score indicates compliance of users to the REACH protocol, average similarity scores indicate non-compliance or partial compliance of users to the REACH protocol and negative or zero similarity scores indicate an ad hoc behavior of users which in turn is a type of non-compliance.

5.3.3.1 Compliance Pattern

The pattern that provides evidence of user completing all the steps inside an activity is considered as a compliance pattern. Let us consider Daily Diary as a reference task. The reference string pattern for this task is *D1D2D3D4*. Let us consider actual pattern observed using log data as *D1D2D2D2D2D2D2D3D3D3D4*. According to NW, the similarity score for these two string sequences is calculated by considering group deletions or single insertions or single replacements from actual pattern.

Reference String Pattern	D1	D2	-	-	-	-	-	D3	-	-	D4
Actual String Pattern	D1	D2	**D2**	**D2**	**D2**	**D2**	**D2**	D3	**D3**	**D3**	D4
Operation			**Deletion**						**Deletion**		

Table 12 Global Optimal Alignment using Needlemen-Wunsch method

Table 12 shows the global optimal alignment of two strings using Needleman-Wunsch method. The group deletion in this case is highlighted in the second row (in bold). After deleting this highlighted substring, we get the reference string of *D1D2D3D4*. Thus the similarity score is calculated as given in the equation 2. The 0.68 similarity score represents that the user took a lot of extra steps but completed the intended task.

$sim_{NW}\ (D1D2D3D4,\ D1D2\mathbf{D2D2D2D2D2}D3\mathbf{D3D3}D4) = 1 - (3.5/\max(4,11)) = 0.68$..(2)

Looking at the pattern *D1D2**D2D2D2D2D2**D3**D3D3**D4,* it can be observed that the user struggled in state D2. The design of the Daily Diary activity as shown in the Figure 5, makes it mandatory for the user to provide the input. This may have been one of the reasons for the struggle observed in the actual pattern.

Figure 5 Daily Diary Feedback Message

5.3.3.2 Non-compliance Pattern

A non-compliance pattern deals with a clickstream pattern providing evidence of user not completing the activities prescribed by the REACH protocol. Another example of similarity estimation of strings can be given that demonstrates user not completing the intended task. Consider a reference task of Worryheads. The reference string pattern for this task is *W1W2W3W4*. Let us consider the actual pattern string as *W1W2W2W2W2*. In this case, the actual pattern is missing 2 states and has W2 more than 1 time (*W1W2**W2W2W2***) that results in 2 replacement and 1 deletion operation as shown in Table 13. The pattern clearly indicates that the user did not complete the steps as per the reference task and was non-compliant.

Reference String Pattern	W1	W2	W3	W4	-
Actual String Pattern	W1	W2	**W2 Replace**	**W2 Replace**	**W2 Delete**

Table 13 Sample Replace and Delete Operations based on Global Alignment

sim_{NW} (*W1W2W3W4*, W1W2W2W2W2) = 1 – (2.5/max(4,5)) = 0.5 ....(3)

5.3.3.3 Ad hoc Behavior Pattern

The third category patterns are the strings representing *ad hoc* interaction sequences. Ad hoc interaction means that the user completed an entirely different activity in the app compared to what was asked from the user. For example, consider the reference task of S.T.O.P. The reference string pattern for this task is *S1S2S3S4*. Let the actual pattern string be *RD1D2D3*. In this case, there will be 4 replacements because the actual string is entirely different from the reference string (as shown in Table 14). The 0 similarity score calculated in Equation 4 clearly indicates that the two strings had nothing in common and represent an ad hoc behavior by the user.

Reference String Pattern	S1	S2	S3	S4
Actual String Pattern	**R Replace**	**D1 Replace**	**D2 Replace**	**D3 Replace**

Table 14 Global Alignment for an Ad Hoc Pattern

sim_{NW} (*S1S2S3S4*, *RD1D2D3D3*) = 1 – (4 / max(4,4)) = 0 ....(4)

These three categories of similarity scores and the patterns of usages they represent are the basis of clickstream analysis in this research. The next section will discuss the relationship between these three methods and how they inform each other on predicting the impact of design features on protocol compliance.

5.4 Relationship between the Methods

Each of the three methods discussed in previous sections provide information that correlates to the information provided by other two. Surveys can be used to get user's perception of the usability of the app as well as positive attitudes towards particular design features of the app (as discussed in section 5.1). Log analysis provides numerical details of the compliance of patients to the protocol activities. The first relationship exists between survey analysis and log analysis.

Let us consider that in a hypothetical user study conducted with the REACH app, 10 subjects participated and 40% compliance was observed for a Daily Diary activity (only 4 subjects completed the Daily Diary activity). Participants were also asked to fill out a survey tailored to get their perception of the design of the Daily Diary activity. Table 15 shows the survey questionnaire tailored towards Daily Diary activity.

	Not at all	Somewhat								Very much
1. It was easy to follow the steps in the Daily Diary.	1	2	3	**4**	5	6	7	8	9	10
2. The buttons made it easy to do the work.	1	2	3	**4**	5	6	7	8	9	10
3. Using the keyboard on the app was easy.	1	2	3	4	**5**	6	7	8	9	10
4. Using the microphone on the app was easy.	1	2	3	4	**5**	6	7	8	9	10
5. Using the numbers on the app to rate situations in the Daily Diary was easy.	1	2	3	4	**5**	6	7	8	9	10

Table 15 Survey Tailored towards the Design of Daily Diary

The user responses of the non-compliant users given in bold indicate that the users do not seem positive about the design of the Daily Diary. The average score of the responses is 4.6 that reinforces the results observed in log analysis. This shows a relationship between the non-compliance observed in log analysis and survey analysis. This relationship can be further extended by doing a clickstream analysis of the patterns of usages for non-compliant youth. The reference pattern for Daily Diary is D1D2D3D4. Let us consider that the actual patterns generated by parser for these non-compliant youth were of the sort D1D1D1D2, D1D1D1, D1D2D3D3D3, and RB. These patterns have similarity scores as follows.

$$sim_{NW}\ (D1D2D3D4,\ D1D1D1D2) = 1 - (3 / \max(4,4)) = 0.25 \quad \text{(4.1)}$$

$$sim_{NW}\ (D1D2D3D4,\ D1D1D1) = 1 - (3 / \max(4,4)) = 0.25 \quad \text{(4.2)}$$

$$sim_{NW}\ (D1D2D3D4,\ D1D2D3D3D3) = 1 - (1.5 / \max(4,5)) = 0.7 \quad \text{(4.3)}$$

$$sim_{NW}\ (D1D2D3D4,\ RB) = 1 - (4 / \max(4,4)) = 0 \quad \text{(4.4)}$$

Low similarity score indicate that the designed task was not completed in the manner intended. The patterns observed for the first three patterns indicate the struggle to complete the steps in Daily Diary activity. The pattern RB shows an ad hoc behavior by the user.

In summary, this chapter explained in detail the three usability validation methods. These methods when used together will give a better understanding of impact of design features of the REACH app on protocol compliance. The next chapter will discuss experimental user studies and validation using mixed methods approach.

6 EXPERIMENTAL STUDIES

This chapter will discuss the results of user studies conducted in public schools. The preliminary user study discussed in chapter 4, section 4.3.4 was a pilot study with the REACH app. Though the study showed positive attitudes of subjects towards the REACH app, it did not measure app feature usage. Two user studies were conducted to measure compliance of users to the REACH protocol. The first user study was a single REACH session based study whereas the second user study was a six week (full length of REACH protocol) based study.

In each of these two studies the usability validation is performed using the mixed methods explained in section 5.3. Log analysis gives the numerical compliance measure of subjects participating in the user studies. This quantified data answers the first research question (chapter 3, section 3.1). Clickstream analysis along with survey analysis provides a causal connection for the non-compliance measures calculated during log analysis thereby answering the second research question (chapter 3, section 3.2).

6.1 User Study 1

With Institutional Review Board approval (Appendix A), a total of 390 parents (primary caregivers, legal guardians) received a letter explaining the nature of this research and the two-week timeline for participation. From those contacted, 34% provided child consent and every child with parent consent provided assent (n = 132) to participate in the usability trial. These rates of consent reflect that the entire study (i.e., send recruitment letter, receive consent/assent, conduct usability evaluation) was scheduled to start and end over the course of two weeks, prior to summer vacation. Children with consent/assent were escorted by a school liaison to a classroom where usability evaluation procedures were implemented by three trained research assistants;

providers assembled at a classroom or office for the study. Usability evaluation activities with children were conducted in a group format. Participants were given an envelope containing a smartphone device preloaded with the REACH app and a questionnaire. Instructions and usability items were read aloud. Participants were directed to: (1) listen to the Relaxation mp3; (2) play the Worryheads game, (3) respond to part 1 of the survey, (4) write a Diary or S.T.O.P. entry, (5) respond to part 2 of the survey, (6) interact with the blob, and (7) respond to part 3 of the survey. Procedures 1, 2, 4, and 6 lasted 2 minutes each while responding to survey items was not timed; each such session of the procedure lasted 20 to 30 minutes.

Compliance criteria for this study was completing activities as instructed by the psychologist during the session. The youth was said to be compliant with the session if they complete all the activities (*Relaxation, WorryHeads, Daily Diary* or *S.T.O.P.,* and *Blob tricks*). Log data was collected for the youth who participated in the study. For each activity, the results from log analysis and clickstream analysis are discussed in the subsequent sections. The discussion will start with the survey analysis using survey responses provided by youth at the end of the study.

6.1.1 Survey Analysis

This section consists of the survey analysis that is derived from a usability study performed by Ryan Stoll and colleagues (Stoll et al., Unpublished manuscript, 2016). System usefulness, satisfaction, and ease were assessed via 22 items from the Usefulness, Satisfaction, and Ease of Use Questionnaire (Lund, 2001) modified for children and adolescents. Youth responded to each item using a 10-point rating scale (1= “not at all” to 10 = “very much”). System ease of use (SYSUSE) was measured via 11 items (e.g., it is easy to use; it is simple to use), quality of support information (INFOQUAL) was measured via 3 items (e.g., instructions and messages are easy to understand; messages to fix problems are clear), system ease of learning (SYSEASE) was measured via 4 items (e.g., I easily remember how to use it; I quickly became good at it), and system satisfaction (SYSSATIS) was measured via 4 items (e.g., I am happy with this app; I would tell a friend about this app).

A total of 132 youth from public schools participated in the present study. Youth ages ranged from 8 to 12 years old (M = 9.65, SD = 0.82), 63% were female and 29% were Hispanic/Latino (32% White; 10% African American/Black; 5% Asian/Pacific Islander; 24% Native American or mixed ethnicity/race). The app was highly and positively rated on overall usability (*M* = 33.30 out of 40, *SD* = 5.88) and each usability dimension (possible range is 0 to 10): system ease of use (*M* = 8.57, *SD* = 1.53), quality of support information (*M* = 8.99, *SD* = 1.52), system ease of learning (*M* =8.96, *SD* = 1.72), and system satisfaction (*M* = 9.18, *SD* = 1.47).

Pragmatically, overall usability scores were transformed into a traditional "grade" scale and showed that the REACH app earned an "A+" grade from 7% of youth, "A" from 27%, "A-" from 14% "B+" from 8%, "B" from 5%, and failing grades of "C-" or less from 17% (or 23 youth). Focusing on youths who rated the app with "C-" or less, 10 youth encountered one or more software, hardware, and/or user knowledge errors during the testing protocol. Of those, 3 youth encountered software errors, 3 hardware error, and 4 user knowledge errors. Software errors included: app suddenly quitting in the middle of use (2 youth) and extraneous notifications or pop ups interfering with using the app (1 youth). Hardware errors included: Android smartphone restarting in the middle of use (2 youth) and headphone jack of smartphone not working properly (1 youth). User knowledge errors included: users having difficulty finding correct buttons or activities within the app (3 youth), users having no knowledge of the Android operating system (4 youth), and users couldn't turn on or unlock the Android smartphone device (2 youth).

The survey analysis presented in this study was similar to the survey analysis presented in the preliminary study and it measured the user perception of the usability of the app, however it did not provide enough details to measure the compliance of users to the REACH protocol. For getting numerical measurement of the compliance, a much more detailed data was considered in this user study in the form of user log data. The next subsections will discuss log analysis as well as clickstream analysis of the log data with respect to activities in the REACH session.

6.1.2 Relaxation

The compliance criteria for *Relaxation* was listening to the audio file completely. The results of log analysis showed that 123 out of 132 youth completed listening to the audio file. The remaining 9 youth did not listen to the audio file completely or they were busy doing something else in the app. The reason of non-compliance by these 9 youth can be found out by clickstream analysis.

The reference string pattern for this task is just one state, *R,* as there are no steps inside *Relaxation* activity. From the total of 132 youth participating in the study, 123 youth showed the actual string pattern as *R,* after parsing the log data. Hence the similarity score was equal to 1 (equation 5) as there was no conversion needed between actual and reference pattern. These 123 youth were compliant with the *Relaxation* activity.

$sim_{NW}(R, R) = 1 – (0 / max(1,1)) = 1$ ……………………………………………………………………(5)

The remaining 9 youth started with Blob Tricks instead of *Relaxation*, so the actual string pattern generated from the log data was *B* and the similarity score was 0 as one replace operation was required to convert actual pattern *B* to reference pattern *R* (equation 6). This shows that these 9 youth played with Blob Tricks when they were instructed to complete Relaxation activity. So the reason of non-compliance here is youth inclining towards Blob Tricks.

sim_{NW} $(R, B) = 1 - (1 / \max(1,1)) = 0$ ……………………………………………………………….(6)

6.1.3 Worryheads

The compliance criteria for Worryheads was completing all four steps inside the activity. Log analysis was performed on the Worryheads log data for N=132 youth. Number of attempts made by the youth and the number of completions are calculated for every youth. For example, if the youth has attempted Worryheads 4 times and completed only 3 times, then the percentage completion is 75. Out of 132 youth, 121 youth completed the Worryheads activity when they attempted it. The remaining 11 youth did not complete the Worryheads activity. The reason behind the non-compliance can be figured out using clickstream analysis.

The reference string pattern for this task was *W1W2W3W4*. From 132 youth, 121 youth showed the actual string pattern as *W1W2W3W4*. Hence the similarity score was 1 for these 121 youth (equation 7). These 121 youth were compliant with the Worryheads activity.

sim_{NW} $(W1W2W3W4, W1W2W3W4) = 1 - (0 / \max(4,4)) = 1$ ……………………………….(7)

The remaining 11 youth had the actual string patterns of *W1, W1W2* and *W1W2W3*. The similarity scores for these three patterns are 0.25, 0.50 and 0.75 respectively (equations 8,9 and 10). These three similarity scores represent 11 youth who were non-compliant. There was no external factor observed in the actual string patterns that may have been a cause of non-compliance. The similarity scores show the types of interactions of these 11 youth who backed out of the Worryheads activity from state W3, state W2 and state W1.

sim_{NW} ($W1W2W3W4$, $W1$) = 1 – (3 / max(4,1)) = 0.25 ....(8)

sim_{NW} ($W1W2W3W4$, $W1W2$) = 1 – (2 / max(4,4)) = 0.5 ....(9)

sim_{NW} ($W1W2W3W4$, $W1W2W3$) = 1 – (1 / max(4,4)) = 0.75 ....(10)

The high compliance observed in Relaxation and Worryheads was reflected in survey analysis as well.

Survey Statement	Average user response for compliant users	Average user response for compliant users
I can use the app without written instructions.	8.8	7
In a few steps the app does what I want.	8.7	7.1
People using it once or many times would like it.	9	7.5
It is easy to understand.	8.9	7.4
I can use it well every time.	9	7.8
Using it requires no effort.	8.8	7.1

Table 16 Survey Responses for Compliant Users of Worryheads and Relaxation

The users had to respond to a survey after completing Relaxation and Worryheads. Users reported high responses for the statements given in Table 16. The average response recorded for the compliant users was 8.86 out of possible 10 points. The average response for non-compliant users was 7.31 out of 10 possible points. These findings from survey analysis corroborate findings from log analysis and clickstream analysis.

6.1.4 Daily Diary

The compliance criteria for Daily Diary was completing all four steps inside the activity. Log data of the user study showed that out of 132 youth, 111 attempted Daily Diary. Out of these 111 youth, only 60 youth successfully completed DD activity. 51 youth were not compliant because they did not complete all the steps of DD. The reasons behind 51 non-compliant youth can be figured from the clickstream analysis of the log data.

The reference string pattern for DD activity was *D1D2D3D4*. Only 43 out of 111 youth showed the actual string pattern of their interaction as *D1D2D3D4*. Hence the similarity score was 1. 17 youth out of 111 showed actual string patterns as *D1D2D3D2D1D2D3D3D3D4*, *D1D1D1D1D1D2D3D4*, *D1D1D1D2D3D2D2D3D3D4,* or *D1D1D2D3D4*. The highlighted substrings in the actual patterns indicate the group deletions performed during similarity calculations. The similarity scores for these patterns are as follows.

Reference String Pattern	D1	D2	D3	-	-	-	-	-	-	D4
Actual String Pattern	D1	D2	D3	**D2**	**D1**	**D2**	**D3**	**D3**	**D3**	D4
				Delete						

Table 17 Global Alignment for the String Pattern

sim_{NW} (*D1D2D3D4*, *D1D2D3***D2D1D2D3D3D3***D4*) = 1 – (3 / max(4,10)) = 0.70 ......(11)

As shown in Table 17, the global alignment shows a gap between reference and actual pattern. This gap can be removed using a group deletion operation as shown in the table. Similarly global alignment is performed for other string patterns and similarity scores are calculated.

sim_{NW} (*D1D2D3D4*, *D1***D1D1D1D1***D2D3D4*) = 1 – (2 / max(4,8)) = 0.75 ....(12)

sim_{NW} (*D1D2D3D4*, *D1***D1D1***D2D3***D2D2D3D3***D4*) = 1 – (3 / max(4,10)) = 0.70 .....(13)

sim_{NW} (*D1D2D3D4*, *D1***D1***D2D3D4*) = 1 – (0.5 / max(4,5)) = 0.90 ....(14)

These four similarity scores generated for 17 youth fall in the range of 0.7 to 1 and are very similar to the reference pattern. The similarity score 1 indicates the perfect completion of the activity. The similarity scores in the range 0.70 to 1 (1 exclusive) shows that these youth struggled in between the steps but at the end still managed to complete the activity. This behavior can be explained by looking at the design of the DD activity in the REACH app. In the design of the DD activity, a feedback message is displayed to the user if he/she tries to go directly to next step without completing the previous step. This feedback message tells the user to respond to the question asked in the step (Figure 5). These 17 youth may have understood this feedback from the app and completed the respective steps in the Daily Diary activity.

The remaining 51 youth out of 111 did not complete the Daily Diary activity. The patterns of actual usages and the similarity scores for these patterns are as follows. The patterns generated by the parser for non-compliant youth are *D1D1D1D1D1D1D1D2*, *D1D1D2D3D3*, and *D1D1D1D1D1D1*. The highlighted substrings in the equations indicate the operations performed during similarity calculations. The similarity scores for these patterns are as follows.

Reference String Pattern	D1	D2	D3	D4	-	-	-	-
Actual String Pattern	D1	**D1**	**D1**	**D1**	**D1**	**D1**	**D1**	**D2**
Operation		Replace	Replace	Replace	Group Deletion			

Table 18 Global Alignment for an Ad Hoc Pattern for Daily Diary

sim_{NW} (*D1D2D3D4*, *D1****D1D1D1D1D1D1D2***) = 1 – (5 / max(4,8)) = 0.375 ....(15)

Table 18 shows the global alignment and the operations that are performed on the actual string to convert it to the reference string. Equation 15 calculates the similarity score for the pattern *D1****D1D1D1D1D1D1D2***.

sim_{NW} (*D1D2D3D4*, *D1****D1****D2D3****D3***) = 1 – (1.5 / max(4,5)) = 0.7 ....(16)

Equation 16 represents a pattern *D1**D1**D2**D3**D3*. One deletion and one replacement is required to match the reference pattern.

sim_{NW} (*D1D2D3D4*, *D1**D1D1D1D1D1***) = 1 – (4 / max(4,6)) = 0.33 ....(17)

Equation 17 represents a pattern *D1**D1D1D1D1D1***. Three replacements and one group deletion is required to match the reference pattern. Thus, similarity score is 0.33.

Actual string patterns of non-compliance of Daily Diary and their respective similarity scores both indicate the struggle of youth in going from state 1 to state 2 of Daily Diary. In reviewing the design of the Daily Diary activity, it is clear that youth struggle to give the input between the steps even with the feedback messages suggesting the users to respond in the respective steps (Figure 5). Youth keeps tapping on the NEXT button inside the Daily Diary activity without providing the input in state 1 which is reflected clearly in the patterns.

The survey analysis of Daily Diary and S.T.O.P. users is consolidated in the section 6.1.5 as users were given a choice to choose between the two activities during the session.

6.1.5 S.T.O.P.

The compliance criteria for S.T.O.P. was completing all four steps inside the activity. According to log analysis, out 132 youth, only 53 attempted S.T.O.P. activity. From these 53 youth only 20 youth completed all the four steps of S.T.O.P. The remaining 33 youth were non-compliant. The reason for their non-compliance can be explained by clickstream analysis of the log data.

The reference string pattern for S.T.O.P. was *S1S2S3S4*. Out of 53 youth who attempted S.T.O.P., only 20 youth completed the activity successfully. Out of these 20 youth, only 4 youth completed the S.T.O.P. with a perfect string pattern of *S1S2S3S4*. The actual string patterns generated by the parser for the remaining 16 youth were *S1S2S2S2S2S3S4*, *S1S1S1S1S2S2S2S3S4* and *S1S2S2S3S2S3S4*.

Reference String Pattern	S1	S2	-	-	-	S3	S4
Actual String Pattern	S1	S2	**S2**	**S2**	**S2**	S3	S4
Operation			Group Delete				

Table 19 Global alignment for a S.T.O.P. Pattern

sim_{NW} (*S1S2S3S4*, *S1S2**S2S2S2**S3S4*) = 1 – (1.5 / max(4,7)) = 0.78 ....(18)

Table 19 shows the global alignment of actual string pattern using NW method. Equation 18 represents the similarity score. One group deletion (**S2S2S2**) is required to match the reference pattern. The similarity score is 0.78.

sim_{NW} (*S1S2S3S4*, *S1**S1S1S1**S2**S2S2**S3S4*) = 1 – (2.5 / max(4,9)) = 0.72 ....(19)

Equation 19 represents a pattern *S1**S1S1S1**S2**S2S2**S3S4*. Two group deletions are required to match the reference pattern. The similarity score is 0.72.

sim_{NW} (*S1S2S3S4*, *S1S2**S2**S3**S2S3**S4*) = 1 – (1.5 / max(4,7)) = 0.78 ....(20)

Equation 20 represents a pattern *S1S2**S2**S3**S2S3**S4*. Two deletions are required to match the reference pattern. Thus, the similarity score is 0.78.

The high similarity scores indicate that the youth are compliant with the S.T.O.P. activity. Even if the youth seem to get stuck in the steps of the activity, they manage to complete the activity eventually. This is similar to the patterns observed in Daily Diary. Similarly, there are feedback messages provided to the youth if they keep tapping on NEXT button in the app. These feedback messages tell the youth to respond and provide input in the respective step. These 16 youth seem to have understood this message and ended up completing each step and being compliant.

Out of 53 youth who attempted S.T.O.P., 33 youth were non-compliant. The actual string patterns for these youth were, *S1S1S1S2S1S1, S1S1S1S1S1S1S1S1S1, S1S1S2S2S3S3S2S1* and *S1S1S2S2S2S2S2S2*.

Reference String Pattern	S1	-	-	S2	S3	S4
Actual String Pattern	S1	**S1**	**S1**	**S2**	**S1**	**S1**
Operation		**Delete**			**Replace**	**Replace**

Table 20 Global alignment for S.T.O.P. Non-compliant patterns

sim_{NW} (*S1S2S3S4, S1**S1S1**S2**S1S1***) = 1 – (3 / max(4,6)) = 0.50 ....(21)

Table 20 shows the global alignment of the pattern and Equation 21 gives the similarity score. One group deletion and two replacements are required to match the reference pattern. Thus, similarity score is 0.50.

sim_{NW} (*S1S2S3S4, S1**S1S1S1S1S1S1S1S1***) = 1 – (4 / max(4,9)) = 0.38 ....(22)

Equation 22 represents a pattern *S1**S1S1S1S1S1S1S1S1***. Three replacements and one group deletion is required to match the reference pattern. Thus, similarity score is 0.38.

sim_{NW} (*S1S2S3S4, S1**S1**S2**S2**S3**S3S2S1***) = 1 – (3 / max(4,8)) = 0.63 ....(23)

Equation 23 represents a pattern *S1**S1**S2**S2**S3**S3S2S1***. Three deletions and one replacement is required to match the reference pattern. Thus, similarity score is 0.63.

sim_{NW} (*S1S2S3S4, S1**S1**S2**S2S2S2S2**S2*) = 1 – (3.5 / max(4,8)) = 0.56 ....(24)

Equation 24 represents a pattern *S1**S1**S2**S2S2S2S2**S2*. Two deletions and two replacements are required to match the reference pattern. Thus, similarity score is 0.56.

The actual string patterns and their corresponding similarity scores show that the youth struggled between the steps of S.T.O.P. activity. For example, the pattern *S1S1S2S1S3S3S2S1* shows that the youth kept going back and forth between the steps and finally backtracked out of the activity. By looking at the design of the S.T.O.P. activity in the REACH app, it is clear that the youth are struggling to provide input in the app, they are tapping on NEXT buttons repeatedly without responding to the feedback messages in that step. This is the major cause of non-compliance observed in S.T.O.P.

The previous two sections presented log analysis and clickstream analysis of users who attempted Daily Diary and S.T.O.P. activities. It was observed that there were many users who were non-compliant in respective activities. After completing Daily Diary and S.T.O.P. activities in the user study, users were asked to respond to a part of survey questionnaire with statements tailored towards these two activities. Table 21 shows the average user response scores for users who were non-compliant during Daily Diary and S.T.O.P. activities. Low scores indicate that the users did not show positive attitudes towards these activities. These low scores corroborate the findings of log analysis and clickstream analysis.

Survey Statement	Average user response for non-compliant users (Daily Diary + S.T.O.P.)
I can use the app without written instructions.	6
In a few steps the app does what I want.	6.5
People using it once or many times would like it.	6.7
It is easy to understand.	6.5

Table 21 Responses to Survey Statements by Non-compliant Users

6.1.6 Blob Tricks

Blob tricks was not part of the REACH protocol. It was a design feature added to the REACH app as a motivation for the youth to complete other activities in the app. The nature of interaction of youth with blob tricks was pretty straightforward. According to log analysis, every youth who participated in the study opened at least 6 Blob tricks out of possible 12 tricks in the app.

In summary, the results of the survey analysis were consistent with the results of preliminary user study (chapter 4, section 4.3.4). The major difference in this study and preliminary study was the number of participants. The log analysis results discussed in this study gave a high level compliance measure with respect to each activity in the session. The clickstream analysis results discussed in this user study provided insights into compliance as well as non-compliance of the youth who participated in the study. This user study was a single session based study. The interaction of the youth with the REACH app was limited in time as well as in scope. The next section will discuss the mixed methods approach for a user study based on the entire length of the REACH protocol (six weeks). This user study will present a new approach of survey analysis that measures the user's perception of the design features in the REACH app.

6.2 User Study 2

N=16 youth from public school participated in experimental study with the REACH app for the full length of the REACH protocol, i.e. 6 weeks. These youth were provided with mobile phones to perform out of session practice at home (after school hours). Log data of the work done by youth was gathered every week in schools. Data gathered from N=10 was included in the study. N=6 data points were excluded from the study because these youth forgot to bring the phones to sessions, or did not charge the phones, or lost the phone, or dropped out of the protocol (two out of six youth dropped out after week 2)).

Based on the REACH protocol schedule, every week, the school psychologist conducted a REACH protocol session in the school and asked the participating youth to practice a particular skill at home (using the app) as an out of session “homework” practice of the skill. Based on this schedule, the log data was parsed and the actual string patterns of the usages were generated. The next subsections discuss the particular week of the REACH protocol, log analysis of the log data of that particular week followed by clickstream analysis results.

6.2.1 Week 1

For week 1, the compliance criteria was completing Daily Diary (DD) at least once per day. Log analysis showed that all 10 youth completed DD during the week but 4 of them did not do it every day. There were several interesting patterns observed after parsing the log data of youth.

The reference string pattern for this week was *D1D2D3D4*. The actual string patterns generated from the parser for the log data of week 1 are, *D1D2D3D4BR* (where R is Relaxation and B is Blob Tricks; refer to Table 10 for states and their meaning), *BRD1D2D3D4*, *RB* and *BD1D2D3D4*. The similarity scores for these patterns are calculated as follows.

sim_{NW} (*D1D2D3D4*, *D1D2D3D4****BR***) = 1 – (1 / max(4,6)) = 0.84 ....,........(25)

sim_{NW} (*D1D2D3D4*, ***BR****D1D2D3D4*) = 1 – (1 / max(4,6)) = 0.84 ....(26)

sim_{NW} (*D1D2D3D4*, ***B****D1D2D3D4*) = 1 – (0.5 / max(4,5)) = 0.80 ....(27)

sim_{NW} (*D1D2D3D4*, **BR**) = 1 – (4 / max(4,2)) = 0 ....(28)

As seen in the equations (25 to 28), the first three strings represent a compliant youth with high similarity score. The last pattern indicates a non-compliant youth that doesn't complete the DD activity but does complete the other two. The complaint youth not only completed Daily Diary, but also explored Blob Tricks and Relaxation regularly. This shows the curiosity of youth to try different activities in the app.

6.2.2 Week 2

In week 2 the compliance criteria for youth was to complete listening to relaxation audio at least once every day during the week. The log analysis showed that the youth were highly compliant and all 10 youth completed listening to Relaxation audios during the week. There were several interesting patterns observed after parsing the log data of youth.

The reference string pattern was *R*. But the actual string patterns generated by the parser were, *D1D2D3D4RBW1W2W3W4, RD1D2D3D4, RW1W2W3W4B, RB and RBD1D2D3D4*. The similarity scores for these pattern are as follows.

sim_{NW} (*R*, ***D1D2D3D4****R****BW1W2W3W4***) = 1 – (4.5 / max(1,10)) = 0.55……………….(29)

sim_{NW} (*R*, *R****D1D2D3D4***) = 1 – (2 / max(1,5)) = 0.80 ……………………………………….(30)

sim_{NW} (*R*, *R****W1W2W3W4B***) = 1 – (2.5 / max(1,6)) = 0.58 ………………………………….(31)

sim_{NW} (*R*, *R****BD1D2D3D4***) = 1 – (2.5 / max(1,6)) = 0.58 ……………………………………(32)

The similarity scores indicate that the actual behavior of youth is similar to the expected behavior in this week. As you can see, every pattern has R as well as some other activity. They have done these activities in a different order. This behavior of youth to complete more than one activity during the week is positive as it indicates that the youth are not only completing the required activity but also completing the activity of their choice.

6.2.3 Week 3

In week 3 the compliance criteria for youth was to complete the Worryheads activity at least once every day during the week. It was observed during the log analysis that the youth have been highly compliant during this week with each youth completing Worryheads at least 4 times during the week.

The reference string pattern for this week was *W1W2W3W4*. The actual string patterns generated by the parser are, *W1W2W3W4D1D2D3D4RB, W1W2W3W4B, W1W2W3W4BBD1D2D3D4* and *W1W2W3W4BR*. These patterns suggest that all youth were compliant with the Worryheads activity. Similarity scores are not required for this week because patterns are clearly indicating that the youth have complied with Worryheads activity. The patterns suggest that the youth kept checking if any new trick was unlocked. These patterns also show that the youth completed DD and Relaxation activities on their own even though they were not asked to do it.

6.2.4 Week 4

In week 4 the compliance criteria for youth was to complete S.T.O.P at least once during week. S.T.O.P. is an activity that is only completed when a youth faces an anxious situation. The reference string pattern for this week was *S1S2S3S4*. The actual string patterns generated by the parser are, *S1S2S3S4, BR, R, B, W1W2W3W4RB* and *W1W2W3W4D1D2D3D4*. The similarity scores for these patterns are as follows.

sim_{NW} (*S1S2S3S4*, *S1S2S3S4*) = 1 – (0 / max(4,4)) = 1 ....(33)

sim_{NW} (*S1S2S3S4*, ***W1W2W3W4RB***) = 1 – (5 / max(4,6)) = 0.17 ....(34)

There are two types of similarity scores. Similarity score 1 indicates that the youth completed S.T.O.P. All other patterns indicate that youth completed activities of their choice during the week. Some of the youth preferred to just practice Relaxation, some of them just played with the Blob Tricks, some of them preferred doing Worryheads instead of going for S.T.O.P. This is because Worryheads is a variation of S.T.O.P and it's easier because the situation and thoughts are already provided in the app and the youth just has to choose between the other thoughts.

6.2.5 Week 5

In week 5, the youth were asked to complete S.T.I.C. activity and get acknowledgement from the teacher or parent. It was observed from the log analysis that the youth were highly non-compliant in this week. The reference pattern for this activity was start state of S.T.I.C. and the end state of the S.T.I.C. that is *STIC1STIC2*. The actual string pattern based on the youth interaction log data is *STIC1*, *STICK1STICK2BR*, *BRS1S2S3S4* and RB. The similarity scores for the actual patterns are as follows.

sim_{NW} (*STIC1STIC2*, *STIC1STIC2*) = 1 – (0 / max(2,2)) = 1 ....(35)

sim_{NW} (*STIC1STIC2*, *STICK1STICK2**BR***) = 1 – (1 / max(2,4)) = 0.75 ....(36)

sim_{NW} (*STIC1STIC2*, *STICK1*) = 1 – (1 / max(2,1)) = 0.50 ....(37)

The patterns not involving either of the STIC1 or STIC2 states are considered non-compliance patterns because they clearly are not similar to the reference pattern. The non-compliance patterns are more than the compliance patterns for S.T.I.C.. The patterns indicate that youth did start the activity but did not complete it. Patterns also show the youth practicing S.T.O.P. and Relaxation activities in this week.

6.2.6 Week 6

Week 6 was a week to practice the skills of the youths' choice. They were asked to complete at least one activity every day in this week. The actual string patterns generated by the parser for the log data provide the evidence of varied use of the app during the week. The pattern *BW1W2W3W4D1D2D3D4STIC1STIC2RBB* indicates that the youth practiced Worryheads, Daily Diary, STIC and Relaxation during the week. They also kept checking blob tricks if any trick was unlocked. It was observed that every youth did complete at least one activity during the week. Some of the youth did complete STIC activity in week 6 but did not do it in week 5. Almost all the youth played with blob tricks. Worryheads was among the favorite activities throughout the protocol and it reflected in week 6 as well.

This user study was of six week duration and it showed the variety of patterns of usages. Some of the patterns generated during this study for the user log data indicate the interest of users and their inclination towards specific activities in the app. Youth were compliant with respect to Daily Diary, Worryheads and Relaxation throughout the six weeks. This gives a new perspective to analyze the log data from a correlation point of view. The next section will discuss the correlation analysis between the activities of the REACH app.

6.2.7 Post Test Survey Analysis of the REACH App

In this user study, the survey questionnaire was tailored to understand user's perception of design feature of the REACH app. The survey questionnaire was given to each participant at the end of the user study. The survey statements were targeting design features like Rewards, Customized Navigation, Consistent Theming, Calls to Action and Input methods. Youth responded to each item using a 10-point rating scale (1= "not at all" to 10 = "very much"). Responses for statements belonging to a particular category of design feature were added together. For example if there are three statements in the survey that are based on Blob tricks, then the responses provided by youth for these three statements are added and the mean score for Blob tricks is calculated. Based on this approach, mean score and standard deviation for every design feature is calculated (Table 22).

Type of design feature:	Questions	Sd	Mean Score
Consistent Theming	7	0.43	9.12
Customized Navigation	7	0.51	6
Rewards	4	0.60	8.48
Calls to Action	7	0.69	8.08
Input Methods	3	0.63	6.5

Table 22 Survey Analysis Results

In the clickstream analysis, several usage patterns were suggesting the struggle of youth while navigating between the states of activities like Daily Diary and S.T.O.P. The survey analysis results in Table 22 also suggest something similar. Low mean scores observed in the table such as 6.5 for Input Methods and 6 for Customized Navigation, indicate that the youth were not feeling very positive about the design of the input methods and customized navigation. On the contrary, youth seem to be feeling very positive about Gamification and Rewards design features. This is observed in the clickstream analysis as well. There were several patterns showing youth checking regularly if a Blob trick has been unlocked or not. Glowing buttons and notifications are design features that are categorized under Calls to Action. The 8.08 mean score suggests that the youth were liking the design feature. Youth seem to like the theme of the app as they reported a very high score for the Consistent Theming design feature. The results presented in Table 22 are the results of the entire population of users who participated in the study (compliant as well as non-compliant). It is important to partition these results and look at the survey responses specific to the categories of compliant and non-compliant users separately. Table 23 shows the results of survey analysis for partially compliant as well as non-compliant users. It is observed that the non-compliant users gave very low scores to the *Input Methods, Calls to Action* and *Customized Navigation* design features. Partially compliant users gave average scores for *Input Methods, Calls to Action* and *Customized Navigation*.

Type of design feature:	Mean Score for Partially Complaint users	Mean Score for Non-compliant users
Consistent Theming	8.86	7
Customized Navigation	7	5.9
Rewards	9.5	6.37
Calls to Action	7.1	5.83
Input Methods	6	3.55

Table 23 Survey Responses for Partially Compliant and Non-compliant Users

6.2.8 Relationship between the Activities

This section deals with recognizing usage patterns that provide evidence of one activity in the REACH app influencing the usage of other activity and thereby impacting compliance.

6.2.8.1 S.T.O.P and Worryheads

In week 4, the youth were asked to practice S.T.O.P. at home. The reference string pattern for this week was *S1S2S3S4* and the actual string patterns generated by the parser were *W1W2W3W4W1W2W3W4RB*, *W1W2W3W4D1D2D3D4* and *RBS1S2S3S4W1W2W3W4*. Most of the youth preferred doing Worryheads instead of doing S.T.O.P. These patterns indicated that there is something about Worryheads which attracts youth towards it.

	
S.T.O.P. steps	Worryheads steps

Table 24 Comparison Between S.T.O.P and Worryheads Design

By looking at the design in Table 23, the preference of Worryheads over S.T.O.P. becomes clear. S.T.O.P. requires the youth to enter what was the situation, by the use of keyboard. The steps in S.T.O.P. also make it mandatory for the youth to enter their response. If the youth does not enter the response then the app does not let the youth to go to the next state. Worryheads in comparison is a straightforward activity. As seen in the table, youth has to read the S (Situation) and accordingly select the O (Other thoughts) from the possible four options. Also there can be situations that in a week's time, the youth actually did not feel anxious. So the S.T.O.P. activity is not relevant in that case and cannot be planned to be done every day.

6.2.8.2 Blob tricks and other Activities

There were patterns observed throughout the six weeks indicating youth were interested in the Blob tricks. There were two types of patterns that show the way Blob tricks affected compliance of other activities. Youth were told during the sessions that, if they practice the skills at home regularly using the app as told by the psychologist, the Blob tricks will start getting unlocked. This thought of getting a reward by completing the homework was seen in the usage patterns as well. The actual patterns generated by the parser were BD1D2D3D4RB, BW1W2W3W4W1W2W3W4R and RBS1S2S3S4B. These patterns indicate that the youth were completing the tasks and they kept checking if a new trick has been unlocked. These patterns were generated for 8 out of 10 youth.

Another type of pattern generated by the parser indicated Blob tricks affecting compliance in a negative way. The actual string patterns observed for 2 out of 10 youth were W1W2W2W1BBB, BSTIC1, D1D2BBBB and BBBBBB. These patterns indicate that the youth kept playing with Blob tricks when they were asked to complete activities like S.T.I.C. and Daily Diary. In these patterns Blob tricks are distracting youth from doing intended activities.

6.2.9 Key Observations

The previous sections discussed week by week compliance measures and provided reasoning behind compliance and non-compliance. This section will mention some of the key observations about the compliance throughout the duration of 6 weeks.

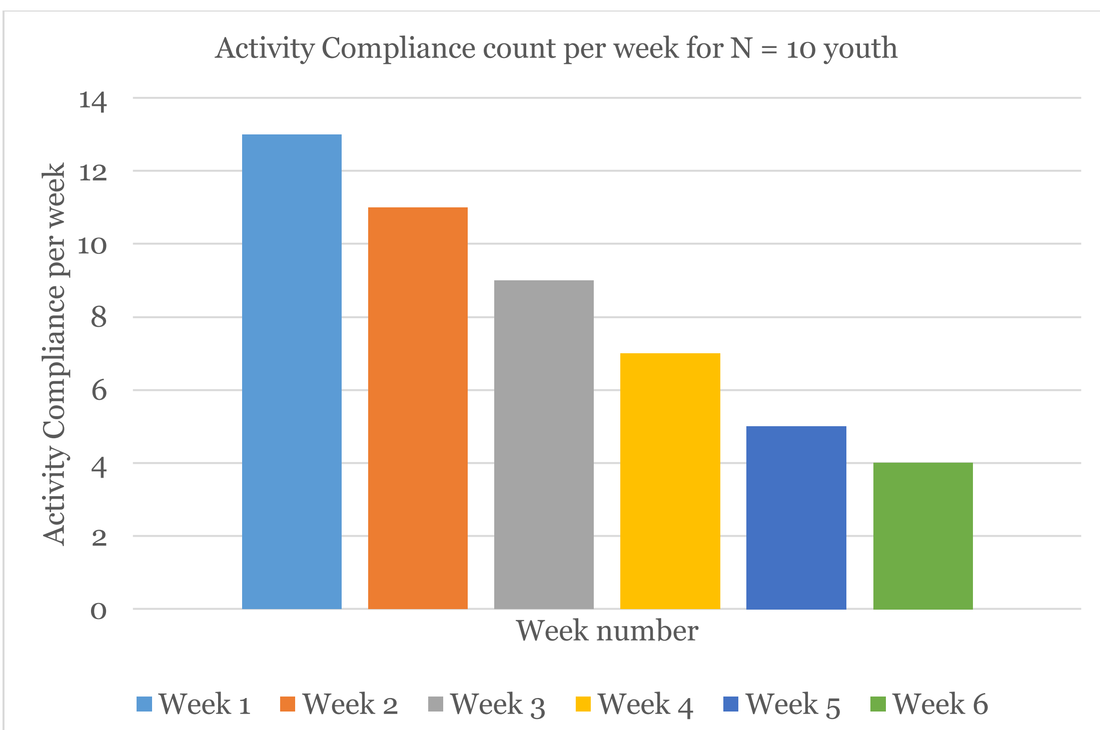


Table 25 Activity Compliance Count per Week

The first key observation is based on the compliance measures of users taken every week for six weeks. As seen in the Table 24, the compliance is highest in week 1 and it goes on decreasing as weeks go by and by the end of the last week it has gone considerably low.

The second key observation regarding compliance of users is regarding activity completion frequency during the week (Table 25). The users tend to complete more activities one or two days prior to the REACH protocol school session. This is an example of a common behaviour in users similar to completing homework, just the day before the due deadline.

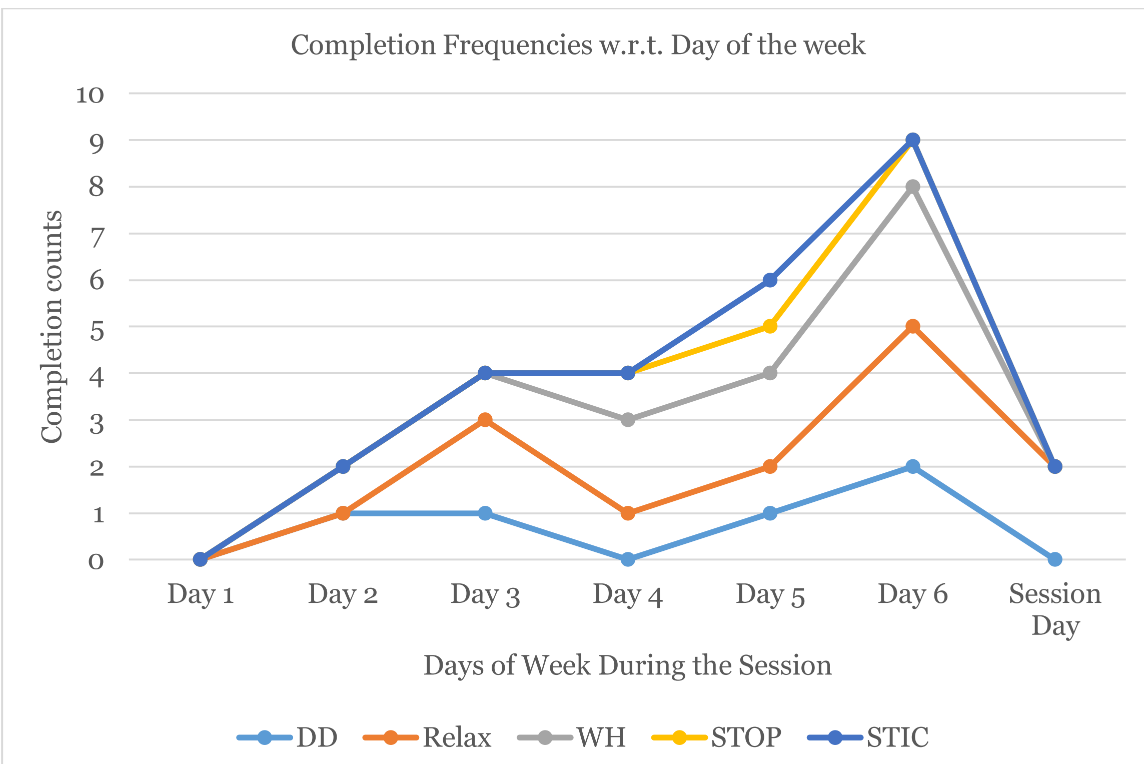


Table 26 Activity Completion Frequency w.r.t Day of the Week

This section presented a full length REACH protocol study and the mixed methods analysis results with respect to weekly user interactions.

In summary, this chapter presented results from two user studies and the application of mixed methods to assess the usability of the application. The survey analysis of preliminary study discussed in chapter 4, section 4.3.4 as well as the survey analysis presented in user study 1 and user study 2 showed that the users showed high ratings for the usability of the REACH app. However the findings of user study 1 and 2 were more focused towards understanding high level compliance measures and the factors affecting compliance.

To discuss the findings of these user studies, it is important to revisit the research questions presented in chapter 3, section 3.1. The first research question of this thesis was to asses, if the introduction of an mHealth app increases compliance of patients to the clinical protocol. To that end, considering the first user study, log analysis showed high compliance for activities like Relaxation and Worryheads. However a low compliance was observed in Daily Diary and S.T.O.P. The clickstream analysis results showed patterns of usages that gave insights about compliant users as well as non-compliant users. It was observed that the non-compliant users struggled between the steps of Daily Diary and S.T.O.P., thereby not completing the required steps in the activity. One of the key findings of the clickstream analysis was the patterns representing distinction between compliance and partial compliance. Compliance patterns were straightforward and showed that the users completed the activities exactly as per the designer's intent. On the other hand, partially compliant users struggled between the steps of the activity but somehow ended up completing the activity. This behavior of users showed that the design of the REACH app was able to lead the users in the direction of the completing the activity. An example of this behavior is presented in user study 1 where a feedback message given by the app in Daily Diary and S.T.O.P. activity may have been the reason for some users to complete these activities, despite struggling in the first few steps of the activity. The survey analysis showed that the responses given by compliant and non-compliant users corroborated the results from log analysis and clickstream analysis. Compliant users reported high scores for the survey statements tailored towards activities like Relaxation and Worryheads, whereas non-compliant users reported low scores for survey statements tailored towards Daily Diary and S.T.O.P. This user study showed the application of mixed methods approach and how these three methods inform on each other to give a better understanding of the usability of the app.

The second research question of this thesis was to identify specific design features of the REACH app responsible for compliance as well as non-compliance. The results of log analysis of the second user study showed that the users were compliant with Daily Diary, Relaxation and Worryheads. The users were non-compliant with S.T.O.P and S.T.I.C. activities. It can be observed that the users are compliant with the REACH protocol in their own way and they kept on completing activities of their choice. This shows that the REACH app was helping the users to practice REACH activities.

To understand the reasons behind non-compliance survey analysis and clickstream analysis results were considered. In the survey analysis, the non-compliant users gave low scores for design features like *Customized Navigation, Input Methods and Calls to Action*. The clickstream analysis showed patterns of users struggling to complete S.T.O.P. and S.T.I.C. activities. The design of the S.T.O.P. and Worryheads showed that, Worryheads provides the user with prepopulated options to choose from, whereas the design of the S.T.O.P. activity makes it mandatory for the user to give an input to complete the activity. This may have been the reason behind the high compliance observed in Worryheads compared to the low compliance observed in S.T.O.P. activities. During the six weeks of the REACH protocol study, the users were continuously checking whether the Blob Tricks are unlocked or not. High compliance with Blob Tricks indicates that the users were motivated by the Blob and they were doing other activities in the app to unlock Blob Tricks.

The discussion in this section shows that the mixed methods approach to validate design of the application does give valuable insights about the impact of design on clinical protocol compliance. The findings also show, how an mHealth app helps to perform out of session practice during a clinical protocol. This was one of the major contributions of this work. The next chapter will discuss lessons learned and future work.

7 CONCLUSIONS AND FUTURE WORK

This chapter will discuss the lessons learnt and future work of this thesis. The discussion will start with a summary of the research method and outcomes followed by the future work possible in this area.

7.1 CONCLUSIONS

This research studies the impact of usability features on clinical protocol compliance and success. The main contribution of this thesis is a mixed methods approach to assess the usability of mHealth applications. This mixed method approach extends the use of surveys by adding user interaction log analysis to determine compliance, and clickstream analysis to attempt to determine patterns of where users get "off-track" (become non-compliant to a design intent). The objective of this research is to evaluate how the design details of an mHealth application impact compliance of patient to the clinical protocol. To that end, first contribution of this thesis is a case study in participatory design of an mHealth app for a pediatric chronic disease. This research focuses on anxiety as a chronic disorder, specifically child anxiety. The first part of this thesis focuses on understanding the anxiety disorder domain by discussing the related work done so far in this area. As a part of case study, REACH protocol for resolving childhood anxiety disorders is identified as the clinical protocol for this research. The second part of this thesis talks about a multidisciplinary team based effort for designing an mHealth application based on the REACH protocol. After sharing the details about the implementation and features of the REACH app, the very basic method of usability validation in the form of survey questionnaires is performed to get initial user perception of the app. To that end, the survey based usability validation method gave positive results. The REACH app system received high and positive rating by the participants.

This was a preliminary study and the number of participants were less. The study was small in scope and did not give any measure of compliance of patients, which was the ultimate objective. The preliminary study showed that the REACH app is a promising option to use as an out of session means of practicing activities in the REACH app.

The second user study was conducted with a larger number of participants. Participants were given mobile phones with REACH app installed and a survey questionnaire to fill out after the session. They were instructed to do activities in the REACH app in a single session based user study. The interactions of users with phones were logged locally on the devices. This data, along with responses to survey questionnaire were collected from the participants. Following the previous trend, REACH app again received high and positive rating by the participants. However, in this study, compliance (as well as non-compliance) of patients was calculated using one of the mixed methods, called as log analysis. The reasons behind compliance and non-compliance are analyzed based on the third method of mixed method approach, called as clickstream analysis. The results of this user study showed that log analysis can be used to measure high level compliance of patients to the clinical protocol and clickstream analysis can be used to identify specific usage patterns of the users to understand the reason behind users getting “off-track” from a design intent.

The third and final user study was conducted for a duration of six weeks (REACH protocol duration). For this user study, survey questionnaire were tailored to assess user’s perception of the specific design features of the REACH app. The mean scores were calculated per category of the design feature based on the responses provided by the users. These scores (out of 10) helped to assess user’s perception of a specific design feature. Similar to that of the second user study, log analysis was used to calculate numerical compliance of users to the REACH protocol. Clickstream analysis gave insights of how design decisions may have impacted user actions, thereby generating unique usage patterns. The duration of this study was six weeks, which is the length of the REACH protocol. The results of this user study were crucial to this thesis because, compliance was measured for the six weeks period based on the activities completed by users at home space. This gave a better understanding of how a mHealth application is used by a user in a real life setting wherein there is no one to govern the app usage. The results showed that the users were compliant in their own way and they were practicing skills of their own choice. This clearly indicates that the app was successful to be used along with the clinical protocol as an out of session practice tool.

In Summary, this thesis' contributions are a case study in participatory design of an mHealth app for a pediatric chronic disease, and a novel method of usability validation that attempts to tie design outcomes to clinical outcomes (namely compliance). To my knowledge, this is the first study of its kind where a combination of methods are used for usability validation and tying the results of these methods to the clinical outcomes. From a clinical protocol compliance perspective, REACH app appears highly useful for increasing between session skill practice and potentiating dosage of active change ingredients (Stoll et al. 2016). This is a promising result for the future of mHealth apps for engaging patients in the treatment process by providing an option for them to perform out of session activities.

Although these contributions are limited to a single domain, protocol, and app, the outcomes are of interest due to the chronic disorder domain (anxiety), the nature of the intervention (preventative-early intervention), the use of an app to increase protocol compliance, and the integration of concepts from innovative design technology (gaming, notifications, user experience design) resulting in improved clinical outcomes.

7.2 FUTURE WORK

From a software engineering point of view, the mixed methods approach discussed in this thesis can be applied to various mHealth apps to validate the usability. It has been observed in the literature that, researchers like Jaspers (Jaspers, 2009) talk about, how more than one method of usability validation complements each other and gives a better understanding of the usability of the app. Mixed methods give minute details of the app usage which can be compared with the design intent. These mixed methods can be applied to other mHealth apps which are used to perform homework activities or used as a tool for an out of session practice during a clinical protocol schedule. However, generalizing the use of such an approach for usability validation is a challenging task. The methods like log analysis and clickstream analysis may not be applied to all the mHealth apps. Every mHealth app is unique in its own context, and the choice of usability evaluation method depends largely on the context of use, type of users and the goals of the mHealth app. Applying these mixed methods approach to other mHealth apps will give a better insight about its validity.

This research used clickstream analysis as one of the mixed methods to validate the usability. A string alignment algorithm by Needlemen-Wunsch (Needleman & Wunsch, 1970) was used to calculate similarity scores between the usage patterns. There are several other methods which can be used to align and calculate similarity of reference pattern versus the actual usage pattern. Again, the choice of string similarity or alignment method largely depends on the nature of strings itself and the context of use. Future researchers in this area will have to understand the nature of app usage patterns, their preprocessing and the categorization of these patterns. Categorization can performed using simple grouping techniques or clustering algorithms (like K-means clustering) can be used to do the same.

One of the promising areas of future work in the usability validation is considering time spent by users while completing a particular task within an app. The string patterns of usage show the actual path taken by users within an app, but including timestamps and the duration of the interaction in the analysis will provide a better understanding of the app usage. The recent literature suggests that, researchers like Lettner (Lettner et al., 2014) have started looking at interaction analysis, a type of analysis which deals with comparing design intent versus the app usage. This thesis is motivated from Lettner's work and it has opened doors to new opportunities in the area of interaction analysis.

One of the key findings of this research was the distinction between partially compliant users and the compliant users. The partially compliant users are of special interest because they give insight about a particular design feature within the app which drives them towards completing the activity. This design feature may be of the form of a feedback message given by the app or a notification guiding the user to do a particular step or visual cues insinuating the users to do the next step towards activity completion. Understanding app usage of these partially compliant users will help the designers to tie back app usage to design details. This will help the future developers of the app to design particular recovery systems within the app to guide the user if they gets "off-track" from a design intent.

In summary, I hope this research contributes to a growing multidisciplinary need to connect clinical research methods with (software) engineering processes.

REFERENCES

Barrett, P. M. (2000). Treatment of childhood anxiety: developmental aspects. *Clinical Psychology Review*, *20*(4), 479–94. http://doi.org/10.1016/S0272-7358(99)00038-0

Bashshur, R., Shannon, G., Krupinski, E., & Grigsby, J. (2011). The taxonomy of telemedicine. *Telemedicine Journal and E-Health : The Official Journal of the American Telemedicine Association*, *17*(6), 484–94. http://doi.org/10.1089/tmj.2011.0103

BJ, A. R. F. (Ed.). (2009). *A Simple, Powerful Way to Change Lives (Stanford University) (Stanford University) Stanford*. CA: Captology Media.

CDC, 2016. Centers for Disease Control and Prevention. http://www.cdc.gov/chronicdisease/

Cramer, J. A., Roy, A., Burrell, A., Fairchild, C. J., Fuldeore, M. J., Ollendorf, D. A., & Wong, P. K. (2008). Medication compliance and persistence: terminology and definitions. *Value in Health : The Journal of the International Society for Pharmacoeconomics and Outcomes Research*, *11*(1), 44–7. http://doi.org/10.1111/j.1524-4733.2007.00213.x

De Land, P. N., & Chase, W. W. (1990). Statistics notebook: Entry I.J, Types of data: Nominal, ordinal, interval, and ratio scales. *Optometry and Vision Science*, *67*(2), 155–156. http://doi.org/10.1097/00006324-199002000-00021

Fiordelli, M., Diviani, N., & Schulz, P. J. (2013). Mapping mHealth research: a decade of evolution. *Journal of Medical Internet Research*, *15*(5), e95.

Fisak Jr, B. J., Richard, D., & Mann, A. (2011). The prevention of child and adolescent anxiety: a meta-analytic review . *Prevention Science : The Official Journal of the Society for Prevention Research*, *12*(3), 255–268. http://doi.org/10.1007/s11121-011-0210-0 [doi]

Free, C., Phillips, G., Felix, L., Galli, L., Patel, V., & Edwards, P. (2010). The effectiveness of M-health technologies for improving health and health services: a systematic review protocol. *BMC Research Notes*, *3*, 250.

FunFriends. (2016). FRIENDS Programs Apps. https://www.friendsprograms.com/friends-programs-apps/

Güyer, T., Atasoy, B., & Somyürek, S. (2015). Measuring disorientation based on the Needleman-Wunsch algorithm. *The International Review of Research in Open and Distributed Learning*, *16*(2).

ISO. (1988). ISO 9241-11: Guidance on Usability

ISO. (1999). ISO 9241-210. Ergonomics of human system interaction-Part 210: Human-centered design for interactive systems (formerly known as 13407).

Istepanian, R., Laxminarayan, S., & Pattichis, C. S. (2006). M-Health: Emerging Mobile Health Systems. *M-Health: Emerging Mobile Health Systems*. Retrieved from http://adsabs.harvard.edu/abs/2006mhem.book.....I

Jaspers, M. W. M. (2009). A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence. *International Journal of Medical Informatics*, *78*(5), 340–53. http://doi.org/10.1016/j.ijmedinf.2008.10.002

Kay, M. (2011). mHealth: New Horizons for Health through Mobile Technologies. *World Health Organization*, *3*, 66–71.

Klasnja, P., & Pratt, W. (2012). Healthcare in the pocket: mapping the space of mobile-phone health interventions. *Journal of Biomedical Informatics*, *45*(1), 184–198.

Lebeau, R. T., Davies, C. D., Culver, N. C., & Craske, M. G. (2013). Homework compliance counts in cognitive-behavioral therapy. *Cognitive Behaviour Therapy*, *42*(3), 171–179.

Lettner, F., Grossauer, C., & Holzmann, C. (2014). Mobile interaction analysis: towards a novel concept for interaction sequence mining. *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services - MobileHCI '14*, 359–368. http://doi.org/10.1145/2628363.2628384

Likic, V. (2008). The Needleman-Wunsch algorithm for sequence alignment. Lecture given at the 7th Melbourne Bioinformatics Course, Bi021 Molecular Science and Biotechnology Institute, University of Melbourne.

Lund, A. (2001). Measuring usability with the USE questionnaire. Usability interface, 8(2), 3-6.

Matthews, M., Voida, S., Abdullah, S., Doherty, G., Choudhury, T., Im, S., & Gay, G. (2015). In Situ Design for Mental Illness. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '15* (pp. 86–97). New York, New York, USA: ACM Press. http://doi.org/10.1145/2785830.2785866

Murray, T., Hardy, D., Spruijt-Metz, D., Hekler, E., and Raij, A. 2013. Avatar interfaces for biobehavioral feedback. Design, User Experience, and Usability. Health, Learning, Playing, Cultural, and Cross-Cultural User Experience Berlin Heidelberg: Springer

Michelle, T. Q. Y., Jarzabek, S., & Wadhwa, B. (2014). CBT Assistant: MHealth App for psychotherapy. In *2014 IEEE Global Humanitarian Technology Conference - South Asia Satellite (GHTC-SAS)* (pp. 135–140). IEEE. http://doi.org/10.1109/GHTC-SAS.2014.6967572

Needleman, S. B., & Wunsch, C. D. (1970). A general method applicable to the search for similiarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, *48*(3), 443–453. http://doi.org/10.1016/0022-2836(70)90057-4

Nielsen, J. Heuristic evaluation, in: J. Nielsen, R.L. Mack (Eds.), Usability Inspection Methods, Wiley, New York, 1994, pp.25–62.

Norman, D. A., and Draper, S. W. 1986. User-Centered System Design: New Perspectives on Human Computer Interacti. Hillsdale N.J. : Lawrence Erlbaum Associates.

Patel, M. S., Asch, D. A., & Volpp, K. G. (2015). Wearable devices as facilitators, not drivers, of health behavior change. *Jama*, *313*(5), 459–460.

Patwardhan, M., Stoll, R., Hamel, D. B., Amresh, A., Gary, K. A., & Pina, A. (2015). Designing a mobile application to support the indicated prevention and early intervention of childhood anxiety. In *Proceedings of the conference on Wireless Health - WH '15* (pp. 1–8). New York, New York, USA: ACM Press. http://doi.org/10.1145/2811780.2811954

Pencil Tool. (2016). http://pencil.evolus.vn/

Pina, A. A., Zerr, A. A., Villalta, I. K., & Gonzales, N. A. (2012). Indicated prevention and early intervention for childhood anxiety: a randomized trial with Caucasian and Hispanic/Latino youth. *Journal of Consulting and Clinical Psychology*, *80*(5), 940–946.

Poushter, Jacob (2016), Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies, Retrieved from http://www.pewglobal.org/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies/

Pramana, G., Parmanto, B., Kendall, P. C., & Silk, J. S. (2014). The SmartCAT: an m-health platform for ecological momentary intervention in child anxiety treatment. *Telemedicine Journal and E-Health : The Official Journal of the American Telemedicine Association*, *20*(5), 419–27. http://doi.org/10.1089/tmj.2013.0214

Price, M., Yuen, E. K., Goetter, E. M., Herbert, J. D., Forman, E. M., Acierno, R., & Ruggiero, K. J. (2014). mHealth: a mechanism to deliver more accessible, more effective mental health care. *Clinical Psychology & Psychotherapy*, *21*(5), 427–36. http://doi.org/10.1002/cpp.1855

Rizvi, S. L., Dimeff, L. A., Skutch, J., Carroll, D., & Linehan, M. M. (2011). A pilot study of the DBT coach: an interactive mobile phone application for individuals with borderline personality disorder and substance use disorder. *Behavior Therapy*, *42*(4), 589–600. http://doi.org/10.1016/j.beth.2011.01.003

Rodgers, A., & Dunsmuir, S. (2015). A controlled evaluation of the "FRIENDS for Life" emotional resiliency programme on overall anxiety levels, anxiety subtype levels and school adjustment. *Child and Adolescent Mental Health*, *20*(1), 13–19. http://doi.org/10.1111/camh.12030

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Amer Psychol., 55,68–78.

Silverman, W. K., Kurtines, W. M., Jaccard, J., & Pina, A. A. (2009). Directionality of change in youth anxiety treatment involving parents: an initial examination. *Journal of Consulting and Clinical Psychology*, *77*(3), 474–485.

Silverman, W. K., Pina, A. A., & Viswesvaran, C. (2008). Evidence-based psychosocial treatments for phobic and anxiety disorders in children and adolescents. *Journal of Clinical Child and Adolescent Psychology : The Official Journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division 53*, *37*(1), 105–130.

Silverman, W. K., & Kurtines, W. M. (1996b). Anxiety and phobic disorders: A pragmatic approach. Springer Science & Business Media.

Smith, T. F., & Waterman, M. S. (1981). Identification of common molecular subsequences. *Journal of Molecular Biology*, *147*(1), 195–197. http://doi.org/10.1016/0022-2836(81)90087-5

Stallard, P., Taylor, G., Anderson, R., Daniels, H., Simpson, N., Phillips, R., & Skryabina, E. (2014). The prevention of anxiety in children through school-based interventions: study protocol for a 24-month follow-up of the PACES project. *Trials*, *15*(1), 77. http://doi.org/10.1186/1745-6215-15-77

Stoll, R., Pina, A.A., Gary, K.G., & Amresh, A. (2016). An initial usability evaluation of a smartphone application designed to support the prevention and early intervention of anxiety in youth. Unpublished manuscript, Cognitive and Behavioral Practice.

Stoll, R., Pina, A.A., Gonzales, N., Gary, K.G., Amresh, A., & Patwardhan, M. (2016, March). Using mHealth technologies to improve child anxiety prevention efforts in real-world settings. Poster presented at the ASU ISSR Graduate Poster session.

Stone, A. A., Shiffman, S., Schwartz, J. E., Broderick, J. E., & Hufford, M. R. (2002). Patient non-compliance with paper diaries. *BMJ (Clinical Research Ed.)*, *324*(7347), 1193–4. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=111114&tool=pmcentrez&rendertype=abstract

Taniguchi, D. (2004). Click stream analysis. Google Patents. Retrieved from https://www.google.com/patents/US20040133671

Vredenburg, K., Mao, J., Smith, P. W., and Carey, T. 2002. A Survey of User-Centered design Practice. Paper presented at the Proceedings of the 2002 ACM Symposium on Computer- Human Interaction (CHI 2002), Minneapolis, MN, April 2002., Minneapolis, MN.

Wang, J., Wang, Y., Wei, C., Yao, N. A., Yuan, A., Shan, Y., & Yuan, C. (2014). Smartphone interventions for long-term health management of chronic diseases: an integrative review. *Telemedicine Journal and E-Health : The Official Journal of the American Telemedicine Association*, *20*(6), 570–83. http://doi.org/10.1089/tmj.2013.0243

Whittaker, R. (2012). Issues in mHealth: findings from key informant interviews. *Journal of Medical Internet Research*, *14*(5), e129.

APPENDIX A

IRB APPROVAL

APPROVAL: EXPEDITED REVIEW

Armando Pina

Psychology 480/965-0357
Armando.Pina@asu.edu

Dear Armando Pina:

On 9/23/2015 the ASU IRB reviewed the following protocol:

Type of Review:	Initial Study
Title:	SMART REACH: Pilot demonstration study
Investigator:	Armando Pina
IRB ID:	STUDY00003245
Category of review:	(7)(b) Social science methods, (7)(a) Behavioral research
Funding:	None
Grant Title:	None
Grant ID:	None
Documents Reviewed:	• Application , Category: IRB Protocol; • Support from the Schools for the study , Category: Off-site authorizations (school permission, other IRB approvals, Tribal permission etc); • Pretest, Category: Measures (Survey questions/Interview questions /interview guides/focus group questions); • Screen, Category: Measures (Survey questions/Interview questions /interview guides/focus group questions); • Consent, Category: Consent Form; • Posttest, Category: Measures (Survey questions/Interview questions /interview guides/focus group questions);

The IRB approved the protocol from 9/23/2015 to 9/22/2016 inclusive. Three weeks before 9/22/2016 you are to submit a completed Continuing Review application and required attachments to request continuing approval or closure.

If continuing review approval is not granted before the expiration date of 9/22/2016 approval of this protocol expires on that date. When consent is appropriate, you must use final, watermarked versions available under the "Documents" tab in ERA-IRB.

In conducting this protocol you are required to follow the requirements listed in the INVESTIGATOR MANUAL (HRP-103).

Sincerely,

IRB Administrator

cc:

Mandar Patwardhan, Ryan Stoll

APPENDIX B

SIMILARITY SCORE CALCULATION METHOD

This section will discuss in detail the method of calculating string similarity scores. This method is similar to that of Lettner's (Lettner et al., 2014) approach of similarity estimation. It uses a concept of global alignment of reference and actual strings. Needlemen-Wunsch (Needleman et al., 1970) (NW) algorithm is used to calculate the global alignment. After getting the global alignment, either delete or insert or replace operations are performed on the aligned string to convert it into reference string. The number of operations are measured and a similarity score is calculated based on the formula derived from Levensthein's formula of calculating similarity score mentioned in Lettner's research (Lettner et al., 2014).

The discussion will start with explaining the concept of global alignment, followed by the details of NW algorithm and finally calculating similarity score based on the operations performed on the aligned string sequence.

```
SIMILARITY
PI-LLAR---
```

Figure 6 Global Alignment Example

Global alignment of two sequences is nothing but the best alignment over the entire length of two sequences. Example of global alignment is given in Figure 6 that shows how two sequences are aligned to get the maximum similarity between them. The "—" indicates the gap introduced in the actual string to align it with reference string. As Guyer explains in his paper,

> *"The Needleman-Wunsch (N-W) algorithm was proposed in 1970 by Saul Needleman and Christian Wunsch. It is commonly used for global sequence alignment and scoring. Although the original purpose of this algorithm was to search for similarities between protein or nucleotide sequences in*

bioinformatics (Needleman & Wunsch, 1970), the algorithm is being used in various interdisciplinary fields to measure the string similarity. This string-matching algorithm includes four parameters and it finds the optimal alignment of two sequences and computes similarities between them. The first two parameters in the algorithm are two strings that ideally should match. The third parameter is a similarity matrix, showing relations between each character of the two strings. The fourth is a gap penalty which is a value designed to reduce the score when the characters do not match" (Güyer et al., Page 190, 2015).

NW algorithm is an example of dynamic programming that builds up the best alignment by using optimal alignments of smaller subsequences. To understand the working of NW algorithm, a simple example including two strings, *SEND* and *AND* is explained. The following example follows presentation in (Likic, V., 2008). There are three steps in NW algorithm as follows.

1. Initialization of the score matrix

		S	E	N	D
	C(1,1)	C(1,2)	C(1,3)	C(1,4)	C(1,5)
A	C(2,1)	C(2,2)	C(2,3)	C(2,4)	C(2,5)
N	C(3,1)	C(3,2)	C(3,3)	C(3,4)	C(3,5)
D	C(4,1)	C(4,2)	C(4,3)	C(4,4)	C(4,5)

Figure 7 Score Matrix by Likic, V. (2008). The Needleman-Wunsch algorithm for sequence alignment 7th Melbourne Bioinformatics Course [Vladimir Likic]. Retrieved from http://www.cs.sjsu.edu/~aid/cs152/NeedlemanWunsch.pdf

The cells of the score matrix are labelled C(i, j) where i = 1, 2, ..., N and j = 1, 2, ..., M (Figure 7). Let us consider the gap penalty for a character mismatch is 2. Based on this assumption the score matrix cells are filled by row starting from the cell C(2, 2).

2. Calculation of scores

The score of any cell C(i, j) is the maximum of:

$q_{diag} = C(i - 1, j - 1) + S(i, j)$

$q_{up} = C(i - 1, j) + g$

$q_{left} = C(i, j - 1) + g$

where S(i, j) is the substitution score for letters i and j, and g is the gap penalty. Let us assume the substitution score is +1 for a match, -1 for a mismatch. So if the characters i and j match, then S(i, j) will be +1 and if they don't match, then the S(i, j) will be -1. The value of the cell C(i, j) depends only on the values of the immediately adjacent northwest diagonal, up, and left cells that is C(i – 1, j – 1), C(i – 1, j) and C(i , j – 1). Based on the formula, the cells in the score matrix are filled by calculating C(i, j) for each row and column. The resulting matrix looks like Figure 8. The arrows indicates the cell containing best score among C(i – 1, j – 1), C(i – 1, j) and C(i , j – 1) chosen at that particular C(i, j) calculation.

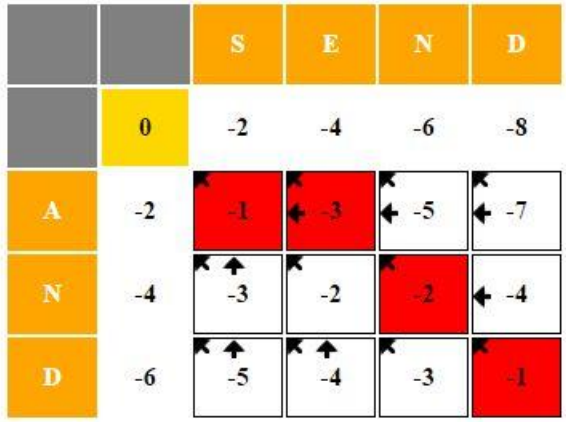


Figure 8 Score Matrix after Calculations: Generated by http://experiments.mostafa.io/public/needleman-wunsch/index.html

3. Deducing the alignment from the score matrix

The alignment is deduced from the arrows indicated in the score matrix. The pattern alignment is generated backwards by considering the character is the rightmost cell in the matrix. In this case the rightmost cell contains a -1 score, represents D character with a diagonal arrow. A diagonal arrow indicates a match in character. A left arrow indicates that the gap should be introduced in the *left* sequence and an upward pointing arrow indicates that the gap should be introduced in the *up* sequence. Based on this assumption the global alignment is derived as shown in Figure 9.

```
SEND
A-ND
```

Figure 9 Global Alignment of Two Strings

Based on the global alignment, delete, insert or replace operations are performed on the actual string to convert it to the reference string. Let us consider *SEND* string as a reference string and *AND* as the actual string. The global alignment as per NW algorithm is *A-ND*. To convert *A-ND* to *SEND*, two operations are required. First will replace *A* with *S* and second will insert *E* at the gap location. So the total number of operations to convert *A-ND* to *SEND* is 2.

The formula to calculate similarity score is derived from Levenstein's formula of similarity score measurement. Levenstein's formula is as follows:

$sim_{LV}(a, b) = 1 - n / (\max(\text{len } a, \text{len } b))$, where *a* is a reference string and *b* is actual string. *Len a* gives the length of string *a* and *n* is the number of operations performed to convert the string *b* to string *a*.

The formula used in this thesis is derived from Levenstein's formula and it is based on the nature of actual string patterns. The formula is given as follows:

$sim(a, b) = 1 - (N / \max(\text{len } a, \text{len } b))$...(38)

Where N = (# of deletions/ (1+ # of recovery actions)) + # of inserts + # of replacements, where # is used to represent *number*.

Let us consider the reference pattern for Daily Diary Activity is D1D2D3D4. If we consider the actual pattern observed by the user as **D1D1**D2D3D4, then according to the design of the app, when user taps on the NEXT button in D1 without completing the D1 step, a feedback message is given to the user that tells the user to complete the D1 step. This action is considered as a recovery action. Thus, *number of recovery actions* in this case is 1. So formula for N for this thesis is,

N = (# of deletions/2) + # of inserts + # of replacements.

Let us consider an actual string of D1D1D1D2D3D4. The global alignment using NW method for the reference pattern D1D2D3D4 is

D1 **--** **--** D2 D3 D4

D1 **D1 D1** D2 D3 D4

Based on the global alignment, it can observed that two deletions are required in actual pattern to convert it to the reference pattern. So the value of N for calculating similarity score using Equation 38 is as follows:

N = (# of deletions/ (1+ # of recovery actions)) + # of inserts + # of replacements

N = (2/ (1+1)) + 0 + 0 = 1. Thus, substituting the value of N in equation 38 gives,

sim(D1D2D3D4, D1**D1D1**D2D3D4) = 1 – (1 / max(4, 6)) = 0.83……………………(39)

In this way similarity scores are calculated using NW algorithm and a similarity estimation formula given by Equation 38. Various examples of similarity score calculations are given in Table 26.

Actual String Patterns (GA indicates Global Alignment)	Similarity scores calculated using Equation 38
*D1D2D3**D2D1D2D3D3D3**D4* *GA: D1D2D3 -- -- -- -- -- -- D4*	N= (6/2 deletions + 0 insertions + 0 replacements) Thus similarity score = 1 – (3/10) = 0.7
*D1**D1D1D1D1**D2D3D4* *GA: D1 -- -- -- -- D2D3D4*	N= (4/2 deletions + 0 insertions + 0 replacements) Thus similarity score = 1 – (2/8) = 0.75
*D1**D1D1D1D1D1D1D1*** *GA: D1**D2D3D4 -- -- -- --*** (replace first three D1s with D2D3D4, delete remaining D1s from actual string)	N= (4/2 deletions + 0 insertions + 3 replacements) Thus similarity score = 1 – (5/8) = 0.375
*D1**D1D1D2D2D2*** *GA: D1 -- -- D2D3D4*	N= (2/2 deletions + 0 insertions + 2 replacements) Thus similarity score = 1 – (3/6) = 0.5

Table 27 Examples of Similarity Score Calculations

To check the validity of the formula given in Equation 38, similarity scores of various types of actual strings were calculated and the results were plotted to check if the similarity scores are consistent or not. The reference string *ABCD* was considered where *A*,*B*,*C* and *D* are the states. The actual patterns considered in the validation contained patterns between *ABCD* and *XYZW* that is the actual string which will have the highest similarity score of 1 and the string which will have the least similarity score of 0. *ABCD* gets the highest similarity score because it is the same string as the reference pattern. *XYZW* will have a 0 similarity score because not a single character is similar to that of the pattern *ABCD*. After plotting the results, it was observed that the formula given by Equation 38 gives consistent results for the variety of actual strings. The graph of strings versus similarity scores is calculated and shown in Figure 10.

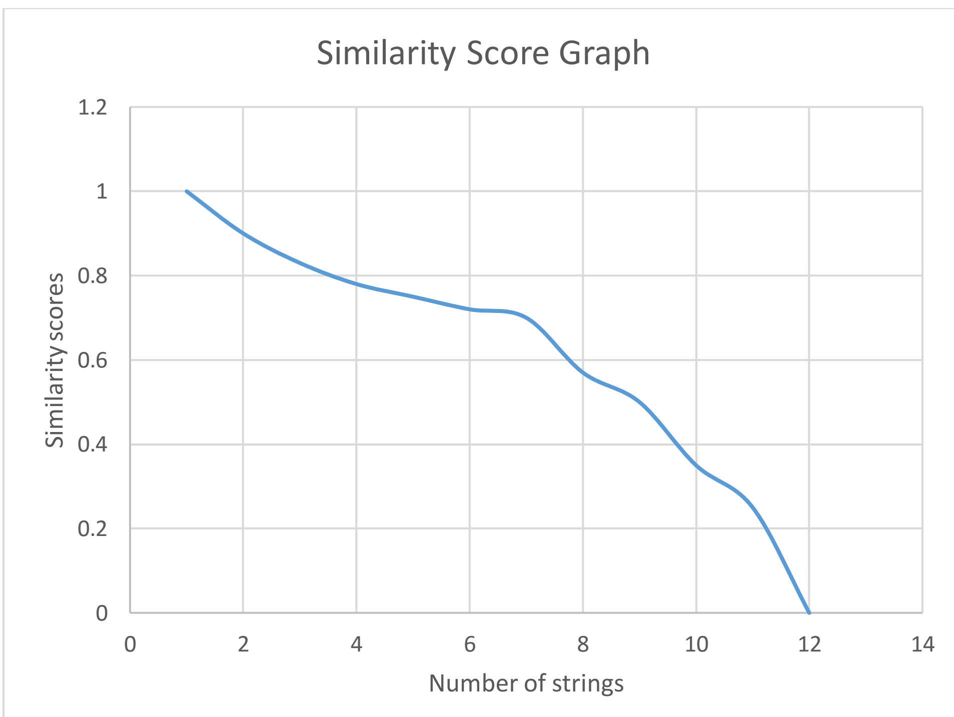


Figure 6 Graph of Similarity Scores vs Actual Strings

APPENDIX C

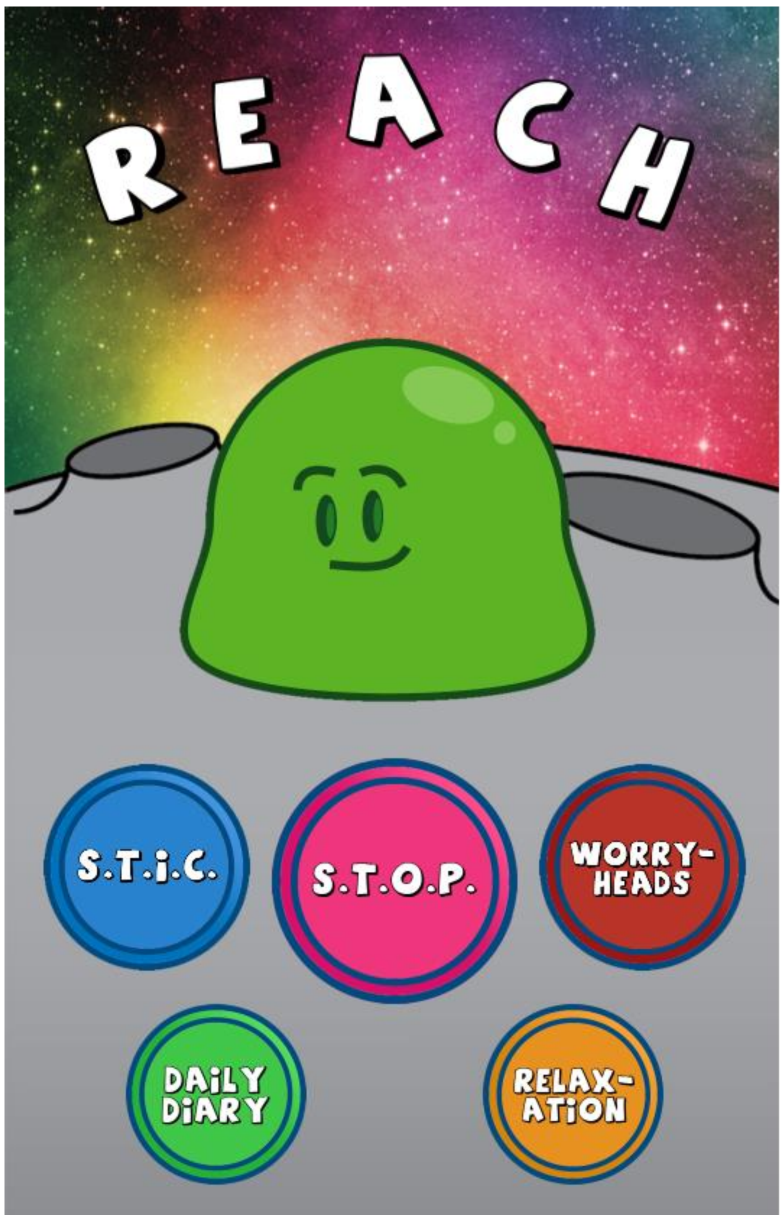
SURVEY QUESTIONNAIRES

User Study 2 Survey Questionnaire

Usefulness, Satisfaction, and Ease (for Android Users Only)

(student/youth)

My name is __________ and I work at ASU. I am doing a study to learn what you think about an app for smartphone we have created.



Part 1

Tap on relaxation and listen to the INTRO using the earbuds.

Tap on WORRYHEADS and try it out.

When you are done, respond to the statements below. There is no right or wrong answer. Just circle the number that best describes what you think. Okay?

Not at all		Somewhat								Very much
1. It is easy to use	1	2	3	4	5	6	7	8	9	10
2. It is simple to use.	1	2	3	4	5	6	7	8	9	10
3. It is easy to understand.	1	2	3	4	5	6	7	8	9	10
4. In a few steps it does what I want.	1	2	3	4	5	6	7	8	9	10
5. It lets me do several things.	1	2	3	4	5	6	7	8	9	10
6. Using it requires no effort.	1	2	3	4	5	6	7	8	9	10
7. I can use it without written instructions.	1	2	3	4	5	6	7	8	9	10
8. I don't notice any problems as I use it.	1	2	3	4	5	6	7	8	9	10
9. People using it once or many times would like it.	1	2	3	4	5	6	7	8	9	10
10. Mistakes can be fixed quickly and easily.	1	2	3	4	5	6	7	8	9	10
11. I can use it well every time.	1	2	3	4	5	6	7	8	9	10

Part 2

Tap on the DAILY DIARY or S.T.O.P. to try it out.

When you are done, respond to the statements below:

	Not at all	Somewhat								Very much
1. The instructions and messages are easy to understand.	1	2	3	4	5	6	7	8	9	10
2. I easily remember how to use it.	1	2	3	4	5	6	7	8	9	10
3. It is easy to learn to use it.	1	2	3	4	5	6	7	8	9	10
4. I quickly became good at it.	1	2	3	4	5	6	7	8	9	10
5. The messages to fix problems are clear.	1	2	3	4	5	6	7	8	9	10
6. The instructions and messages are clear.	1	2	3	4	5	6	7	8	9	10
7. I learned to use it quickly.	1	2	3	4	5	6	7	8	9	10
8. It is easy to use.	1	2	3	4	5	6	7	8	9	10
9. It is simple to use.	1	2	3	4	5	6	7	8	9	10
10. It is easy to understand.	1	2	3	4	5	6	7	8	9	10
11. In a few steps it does what I want.	1	2	3	4	5	6	7	8	9	10
12. It lets me do several things.	1	2	3	4	5	6	7	8	9	10
13. Using it requires no effort.	1	2	3	4	5	6	7	8	9	10
14. I can use it without written instructions.	1	2	3	4	5	6	7	8	9	10
15. I don't notice any problems as I use it.	1	2	3	4	5	6	7	8	9	10
16. People using it once or many times would like it.	1	2	3	4	5	6	7	8	9	10
17. Mistakes can be fixed quickly and easily.	1	2	3	4	5	6	7	8	9	10

Part 3

Tap on BOB THE BLOB to try it out.

When you are done, respond to the statements below:

	Not at all	Somewhat								Very much
1. I am happy with this app	1	2	3	4	5	6	7	8	9	10
2. I would tell a friend about this app	1	2	3	4	5	6	7	8	9	10
3. This app is fun to use	1	2	3	4	5	6	7	8	9	10
4. This app works the way I would want it to work	1	2	3	4	5	6	7	8	9	10

	Not at all	Somewhat								Very much
1. Would you be embarrassed to have this app?	1	2	3	4	5	6	7	8	9	10
2. Would you get tease or picked-on by other kids for having this app?	1	2	3	4	5	6	7	8	9	10
3. Would you get any criticism or hassles at home for having this app?	1	2	3	4	5	6	7	8	9	10
4. Would you get any criticism or hassles at school for having this app?	1	2	3	4	5	6	7	8	9	10

How old are you________________		**Gender:**	**Girl**	**Boy**
What grade are you in ___________				
Are you: White	**Hispanic**	**Black**	**Asian**	**Other**
Do you know how to use an Android/Google phone?	**Yes**	**No**		
How often do you play games on the Phone?	**Never**	**Sometimes**	**Often**	**All the time**

User Study 3 Survey Questionnaire

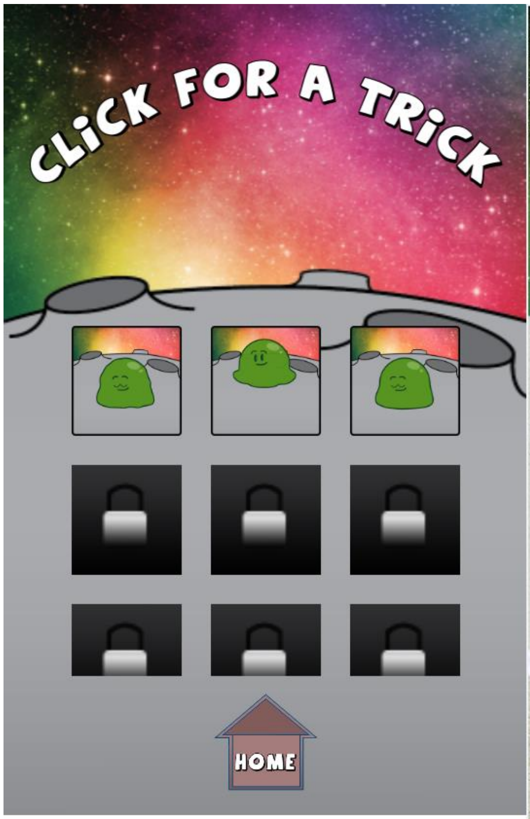
REACH app: User Satisfaction and Ease (for Android Users Only)

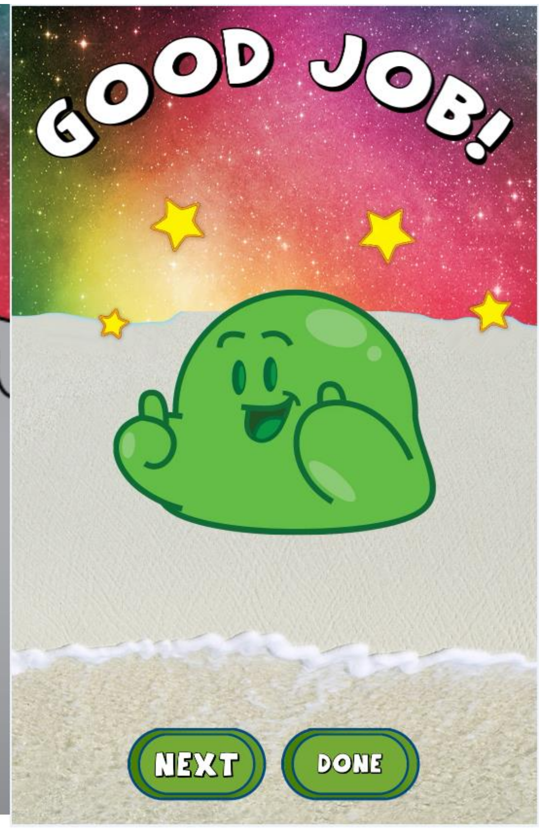
Below are some statements about the app you used during the REACH group. There is no right or wrong answer. Just circle the number that best describes what you think. Okay?

HOME SCREEN:

Not at all		Somewhat								Very much
1. I liked the colors of the buttons.	1	2	3	4	5	6	7	8	9	10
2. I liked the background.	1	2	3	4	5	6	7	8	9	10
3. I liked the smiling Blob.	1	2	3	4	5	6	7	8	9	10
4. It was easy to find the Blob tricks.	1	2	3	4	5	6	7	8	9	10

BLOB:





	Not at all	Somewhat							Very much
5. I liked playing the Blob's tricks.	1	2	3	4	5	6	7	8	9	10
6. The Blob doing tricks made me work harder to get more tricks.	1	2	3	4	5	6	7	8	9	10
7. I liked the sounds the Blob makes.	1	2	3	4	5	6	7	8	9	10
8. I liked the way Blob moves around.	1	2	3	4	5	6	7	8	9	10
9. I liked the Blob saying "Good Job!"	1	2	3	4	5	6	7	8	9	10
10. The Blob telling me that I did a good job made me try it again.	1	2	3	4	5	6	7	8	9	10
11. Getting thumbs-up from the Blob made me work harder.	1	2	3	4	5	6	7	8	9	10

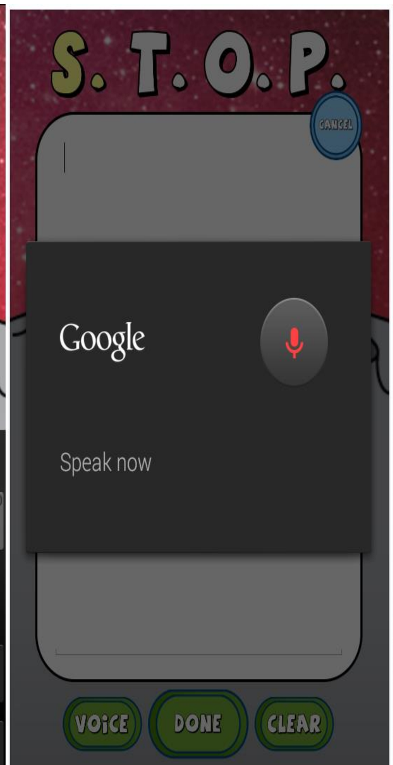
PROGRESS BARS:

	Not at all	Somewhat								Very much
12. It was easy to open the progress bar.	1	2	3	4	5	6	7	8	9	10
13. I liked the colors in the progress bar.	1	2	3	4	5	6	7	8	9	10
14. It was easy to check my progress using the bars.	1	2	3	4	5	6	7	8	9	10
15. The progress bar showed me how much work I did.	1	2	3	4	5	6	7	8	9	10
16. The progress bars made me practice.	1	2	3	4	5	6	7	8	9	10

RELAX:

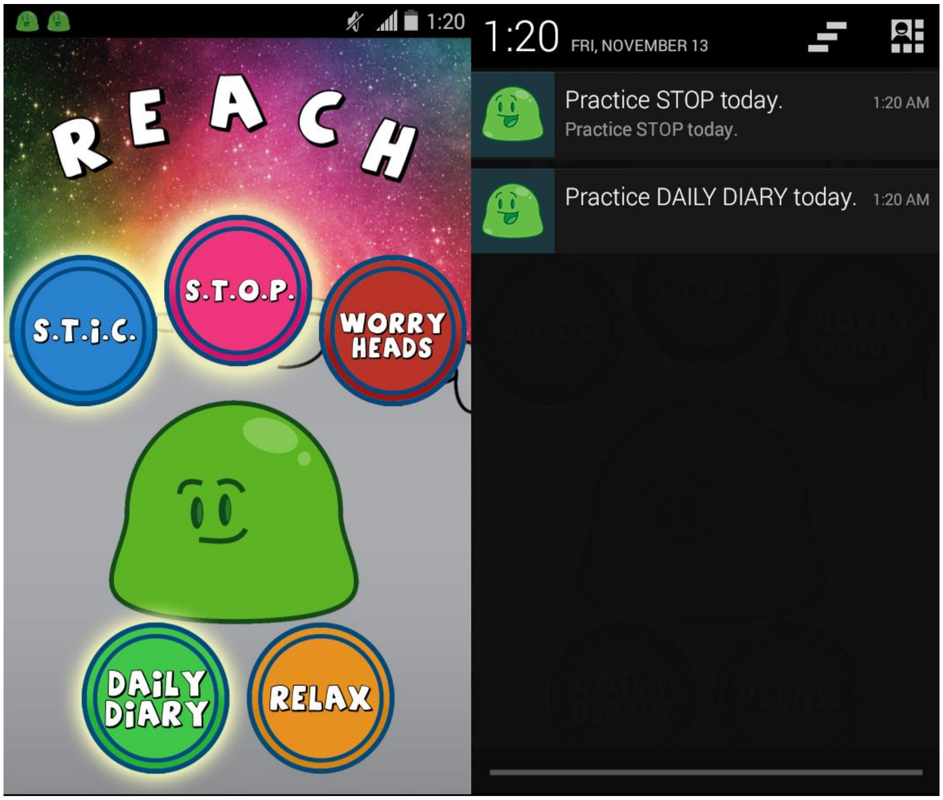
Not at all		Somewhat								Very much
17.I liked the beach background.	1	2	3	4	5	6	7	8	9	10
18.It was easy to listen to the relaxation.	1	2	3	4	5	6	7	8	9	10
19.I liked the dancing red crab.	1	2	3	4	5	6	7	8	9	10

APP TOOLS:



	Not at all	Somewhat								Very much
20. It was easy to follow the steps in STOP.	1	2	3	4	5	6	7	8	9	10
21. It was easy to follow the steps in the Daily Diary.	1	2	3	4	5	6	7	8	9	10
22. The buttons made it easy to do the work.	1	2	3	4	5	6	7	8	9	10
23. Using the keyboard on the app was easy.	1	2	3	4	5	6	7	8	9	10
24. Using the microphone on the app was easy.	1	2	3	4	5	6	7	8	9	10
25. Using the numbers on the app to rate situations in the Daily Diary was easy.	1	2	3	4	5	6	7	8	9	10

MESSAGES:



	Not at all	Somewhat								Very much
26. The Blobs lined-up on the top of the screen reminded me to do the work.	1	2	3	4	5	6	7	8	9	10
27. I liked the glowing buttons.	1	2	3	4	5	6	7	8	9	10
28. The glowing buttons helped me figure out what to practice.	1	2	3	4	5	6	7	8	9	10
29. The glowing buttons made me practice more.	1	2	3	4	5	6	7	8	9	10
30. The reminders I see on the right side, helped me practice.	1	2	3	4	5	6	7	8	9	10
31. I liked the Blob telling me what to do.	1	2	3	4	5	6	7	8	9	10

ABOUT THE APP:

Not at all		Somewhat								Very much
5. I was able to use the app on my own without any help.	1	2	3	4	5	6	7	8	9	10
6. I would want to continue working with the app.	1	2	3	4	5	6	7	8	9	10