Designing Self-Monitoring Technology to Promote Data Capture and Reflection

Eun Kyoung Choe

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

Doctor of Philosophy

University of Washington

2014

Reading Committee:

Julie A. Kientz, Co-Chair

Wanda Pratt, Co-Chair

Carol A. Landis

Sheelagh Carpendale

Program Authorized to Offer Degree:
The Information School

University of Washington

Abstract

Designing Self-Monitoring Technology
to Promote Data Capture and Reflection

Eun Kyoung Choe

Co-Chairs of the Supervisory Committee:

Assistant Professor Julie A. Kientz
Human Centered Design and Engineering

Professor Wanda Pratt
The Information School

Self-monitoring is a powerful means for self-reflection, which is important in health behavior change. More recently, researchers and companies began to offer numerous self-monitoring technologies for health. However, people—even experienced self-trackers such as Quantified-Selfers—have difficulty with continued tracking and gaining insights from their personal data. The goal of this dissertation is to provide insights on designing effective self-monitoring technology. In particular, I examine ways to support easy capturing of data and create persuasive feedback to nudge people toward positive behaviors.

As part of this research, I introduce a mobile self-monitoring technology, SleepTight, a lightweight self-monitoring application widget that helps people capture and reflect on sleep behaviors. The SleepTight system was designed based on two formative studies and theories of reactivity in self-monitoring research. I leveraged the Android platform's lock screen and home screen widgets to lower the capture burden and increase awareness. I conducted a 4-week deployment study to evaluate the efficacy of the SleepTight system and found that SleepTight's widgets served as visual reminders and helped participants collect more data, more accurately.

Participants were also able to reflect on sleeping patterns and relationships among the captured factors.

I further examine effective ways to provide self-monitoring feedback by leveraging the Framing effects with an aim to nudge people toward positive health behaviors. The goal of self-monitoring is not simply to quantify individuals' behaviors, but to improve it. Therefore, self-monitoring feedback needs to convey information to help people make health-enhancing, self-beneficial decisions. To identify the type of framing that could best convey self-monitoring feedback, I conducted an online experiment to test the effects of three types of performance feedback framing—(1) valence, (2) presentation type, and (3) data unit—on individuals' self-efficacy. I identified that it is better to use a positive framing with data units (e.g., raw data, rate, percentage) that can increase the perception of one's performance capabilities to enhance individuals' self-efficacy. This work provides empirical guidance for creating influential, persuasive performance feedback, thereby helping people designing self-monitoring technologies to promote healthy behaviors.

In this dissertation, I discuss how we can successfully design self-monitoring technology to help people collect data easily, learn their behavioral patterns, and develop positive changes for improving health. Effective self-monitoring technology eases the capture burden, supports customization, prevents backfilling, and provides feedback in a positive light. Self-monitoring technology that adheres to these guidelines can enhance tracking adherence, data accuracy, data awareness, self-reflection, and self-efficacy. I verify these thesis statements through a mixed-method approach including formative studies, technology deployment study, and experimental study. This dissertation research expands our knowledge of how consumer health information technology should be designed to support self-monitoring and reflection. Once a motivated individual meets a well-designed self-monitoring technology, exciting possibilities will arise for gaining insights for health, wellness, and other aspects of life.

Dedication

To my parents, Chung Lim Lee & Sang Hoon Choe,
for their continued support and love

Table of Contents

List of Tables

List of Figures

Glossary

Term	Definition
App widget	A miniature application view that can be embedded in other applications (such as the home screen or lock screen) and receive periodic updates. Also known as a “widget.”
Framing (in psychology)	Logically equivalent alternatives portrayed in different ways.
Framing effect	A type of cognitive bias in which people react to a particular choice in different ways depending on whether it is presented as a loss or as a gain.
Performance feedback	Real-time feedback on a person's current status or progress provided by self-monitoring technology.
Personal Informatics	A class of systems that help people collect personally relevant information for the purpose of self-reflection and gaining self-knowledge.
Persuasive technology	Technology that is designed to change attitudes or behaviors of users through persuasion and social influence but not through coercion.
Prospect theory	A behavioral economics theory that describes the way people choose between probabilistic alternatives that involve risk, where the probabilities of outcomes are known. The theory states that people make decisions based on the potential value of losses and gains rather than the final outcome, and that people evaluate these losses and gains using certain heuristics.
Quantified Self	A Meetup community sharing personal stories of self-tracking. It is also used to describe the practice of self-monitoring.
Quantified-Selfer	A person who is a member of Quantified Self or who practices Quantified Self.
Reactive effect	The change in frequency of the behavior resulting from the practice of self-tracking, which often occurs in desirable, therapeutic directions. Also known as “reactivity.”

Self-efficacy	The measure of a person's belief in one's own ability to complete tasks and reach goals.
Self-monitoring	An individual recording the occurrences of his or her target behavior.
Self-monitoring technology	Technology that facilitates capturing of the occurrences of target behavior and provides feedback to help people increase awareness and self-reflection.
Sleep hygiene	A variety of different practices that are necessary to have normal, quality nighttime sleep and full daytime alertness.

Chapter 1
Introduction

For this research, I was motivated by the importance of self-monitoring in personal health management. Self-monitoring refers to an individual recording the occurrences of his or her own target behavior (Nelson & Hayes, 1981). Self-monitoring induces ***positive reactivity effect***—the change in frequency of the behavior resulting from the practice of self-tracking, which often occurs in desirable, therapeutic directions (Kopp, 1988). People can become aware of their behaviors, draw meaningful inference from aggregated data, and identify unrecognized problems. When combined with other behavioral techniques, such as goal setting, self-monitoring becomes a powerful motivator toward behavior change (Kazdin, 1974; Mace & Kratochwill, 1985).

1.1. Motivation

In a 2013 nation-wide survey on people's health tracking practice, researchers reported that seven in ten U.S. adults tracked a health indicator for themselves or for a loved one (Fox & Duggan, 2013). However, among those who tracked one or more health indicators, only 21% used some form of technology for tracking while 49% kept track of progress "in their head" and 34% tracked data on paper (Fox & Duggan, 2013). Although this report showed people's high interests in self-monitoring, it also indicates a high level of unmet needs for supporting self-monitoring .

One of the promises of ubiquitous computing is to provide the ability to easily collect self-monitoring data in the wild. Mobile phones have already become a powerful personal health monitoring platform (Klasnja & Pratt, 2011). Moreover, many consumer wearable-sensing products are available in the market to help with collecting sleep, activity, heart rate, and other biobehavioral data. Self-monitoring data collected from these devices outside the doctor's office could help people monitor their health, notice abnormalities, set and achieve health goals, and maintain a healthy lifestyle. It could also help doctors accurately diagnose and track patients' progress over time. I call a class of tools that help people collect and reflect on personal data "***self-monitoring technology***" and define them as follows:

> ***Self-monitoring technology***: technology that facilitates ***capturing*** of the occurrences of ***target behavior*** and provides ***feedback*** to help people increase ***awareness*** and ***self-reflection***

To design an effective self-monitoring technology, I am concerned with ways to improve ***capturing data*** and ***providing feedback*** to enhance awareness and self-reflection.

Many self-monitoring technologies—both manual tracking tools (e.g., electronic sleep diary) and automatic sensing (e.g., fitbit[1]) tools—exist. Each type has its own advantages and drawbacks. Although manual tracking tools could increase self-awareness due to direct engagement with data collection (Bentley et al., 2013; Choe et al., 2014a), people are prone to forgetfulness. Automated sensing could reduce mental workload and increase data accuracy but are cumbersome to wear (in the case of wearable sensing) and could reduce awareness of the data collected (Li, 2009).

---

[1] fitbit is a wireless-enabled wearable device that measures activity, calories burned, and sleep. http://www.fitbit.com.

For my thesis, I examined ways to design self-monitoring technology that supports effective ***manual*** tracking. I chose to improve the manual tracking practice because self-monitoring's positive reactivity is enhanced when each occurrence of the target behavior is self-monitored and an obtrusive recording device is used (Korotitsch & Nelson-Gray, 1999). In manual tracking, ***people*** play an important role in entering data and making sense of it. Thus, manual tracking increases awareness of behaviors, but it also increases tracking burden. As a result, many people fail to manually track items for the long term despite the benefits. ***Sleep*** is an interesting yet challenging application area for manual tracking because many factors that are hard to automatically track could affect individuals' sleep quality. My goal for this research was to provide design guidelines on ways to design effective manual self-monitoring technology based on formative studies and empirical findings.

1.2. Thesis Statements

My thesis claims are summarized in the following statements:

> *It is important to understand the current practices and design space of self-monitoring to identify design opportunities for effective self-monitoring technology. Effective self-monitoring technology should be easy to access, quick to capture data, customizable, prevent backfilling, and provide feedback in a positive light. By adhering to these guidelines, self-monitoring technology can enhance tracking adherence, data accuracy, data awareness, self-reflection, and self-efficacy.*

1.3. Research Questions and Approaches

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

> **RQ1:** How do people currently practice self-monitoring?
>
> **RQ2:** What is the design space for sleep technologies?

RQ3: How should we design manual self-monitoring technology for capturing and reflecting on sleep behaviors to enhance tracking adherence, data accuracy, data awareness, and self-reflection?

RQ4: How should we design persuasive performance feedback that could enhance individuals' self-efficacy?

To answer **RQ1** and **RQ2** and understand the current landscape of self-monitoring technology, I conducted two formative studies. My goal was to identify challenges and opportunities for designing self-monitoring technology to promote healthy sleep behaviors. In addressing **RQ3,** I designed and developed *SleepTight*—manual self-monitoring technology to help people collect sleep and sleep-related behaviors—focusing on lowering the capture burden and increasing awareness and self-reflection. In addressing **RQ4**, I studied the effects of different feedback designs on self-efficacy.

1.3.1. Two Formative Studies

The first formative study was about understanding people's current self-monitoring practices. My approach was to study the group of experienced self-trackers—in this case, the Quantified Self[2] community. Quantified Self is the name of a community consisting of people who diligently track many kinds of personal data. I studied Quantified Selfers (Q-Selfers) to learn their motivations for tracking, tools for collecting and exploring data, gained insights, challenges they have, and workarounds to overcome the challenges. Although Q-Selfers' stories might not be generalizable or applicable to the broader population, their perspective could give us distinct insights because they have used existing technologies and spent numerous hours to build their own workarounds when facing problems. The results of this formative work helped me identify design opportunities and guidelines for self-monitoring technology.

The second formative study was about understanding the design space of technology for promoting healthy sleep behaviors. To examine the design space, my colleagues and I

[2] Quantified Self. http://quantifiedself.com

conducted contextual inquiry, survey, and interview studies. We found a broad interest in technologies for sleep, with a majority expressing interest in tracking sleep data over time. Simplicity, unobtrusiveness, and privacy were identified as crucial qualities of sleep technologies.

1.3.2. The SleepTight System

I designed, developed, and evaluated the SleepTight system (Figure 1) to study the efficacy of a low-burden, manual self-monitoring technology for capturing ***sleep and sleep-related behaviors***. The two formative studies and theories of reactivity in self-monitoring research informed the design of SleepTight. Tracking multiple factors, such as target behaviors (sleep) and contributing factors (daytime and nighttime activities), at the same time could help people assess their sleep behaviors holistically. However, existing sleep-monitoring tools rarely support capturing both target behaviors and contributing factors. Thus, people would miss vital information on how to improve the target behaviors. Moreover, many contributing factors are user-defined and individualistic, so people need to customize the items to be tracked. The need for flexible and customizable tracking tools makes a case for sleep as an appropriate application area for manual self-monitoring technology.

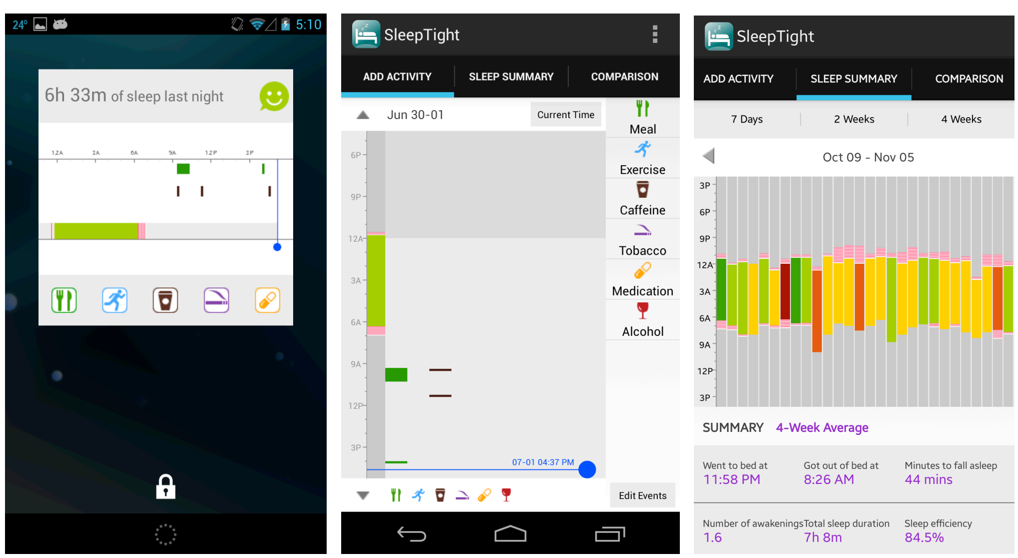


Figure 1. Components of the SleepTight system. Lock screen widget (left), Add Activity tab (middle) of the app, and Sleep Summary tab (right) of the app.

I leveraged the Android's widgets, which can be installed on home screen and lock screen to lower the capture burden and increase awareness. The lock screen widget, which is used to capture and present data, can be accessed without unlocking the phone. To further lower the capture burden, an activity is captured through a single-tap process: tap an icon from the widget or drag the time handle to specify time and tap an icon from the app. To prevent backfilling, SleepTight has a time limit to enforce people to log daily sleep diary before midnight. Once people enter sleep diary and sleep-related factors, SleepTight presents the feedback on the widgets as well as on the app's detailed summary and comparison pages. It shows aggregated sleep summary and frequency and timestamps of behavioral factors across individuals' sleep quality. I evaluated the SleepTight system through a 4-week field deployment study with 22 individuals and showed its efficacy on enhancing capturing adherence, data accuracy, awareness, and self-reflection.

1.3.3. Persuasive Performance Feedback

Self-monitoring technology has varied ways of providing feedback on people's progress. ***Framing effects*** from the behavioral economics literature (Tversky & Kahneman, 1981) suggest that how information is framed affects the persuasiveness of a communication. However, we have a dearth of knowledge of the framing effects on the feedback that self-monitoring technology provides. With an aim to create influential, persuasive performance feedback that will nudge people toward healthy behaviors, I identified three types of framing—namely, valence of performance, presentation type, and data unit. These framings can be applicable in presenting performance feedback in conjunction with a daily goal. I studied the effects of these framings on people's self-efficacy of goal achievement using a hypothetical scenario of a person receiving his or her daily step count from a pedometer. This work provides empirical guidance for creating persuasive performance feedback, thereby helping people design self-monitoring technologies to promote healthy behaviors.

1.3.4. A Summary of Research Questions and Approaches

Table 1 summarizes my research questions and the studies I conducted to answer them.

Table 1. A summary of research questions and studies to address the research questions.

#	Research Question	How I Addressed It
1	How do people currently practice self-monitoring?	• Formative study—video analysis of experienced self trackers (Chapter 4)
2	What is the design space for sleep technologies?	• Formative studies—a survey with general public, contextual inquiry with sleep clinicians, interviews with people who are interested in improving sleep behaviors using technology (Chapter 5)
3	How should we design manual self-monitoring technology for capturing and reflecting on sleep behaviors to enhance tracking adherence, data accuracy, data awareness, and self-reflection?	• Design, development, and deployment study of SleepTight (Chapter 6)
4	How should we design persuasive performance feedback that could enhance individuals' self-efficacy?	• Online experiment—a scenario-based survey to evaluate three types of performance feedback (Chapter 7)

1.4. Contributions

In this dissertation, I present three types of contributions—artifact, empirical finding, and methodology. Specifically, my research resulted in the following outputs:

1. Design guidelines for self-monitoring technology based on both empirical findings (video analysis of experienced self trackers) and theory. These design guidelines on self-monitoring technology include: (1) helping people determine what to track at the beginning of tracking practice, (2) supporting people in conducting self-experimentation with rigor, (3) maximizing the benefits of manual tracking, and (4) promoting self-reflection.

2. A design framework for mapping the design space of technologies to encourage and support healthy sleep behaviors based on empirical evidence (survey, contextual inquiry, and interview). Promising research opportunities for human-computer interaction (HCI) research include: (1) in-home tracking technology for long-term sleep behaviors, and (2) persuasive technology for healthy sleep behavior.

3. Design, implementation, and evaluation of mobile-based manual self-monitoring technology for tracking sleep and sleep-related behaviors. I showed that direct

engagement with data collection and exploration can enhance tracking adherence, data accuracy, self-awareness, and self-reflection.

4. Design guidelines for creating influential, persuasive performance feedback that can enhance self-efficacy, thereby helping people designing self-monitoring technologies to promote healthy behaviors. I showed that a positive framing presented with data unit (e.g., raw data, rate, percentage) that can increase the perception of one's performance capabilities enhances individuals' self-efficacy.

1.5. Dissertation Overview

This dissertation is divided into eight chapters.

In Chapter 2, Theories, I provided the theoretical underpinnings of my thesis research. I highlighted aspects of the theories that are most relevant to both the design of the SleepTight system and persuasive performance feedback in general. In particular, I covered the theoretical background of self-monitoring and framing effects.

In Chapter 3, Related Work, I summarized several empirical investigations of self-monitoring in HCI research and consumer self-monitoring technology. I also provided background on sleep and sleep-monitoring technology.

Chapter 4, Understanding Quantified-Selfers' Practices in Collecting and Exploring Personal Data, forms the first part of the formative work, in which I aimed to understand the current landscape of self-monitoring practices. I reported on qualitative and quantitative findings based on the analysis of 52 Quantified-Selfers' presentations on their self-monitoring practices.

Chapter 5, Opportunities for Computing Technologies to Support Healthy Sleep Behaviors, forms the second part of the formative work, in which, I aimed to identify the design considerations, challenges, and opportunities for using computing to support healthy sleep behaviors.

In Chapter 6, Design and Evaluation of the SleepTight System, I illustrated the design and evaluation study of a self-monitoring technology for collecting and reflecting on sleep behaviors. I reported on qualitative and quantitative results that show how SleepTight helped people easily collect sleep behaviors and reflect on them.

In Chapter 7, Persuasive Performance Feedback: The Effect of Framing on Self-Efficacy, I illustrated three different types of framing that are applicable to presenting performance feedback. Based on an online experimental study, I reported on the characteristics of persuasive performance feedback that can increase people's self-efficacy, thereby nudging people toward positive behaviors for health.

In Chapter 8, Contributions and Opportunities for Future Work, I summarized the contributions of my thesis and opportunities for future work.

Chapter 2
Theories

In this thesis research, I was motivated by several theoretical frameworks with a premise that theory-driven designs are effective in designing interventions to change people's health behaviors and attitudes. Michie et al. (2008) discussed the following three reasons for advocating the use of theory in designing interventions:

> *"First, interventions are likely to be more effective if they target causal determinants of behavior and behavior change; this requires understanding these causal determinants, i.e. theoretical mechanisms of change. Second, theory can be tested and developed by evaluations of interventions only if those interventions and evaluations are theoretically informed. Third, theory-based interventions facilitate an understanding of what works and thus are a basis for developing better theory across different contexts, populations, and behaviors."* (Michie et al., 2008)

Although technologies designed with an intent to change people's behaviors are a relatively new concept (Fogg, 2003), theories on health behavior change have long been studied in the field of psychology. Behavior change theories specify determinants that influence people's behavior, explain relationships among the determinants, provide reliable assessment measures, have predictive validity, and indicate how changes of determinants can lead to behavior change.

Thus, it is critical for the HCI community to learn theoretical foundations from other fields to design better health technologies with a deeper understanding of why and how people change their attitudes and behaviors. The challenge in adopting theories in design, however, is that many of the theories do not provide helpful insights as to *how* to design theory-based technology. Moreover, researchers often design technologies that incorporate a combination of various behavior change techniques and theories. This combination makes evaluation challenging as it is difficult to determine which design elements contributed to bringing about the intended behavior outcome (Klasnja, Consolvo, & Pratt, 2011). Recognizing the ambiguity and complexity of applying theories in design, Hekler et al. (2013) identified three ways of using theories in HCI research. First, theories can be used to generate “design hypotheses” (as opposed to “design guidelines”). Design hypotheses require subsequent testing to validate the efficacy of technology that embodies the design hypotheses. Second, theories can be used to guide the evaluation strategies of behavior change technologies, for example, by delineating conditions and defining measures. Lastly, theories can be used to select target users. This case is exemplified by the use of the transtheoretical model (Prochaska et al., 1992) to screen individuals in a certain phase of the model.

In what follows, I summarize the theoretical background of self-monitoring and framing effects. I then clarify how I used each theoretical background in my own work in light of Hekler et al.’s suggestion on three ways of using theories in HCI research.

2.1. Self-monitoring

Self-monitoring refers to an individual recording the occurrences of his or her own target behavior (Nelson & Hayes, 1981). The types of information that self-monitoring collects can be diverse, ranging from subjective information (e.g., a problem, situation, symptom, or disruption that symptoms may produce, inner thoughts or feelings) to objective information (e.g., frequency or intensity of a behavior under observation). Clinicians value self-monitoring because it could be used for all stages of assessment, such as diagnosis, target behavior selection for treatment, functional assessment, and treatment monitoring.

Theoretical work related to self-monitoring extends back to the 1970s. The depth of theoretical work from psychology provides HCI and Ubiquitous Computing (UbiComp) researchers with enormous insights into how to design effective self-monitoring technology. However, although many HCI and UbiComp researchers have implemented some aspects of self-monitoring, it is unfortunate that many of the self-monitoring technologies designed today are not grounded on lessons from the past. Therefore, it is critical for self-monitoring technology designers to understand theoretical foundations and implications of self-monitoring literature focusing on the determinants that contribute to positive behavior change. Thus, the focus of this literature review is to identify design opportunities for HCI and UbiComp research from insightful findings that previous self-monitoring research provides. In particular, Nelson (1977), Kopp (1988), McFall (1977), and Korotitsch & Nelson-Gray (1999) provided an excellent overview of self-monitoring, illustrating the factors thought to predict the accuracy as well as reactive effects of self-monitoring.

2.1.1. Purpose of Self-monitoring

Self-monitoring has been traditionally employed in clinical and research settings to serve an ***assessment function*** as well as a part of a ***treatment function*** within behavior therapy (Korotitsch & Nelson-Gray, 1999). While self-monitoring provides clinicians or therapists with data to assess an individual's progress, self-monitoring also results in changes in behavior under observation. ***Reactivity*** or ***reactive effect*** refers to the act of self-monitoring causing the change in frequency of that behavior. This reactivity often occurs in a desirable direction, which is why we consider the reactivity to be therapeutic in general. When self-monitoring is used as an assessment tool, it is important to enhance the accuracy of the data collected. When self-monitoring is used for a therapeutic purpose, implementing ways to enhance the reactive effects to maximize the therapeutic outcome is important. The more reactive a self-monitoring procedure is, the less suitable it is for assessment purposes, but the more attractive it becomes as a potential treatment component (McFall, 1977). However, the distinction is less important when self-monitoring is used for the purpose of self-management without therapist or clinician involvement. In this context, it is common to use self-monitoring for the purposes of assessment and treatment at the same time. In this case, a tool should help enhance data accuracy to better assess a person's state *and* maximize the reactive effects to gain a therapeutic outcome.

2.1.2. Process of Self-monitoring

Self-monitoring is composed of two steps: (1) discriminating the occurrence of the behavior and then (2) systematically recording the observation (Thoresen & Mahoney, 1974). Nelson emphasized that the person employing the self-monitoring must perform both of these behaviors to produce accurate self-recordings although performing just the first part—determining that the target behavior has indeed occurred—might result in reactive behavior changes (Nelson, 1977). Researchers designing self-monitoring technology have often attempted to automate both of these processes, which reduces the mental and physical burden of self-monitoring. However, if the purpose of self-monitoring is to enhance awareness and change behavior, automating the process of self-monitoring might not bring about the intended results of reactivity unless extra cues and feedback are built into the system.

Passive self-monitoring refers to self-monitoring tools that automatically collect self-monitoring data so that people need not make a self-recording response after each occurrence of the target behavior. An example of passive self-monitoring is a cigarette case that automatically records the number of times it is opened (Azrin & Powell, 1968). However, it is not yet known whether such passive procedures produce reactive behavior changes over the long term. On the other hand, if the self-monitoring device is too cumbersome or the process is too difficult (e.g., self-monitoring a high-frequency or nearly continuous behavior), people might cease to self-monitor.

2.1.3. Accuracy of Self-monitoring Data

The accuracy of self-monitoring data matters when self-monitoring is used for all stages of assessment—for diagnostic-descriptive assessment, for functional assessment, or for treatment monitoring (Korotitsch & Nelson-Gray, 1999). Previous studies have identified a number of variables that are shown to influence the accuracy of self-monitoring data. Based on Nelson

(1977), McFall (1977), and Korotitsch & Nelson-Gray (1999), I provide the following list of factors that are known to affect the accuracy of self-monitoring:

1. **Awareness of accuracy assessment**: more accurate data is captured when self-recorders are aware that their accuracy is being overtly monitored (e.g., reliability checks by a trained observer) than when their accuracy is monitored covertly.

2. **Reinforcement on accurate self-recorded data**: more accurate data is captured when positive reinforcement is provided.

3. **Commitments**: compliance with self-monitoring procedures is enhanced if verbal commitments or contracts are made.

4. **Concurrent response requirements**: accuracy is lower when self-recorders are required to engage in other responses concurrently.

5. **Schedule of self-monitoring**: more accurate data is captured when the data is recorded closer in time to the actual target behavior.

6. **Valence of target behavior**: accuracy is lower for negatively valenced (inappropriate) behaviors than for positively valanced (appropriate) behaviors.

7. **Training**: training (e.g., giving explicit definitions of target behaviors) improves the accuracy of self-monitoring data.

Studies showed that long-term manual tracking is a challenging task—adherence to manual tracking is typically low while backfilling (i.e., batch completing) is common (Stone et al., 2003), and this backfilling risks data accuracy. In this light, the previous list provides us with design insights for enhancing data accuracy for manual tracking. For example, to enhance data accuracy, self-monitoring technology should support near real-time capture (#5). The valence of the target behavior should be positively framed (#6). Providing positive reinforcement (#2) and supporting accountability (#1) might also help enhance data accuracy.

2.1.4. Reactivity of Self-monitoring

Reactive effects can occur in the therapeutic and desirable direction, and therefore provide benefits toward behavior change (Korotitsch & Nelson-Gray, 1999; Mattila et al., 2008). To design self-monitoring technology for the purpose of treatment, technology designers must understand what factors enhance positive self-monitoring reactivity. The reactive effects can be explained through self-monitoring's feedback loop and self-regulatory behavior. Self-monitoring requires an individual to deliberately attend to his or her behavior and to become aware of whether the behavior departs from a normatively or personally acceptable standard of performance (Kazdin, 1974). Kanfer posited that when one's behavior departs from the standard, a self-regulatory process is triggered (Kanfer & Gaelick-Buys, 1991). When the departed behavior is brought back to an acceptable range, the self-regulatory process is ceased. Understanding the determinants of self-monitoring outcomes provides insightful implications as to how to design self-monitoring technology to draw the most effective behavior change. Below is a list of factors that influence the occurrence, magnitude, or direction of reactive effects associated with self-monitoring.

1. **Motivation**: those who are already motivated to change behavior are more likely to show reactive effects when they monitor their behavior. On the other hand, for those who are not motivated to change behavior, self-monitoring might have no effect or could even backfire on them. This effect was studied in the context of cigarette smoking: when people are not motivated to stop smoking, self-monitoring of the number of cigarettes smoked increases smoking rates (McFall, 1970) or does not show overall change (Lipinski, Black, Nelson, & Ciminero, 1975).

2. **Valence of target behavior**: the valence of target behavior could determine the direction of the reactive changes. Positively valenced behaviors are desirable behaviors that people want to increase and negatively valenced behaviors are undesirable behaviors that people want to decrease. Neutrally valenced behaviors are neither positively nor negatively valued (Kopp, 1988). In general, self-monitoring

increases the frequency of positively valenced behaviors, decreases the frequency of negatively valenced behaviors, and induces no change for neutrally valenced behaviors (Kazdin, 1974; Sieck & McFall, 1976).

3. **Goals, reinforcement, and feedback**: providing a performance goal, feedback to self-monitoring, and a reinforcement that is contingent on behavior augment the reactive effects (Kazdin, 1974).

4. **Number of behaviors concurrently self-monitored**: although self-monitoring of multiple behaviors results in reactive effects, simultaneously recording more than one behavior reduces the reactive effects compared to recording a single behavior (Hayes & Cavior, 1977). However, the behaviors under observations in Hayes and Cavior's study were not related to each other (face touching, non-fluencies, and value judgments). If there are clear relationships between target behaviors and a single outcome measure of overriding concern, the combined effects could be greater. For example, Romanczyk found that self-monitoring produced greater weight loss when both daily weight and caloric intake were monitored than if only weight was self-recorded (Romanczyk, 1974).

5. **Nature of the Target Behavior**: how the target behavior is defind could affect the reactivity. For example, instructions to record the number of resisted urges to smoke decreased smoking rates, whereas instructions to record the number of cigarettes smoked increased smoking rates (McFall, 1970). McFall (1970) presumed that tracking each smoking act might be the equivalent of recording negative events, or failure experiences, and thus could be aversive; whereas recording each instance of resistance to temptation (urges) might be equivalent to recording successful experiences, and thus be positively reinforcing.

6. **Nature of the self-monitoring device**: while small unobtrusive recording devices are recommended in general, Nelson et al. (1978) reported that obtrusive devices (e.g., a wrist counter, slip, toothpick) could become a cue or discriminative stimulus for the

behavior under observation and produce greater reactivity than a less obvious device. For example, Nelson et al. observed greater reactivity when participants used obtrusive hand-held counters to self-record appropriate classroom verbalizations than when they used less-obtrusive belt-worn counter (1978).

To summarize, reactivity is more likely to occur when (1) people are motivated to change behavior; (2) people are given a performance standard, goal, or feedback; (3) correspondingly positive or negative valence is assigned to the behavior under observation; (4) a single behavior is monitored; (5) multiple related behaviors are monitored; (6) positive events (e.g., resistance of negative urges) rather than negative events are recorded; and (7) the recording devices are moderately obtrusive.

2.1.5. Accuracy and Reactivity

One of the most striking findings to HCI and UbiComp researchers from the prior self-monitoring research is that self-monitoring data need not be accurate for reactive effects to occur. Broden et al. (1971), Fixsen et al. (1972), Hayes & Cavior (1977), Herbert & Baer (1972), and Lipinski & Nelson (1974) showed that self-monitoring produces reactivity even though self-monitoring data is inaccurate or unreliable compared to external observations. These findings suggest that inaccurate monitoring could have therapeutically valuable effects. However, if the self-monitoring has a dual purpose of promoting change (therapeutic purpose) and gathering data (assessment purpose), then accuracy is important. Nelson (1977) reported that researchers do not routinely measure both behavior changes produced by self-monitoring and the quality of self-recorded data. Identifying the factors influencing both accuracy and reactivity warrants further research efforts, and we need to reinterpret the relationship between data accuracy and reactivity in the light of the advancement of mobile and sensing devices where data accuracy could be highly enhanced and various feedback could be delivered to the self-trackers.

2.1.6. Implications for Design

Designers of self-monitoring technology should be concerned with ways to enhance both data accuracy and reactive effect so that people can accurately assess their current state and gain therapeutic effects at the same time. In this regard, prior research on ways to enhance data accuracy and reactive effects motivated me to examine various ways to leverage these findings through technology. For example, implementing reminders to help people record activities closer in time to the actual target behavior could enhance data accuracy. Helping people track multiple *related* factors could result in high reactivity. The *valence of target behavior* affects both data accuracy and reactivity. Because technology solutions could induce a new problem, the embodied technology solution drawn from theoretical insights should be implemented and tested before those insights should be claimed as "design guidelines."

2.2. Framing Effects

Self-monitoring technology should provide feedback in a way that nudges people toward positive behaviors. Framing Effects literature is one line of research that provides numerous insights about how to provide information in a persuasive manner. I provide a summary of the theory and discuss how the theory can be used to design persuasive self-monitoring feedback.

2.2.1. Prospect Theory

Tversky & Kahneman (1981) revealed that presenting the same option but varying the framing of acts, contingencies, or outcomes alters people's decisions. Tversky & Kahneman proposed *Prospect Theory* to explain the framing effects, stating that people have an irrational tendency to be less willing to take a risk with profits than with losses. In other words, people value a sure gain over a probable gain with an equal or greater expected value. In contrast, people prefer a probable loss to a smaller loss that is certain when focusing on the prospect of a loss.

Since Tversky & Kahneman first explained how valence framing influences people's willingness to take risk, framings have been studied in many domains. To better understand when and why different types of framing will have an effect, Levin and colleagues developed a typology of framings and distinguished between three different kinds of framings—risky choice, attribute, and goal framing (Levin, Schneider, & Gaeth, 1998). My thesis work was particularly inspired by the *attribute framing*, which affects the evaluation of an object or event characteristic. An example of attribute framing is how we describe the attribute of ground beef, which can be labeled as either "75% lean" or "25% fat." A study showed that people favor the former even though the two labels convey the same information (Levin & Gaeth, 1988). The most common finding in the attribute framing literature is that positive framing leads to more favorable evaluations than negative framing (Levin & Gaeth, 1988; Marteau, 1989).

2.2.2. Prospect Theory in Health Domains

In the health domain, Prospect Theory has been used to understand health-relevant judgment and behaviors. Rothman & Salovey (1997) classified health behaviors into (1) detection behaviors, (2) prevention behaviors, and (3) recuperative behaviors and stated that the influence of framed information on decision-making is contingent on the degree to which performing a health behavior is perceived as risky. For example, early detection behaviors (e.g., mammography, HIV testing) could be regarded more as "uncertain or risky" behaviors than prevention behaviors (e.g., smoking cessation, exercise) because there is a possibility of discovering that one is sick. In accordance with Prospect Theory, empirical studies showed that loss-framed messages tend to be more persuasive for promoting detection behaviors whereas gain-framed messages tend to be more persuasive for promoting prevention behaviors (Krishnamurthy, Carter, & Blair, 2001). Therefore, in the case of promoting physical activities (e.g., walking), for example, these findings suggest the use of gain-frame (i.e., emphasizing the benefits of physical activities) because conducting physical activities is considered as a prevention behavior. However, prior research does not address how to best present *daily near real-time performance feedback* that can lead to health-enhancing, self-beneficial decisions.

Framing is also used in a broad sense, such as varying the *presentation type* or *data unit*. Lipkus & Hollands (1999) and Ancker et al. (2006) provided an extensive review of literature around the use of visuals to enhance health risk communication. Although some visuals help reduce the amount of mental computation, the authors argued that not all graphics are more intuitive than text. Ancker et al. (2006) found that the use of visuals should depend on the purpose of risk communication because some types of visuals are more appropriate for enhancing the accuracy of quantitative reasoning whereas others are more suitable for promoting behavior change. In addition, data presented with different units (e.g., raw data, rate, percentage) could have a significant impact on people's perception of the data. One study showed that rates (e.g., three per 1000) were easier to understand than proportions (e.g., one in 333) when patients were presented with the risk of genetic abnormalities (Grimes & Snively, 1999). This finding suggests that the choice of data unit in designing performance feedback could alter people's health decisions.

2.2.3. Implications for Design and Evaluation

Framing effects motivated me to examine different ways to frame the self-monitoring data. For example, manipulating the valence of a target behavior, data unit, and presentation type could influence people's decisions toward a health behavior. In addition, the framing effects literature provided insights into the design of an evaluation study to assess which framing is most effective in delivering self-monitoring feedback. I learned that self-efficacy is a reasonable proxy for an actual behavior (Bandura, 1990). Numerous studies have shown that self-efficacy can be enhanced and that this enhancement positively influences health behavior change (Bandura, 1990; 2006; Strecher, McEvoy DeVellis, Becker, & Rosenstock, 1986). Thus, the prior research suggests that different framings can influence people's self-efficacy and that we can identify a better framing by measuring self-efficacy.

2.3. Chapter 2 Summary

I covered the theoretical background of ***self-monitoring*** and ***framing effects***. Each theory informed various aspects of the SleepTight system and persuasive performance feedback.

Learning about the self-monitoring theories, I was motivated to design self-monitoring technology[3] that can enhance both reactivity and data accuracy. Manual tracking can reinforce the reactivity effect because it is more obtrusive than passive tracking. Although data accuracy is not that important for the reactivity effect to occur, manual tracking can risk data accuracy, but the risk can be reduced through incorporating the findings from the accuracy of self-monitoring literature.

Learning about the framing effects literature, I was motivated to identify types of framing that can best convey performance feedback[4] to enhance individuals' self-efficacy. I also learned that I should carefully choose the valence of a target behavior when designing self-monitoring technology because it could greatly influence people's self-efficacy and thus their health behavior choices.

The design choices I made for the SleepTight system were "design hypotheses," which were subsequently tested during the SleepTight evaluation study. On the other hand, the framing effects literature informed the persuasive performance feedback design as well as the evaluation study design[5].

---

[3] In Chapter 6, I explain the SleepTight system design and evaluation in detail.

[4] In Chapter 7, I explain how I leveraged the concept of Framing effects for designing and evaluating real-time performance feedback.

[5] Drawn from prior literature on framing effects, I designed a pair of complementary feedback designs and conducted a between-subjects design.

Chapter 3
Related Work

In this chapter, I provide an overview of related work in self-monitoring technology research. I begin with a note on terminology. I then describe research prototypes and commercial products in the area of self-monitoring technology for *health*, focusing on their target behavior, capture mechanism, and feedback mechanism. Because my work is particularly related to *sleep*, I summarize the literature on sleep as well as technology to promote healthy sleep behaviors.

3.1. Terminology

It is worthwhile to clarify the three similar terminologies that are used in self-monitoring research—(1) self-monitoring technology; (2) personal informatics; and (3) Quantified Self—all of which similarly denote technology that facilitates self-monitoring.

3.1.1. Self-monitoring Technology

In *Chapter 1, Introduction*, I defined self-monitoring technology as "***technology that facilitates capturing of the occurrences of target behavior and provides feedback to help people increase awareness and self-reflection***." According to this definition, technology that only facilitates capturing but does not provide feedback is not self-monitoring technology. Similarly, technology that provides feedback based on an individual's data that was collected without the person knowing is not self-monitoring technology. In this dissertation, I am mostly concerned with self-monitoring technology for *health*.

3.1.2. Personal Informatics

Li and colleagues (2010) coined the term personal informatics, which is defined by a class of systems that "***help people collect personally relevant information for the purpose of self-reflection and gaining self-knowledge.***" Two core aspects of personal informatics systems are ***collection*** and ***reflection***. Self-monitoring technology and personal informatics have similar definitions such that they both denote technology that aids with personal data collection and reflection. However, personal informatics covers a broader range of technology than self-monitoring technology. Li et al. argued that personal informatics includes having access to automatically collected data—such as bank statements, email history, credit card bills, and phone call history. Self-monitoring technology, on the other hand, does not include infrastructure that collects these types of data. Data collection from these personal informatics services is mandatory as opposed to voluntary, and the data is collected regardless of an individual's intention. Having said that, an application that uses the readily available information to create real-time feedback to help people reflect on their behavior is considered self-monitoring technology. In this sense, self-monitoring technology is a subset of personal informatics.

3.1.3. Quantified Self

Quantified Self (QS) refers to the name of a community of self-trackers. The community is composed of a diverse group of life hackers, data analysts, computer scientists, early adopters, health enthusiasts, productivity gurus, and patients. Believing in the notion of "self-knowledge through numbers," Wired Magazine editors Gary Wolf and Kevin Kelly created a blog called quantifiedself.com in 2007, which has become the repository for people to share self-monitoring practices. Sometimes, QS refers to a class of tracking technology similar to self-monitoring technology or personal informatics, but in this dissertation, I use the term Quantified-Selfers (Q-Selfers) to denote individuals who are experienced self-trackers.

The prevalence of low-cost monitoring sensors has accelerated the rise of the Quantified Self movement. Initially started in the Silicon Valley area among technology enthusiasts, QS has become an international community of people practicing self-monitoring and building self-monitoring technology. Quantified-Selfers promote sharing of self-monitoring practices through in-person Meetups (meetings), blogs, and annual conferences. As of January 2014, QS is an active, international community, with Meetups held in 106 cities in 36 countries. They have held an annual conference since 2011. Identifying health tracking as a promising area for growth, toolmakers of self-monitoring devices and software attend Meetups to promote their products and sponsor the annual QS conference.

3.2. Self-Monitoring Technology for Health

HCI researchers and designers have developed and studied many self-monitoring technologies in the domain of health and wellness (Klasnja & Pratt, 2011; Swan, 2009). Sensors have become smaller and better integrated with mobile devices, making it easy for people to track numerous types of data. Recognizing the power of self-monitoring in promoting health behavior change, researchers and designers often incorporate automated sensing or manual tracking features in designing self-monitoring technology.

According to the definition, each self-monitoring technology has one or more target behaviors of interest (e.g., amount of sleep, daily step count, time sitting on a desk chair). The target behaviors are recorded through different capture mechanism. The feedback is provided to increase awareness and encourage self-reflection. Self-monitoring technology consists of three design dimensions—(1) target behavior, (2) capture mechanism, and (3) feedback mechanism. I provide key prior work in the area of self-monitoring technology for health according to the three design dimensions. I summarize these technologies in Table 2 and describe them in detail in the following three subsections.

Table 2. A summary of self-monitoring technology from HCI research and commercial products.

	Project / Product Name	Self-Monitoring Technology Dimensions		
		Target Behaviors	**Capture Mechanism**	**Feedback Mechanism**
Commercial Products	Fitbit (fitbit)	Sleep (sleep states) Physical Activity	**Wearable sensing** (accelerometer) **Manual tracking**	Mobile phone + website + wearable device (chart, count, visual)
	Jawbone UP (Jawbone UP)	Sleep (sleep states) Physical Activity	**Wearable sensing** (accelerometer) **Manual tracking**	Mobile phone (chart, count)
	ZEO	Sleep (sleep duration, sleep quality, sleep phase)	**Wearable sensing** (EEG sensor)	Stand-alone + website (chart, count)
	SleepCycle (Sleep Cycle)	Sleep (sleep duration, sleep phase)	**Embedded sensing** (Mobile phone)	Mobile phone (chart, count)
Research Projects	BuddyClock (Kim et al., 2008)	Sleep (sleep states)	**Manual tracking**	Stand-alone (visual)
	Somnometer (Shirazi et al., 2013)	Sleep (sleep quality, sleep states)	**Manual tracking** **Embedded sensing** (Inferring sleep duration from alarm clock use)	Mobile phone (chart, count) Facebook
	Sleepful app (Lawson et al., 2013)	Sleep (sleep state)	**Manual tracking** (user initiates start/stop)	Sleep efficiency was provided, but the paper did not describe the feedback mechanism
	SWP (Chen et al., 2013)	Sleep (sleep duration)	**Embedded sensing** (Inferring sleep duration from natural smartphone use)	Mobile phone (count)
	Lullaby (Kay et al., 2012)	Sleep (sleep state) Environmental sleep disruptors (temperature, light, motion, audio, video)	**Wearable sensing** **Embedded sensing**	Stand-alone (chart, count, visual)
	Health Mashups (Bentley et al., 2013)	Sleep, Physical Activity (step count), weight, location, weather, calendar, food, mood, pain	**Wearable sensing** **Manual tracking** **Automatic feed**	Mobile phone (chart, text)
	MONARCA (Bardram et al., 2013)	Sleep, Physical activity, mood, subjective activity, medicine adherence, optional self-assessment items (alcohol, universal warning signs, early warning signs, stress, free text)	**Manual tracking** (mood, sleep, alcohol) **Embedded sensing** (activity monitor by sampling sensor data from the phone)	Mobile phone + website (chart, count)
	Wellness Diary (Mattila et al., 2008)	Weight, step counts, exercise, feeling, fat%, blood pressure, stress, sleep, health event (people chose the variables they liked to track)	**Manual tracking** **Wearable sensing** (Pedometer) **Stand-alone device** (Blood pressure meter)	Mobile phone (chart)
	Fish'n'Steps (Lin et al., 2006)	Physical activity (step count)	**Wearable sensing** (Pedometer) **Manual uploading**	Public kiosk + website (visual)
	Houston (Consolvo et al., 2006)	Physical activity (step count)	**Wearable sensing** (Pedometer) **Manual uploading**	Mobile phone (count)
	Shakra (Anderson et al., 2007)	Activity level (stationary, walking, driving)	**Wearable sensing** (Mobile phone)	Mobile phone (chart, visual)
	Ubifit (Consolvo et al., 2008)	Physical activities (cardiovascular exercise, strength training, etc.)	**Wearable sensing** (Mobile sensing platform—Choudhury et al., 2008). **Manual tracking**	Mobile phone (visual, count)
	RecoFit (Morris et al., 2014)	Physical activities (weight training, calisthenics)	**Wearable sensing**	Exercise type and counts were provided, but the paper did not describe the feedback mechanism

3.2.1. Target Behaviors

In 2013, sixty percent of U.S. adults said that they tracked their weight, diet, or exercise routine (Fox & Duggan, 2013). One-third of U.S. adults tracked health indicators or symptoms, like blood pressure, blood sugar, headaches, or sleep patterns (Fox & Duggan, 2013). People usually had a specific reason to track these behaviors—such as to cure or manage a health condition (e.g., track blood glucose to hit the target range), achieve a goal (e.g., track weight to get back to the ideal weight), find triggers (e.g., log potential triggers for a food allergy), or answer a specific question (e.g., what is the right dosage of medications?) (Choe et al., 2014a). However, novice trackers found it challenging to know exactly what to track in the first place to meet their needs; thus, novice trackers made the mistake of missing essential items that they needed to track to achieve their tracking goal (Choe et al., 2014a). What people track is constrained by the functionalities technology provides.

Because my dissertation is about designing sleep-monitoring technology, I cover key commercial products and HCI research projects that were developed for the purpose of tracking sleep in Table 2 (Bardram et al., 2013; Bentley et al., 2013; Chen et al., 2013; *fitbit*; *Jawbone UP*; Kay et al., 2012; Kim et al., 2008; Lawson et al., 2013; Mattila et al., 2008; Shirazi et al., 2013; *Sleep Cycle*). In addition, physical activity has long been a popular topic for self-monitoring research in HCI, so I include 5 key research projects that were designed for the purpose of tracking physical activity (Anderson et al., 2007; Consolvo et al., 2006; Consolvo et al., 2008; Lin et al., 2006; Morris et al., 2014).

Sleep. When tracking sleep, the target behavior could mean a number of different things. In general, sleep-monitoring technology captures ***objective sleep measures***—such as to-bed time, wake-up time (sleep state), and the number of awakenings during the sleep. Based on this information, sleep duration and sleep efficiency can be calculated. In addition, some self-monitoring technology can track different sleep phases (such as deep sleep, REM sleep, light sleep). ZEO[6] has been cross-validated with polysomnography (PSG)[7] where the agreement between the two measurements was about 75% (Shambroom & Fabregas, 2012). Some sleep-monitoring technology also provides an ability to track ***subjective sleep quality***. Somnometer (Shirazi et al., 2013) allows people to manually rate subjective sleep quality while the duration is estimated based on tracking a person's explicit interactions with the mobile app. To have a holistic perspective on a person's sleep, sleep-monitoring technology should help capture both objective sleep measures as well as subjective sleep quality.

Physical activity. What I included in Table 2 are just a few among numerous research projects and commercial products that have been developed in the design space of physical activity tracking. Interestingly, as Table 2 shows, many accelerometer-based technologies capture both sleep and physical activity at the same time.

Researchers use daily step count as a measure of physical activity. Health Mashups (Bentley et al., 2013), Wellness Diary (Mattila et al., 2008), Houston (Consolvo et al., 2006), and Fish'n'Steps (Lin et al., 2006) leverage existing accelerometer-based pedometers to incorporate step counts to show feedback of people's daily physical activity. Commercial pedometers—such as fitbit and Jawbone UP—are also accelerometer-based pedometers that can infer calories burned based on the step count tracking data. Ubifit (Consolvo et al., 2008) employs the Mobile Sensing Platform (MSP), which can infer the type of physical activity (i.e., walking, running, cycling, elliptical training) from its 3-d accelerometer and barometer (Choudhury et al., 2008). More recently, Morris et al. (2014) developed RecoFit, a system for automatically tracking repetitive exercises—such as weight training—via an arm-worn inertial sensor.

---

[6] ZEO is one of the early pioneers of self-monitoring technology, but the company went out of business in 2013.

[7] Polysomnography is the gold standard for assessing sleep measures combining measures of electroencephalogram (EEG), eye movements (electrooculogram, EOG), muscle activity (electromyogram, EMG), and heart rhythm (electrocardiogram, ECG). Patients suspected of having a sleep disorder come to sleep lab to conduct polysomnography. However, the process is costly, cumbersome, and does not provide an accurate picture of sleep habits in situ. They are therefore not ideal for daily basis sleep tracking over the long term.

3.2.2. Capture Mechanism

Capture mechanisms are concerned with ways in which self-monitoring technology captures its target behaviors. On one extreme, people use ***manual tracking*** tools (e.g., paper or electronic sleep diaries) for documenting self-report measures. Although manual tracking tools could increase self-awareness due to direct engagement with data collection, people are prone to forgetfulness. On the other extreme, people use sensing—either ***wearable*** or ***embedded***—to automatically track sleep or physical activity. These sensing tools have the potential to reduce mental workload and increase data accuracy, but can be cumbersome to wear. In addition, these automated tools reduce people's awareness of the data collected (Li, 2009). Most of the self-monitoring technologies in Table 2 employ more than one capture mechanism. For example, Fitbit and Jawbone UP employ wearable sensing (accelerometer) to track sleep and physical activity, but they also allow people to manually add food, mood, and water consumed. Similarly, Ubifit (Consolvo et al., 2008) and Wellness Diary (Mattila et al., 2008) allow people to manually add other physical activities that cannot be automatically detected. Somnometer employs a combination of manual tracking (for capturing subjective sleep rating) and embedded sensing (for capturing sleep duration).

Mobile phones, in particular, have become an important self-monitoring technology platform. Depending on the capture mechanism, a mobile phone can be categorized as a platform for (1) manual tracking, as in the case of Wellness Diary (Mattila et al., 2008), (2) wearable sensing, as in the case of Shakra (Anderson et al., 2007), or (3) embedded sensing, as in case of Somnometer (Shirazi et al., 2013) or SWP (Chen et al., 2013).

With regards to sleep tracking, researchers have examined ways to lower the capture burden by leveraging mobile phones and sensing technology. Some researchers used people's direct input to detect sleep measures. Shirazi et al. (2013) designed Somnometer, a social alarm clock on a mobile phone, which estimates sleep duration based on people's interaction with the use of the alarm clock app. Lawson et al. (2013) designed Sleepful app, which uses a stimulus response paradigm[8] to detect sleep and wakefulness state. I categorized Sleepful app as a manual tracking tool because a person needs to initiate and end the sleep tracking. Other researchers devised unobtrusive ways to infer sleep measures from people's normal interaction with the smartphone. For example, Chen et al. (2013) built models from smartphone embedded sensor data (e.g., total duration of phone-lock, phone-off, phone-charging, phone in darkness, phone in a stationary state, and phone in a silent environment) that could detect sleep duration (±42 minutes) in an unobtrusive way. This approach is very promising because they do not require special interventions to detect sleep measures. Although I share similar goals to lower the capture burden, I focus on lowering the manual capture burden while keeping the flexibility that manual capture provides (e.g., being able to define a new tracking item).

3.2.3. Feedback Mechanism

Feedback mechanisms are concerned with the ***device*** where the feedback is provided and the ***form*** of the feedback.

The devices include a mobile phone (glanceable display or native app), website, stand-alone device, wearable device, and public kiosk. Providing feedback on a mobile device is the most common way because often, a mobile device is used to capture a target behavior, which can be conveniently used to display feedback. A glanceable display shows feedback on a mobile phone's lock screen or home screen thereby lowering the access burden (Consolvo et al., 2008). Wearable devices that have a display (e.g., Fitbit) can provide real-time feedback, but the screen size is often so small that it can show only limited information. Websites are used for data exploration and sharing purposes (e.g., patients and doctors) (Bardram et al., 2013). ZEO, BuddyClock (Kim et al., 2008), and Lullaby (Kay et al., 2012) have a form factor of a stand-alone device, which comes with a display to operate the device and provide feedback. Fish'n'Steps projects people's step count data onto a public display for a broad sharing while the data is also uploaded to a website (Lin et al., 2006) for personal access.

[8] The basic concept of stimulous response paradigm is that is a person is awake, they will respond to the tone via a recording device whilst if they are asleep they will not.

The feedback can be provided in the form of count, text, chart, and visual. The choice of form depends on the display size. For example, Fitbit's small display can accommodate only counts and visual feedback, and thus, it shows more detailed information on its website using a combination of count, charts, and visuals. The Mobile Health Mashups system communicates the complex statistical relationships among nine tracking items using text (e.g., "You are happier on days when you sleep more") (Bentley et al., 2013). Charts (e.g., line chart, bar chart) are used to show aggregated feedback and trends, but they do not often provide the ability to explore data. Visuals convey positive and negative reinforcement. When a person achieves a goal, positive reinforcement conveyed with visuals can be provided to encourage positive behaviors. When a person fails to meet a goal, negative reinforcement can be given to discourage negative behaviors. For example, in Fish'n'Steps, individuals' step count is mapped to the growth and emotional state of a virtual pet. In Ubifit, people's physical activity level and goal achievement state are mapped to flowers and butterflies, respectively. While Fish'n'Steps provides both positive and negative reinforcement, Ubifit provides positive reinforcement only.

In summary, self-monitoring technology designers should understand the advantages and disadvantages of various feedback machanism and carefully choose the device and form of the feedback in order to draw most effective results of self-monitoring.

3.3. Background on Sleep

In this section, I provide background on the nature of healthy sleep behaviors, including sleep hygiene, types of sleep, and different sleep disorders. This section is intended to help readers understand important aspects of sleep. I also include other sleep-related technology that I have not covered in *Section 3.2*.

3.3.1. Significance of Sleep

Getting the appropriate amount of sleep has been correlated with numerous health benefits, including reduced fatigue and stress. Likewise, regularly getting less than 6 to 7 or more than 9 hours of sleep per night correlates with an increase in a number of diseases, including diabetes

(Gottlieb et al., 2005) and heart disease (Ayas et al., 2003). In addition to one's physical health, there are other important consequences of poor sleep habits. When people do not get enough sleep, their alertness is greatly reduced (Bonnet & Arand, 1995), and they often put themselves at a greater risk of a car accident, with estimates of as many as 36% of all fatal car accidents resulting from driver drowsiness (Leger, 1994). Finally, a poor night's sleep can affect memory (Maquet, 2001; Wagner et al., 2004), and cognitive functioning (Faubel et al., 2009), which can result in poor work performance. However, chronic sleep deprivation is common in the developed world, with 28–29% of all young adults reporting only 6.5 hours of sleep each night (Bonnet & Arand, 1995). Sleep disorders are often undiagnosed, and many people are unaware of how their activities or environments affect sleep.

3.3.2. Background on Sleep

Sleep doctors and the sleep research literature make consistent recommendations to maximize one's chance of getting a good night's sleep (Stepanski & Wyatt, 2003). Called ***sleep hygiene***, these recommendations help people prepare their sleep environment and engage in behaviors that prepare them physically and mentally for sleep. According to the sleep literature (Stepanski & Wyatt, 2003) and recommendations from the sleep center involved in our research, basic sleep hygiene recommendations consist of the following:

- Sleep only as much as you need to feel refreshed during the following day
- Keep a consistent wake time and amount each day, seven days per week
- Do not eat or exercise within three hours of bedtime
- Make sure your bedroom is comfortable, free of light and noise, and is at a comfortable, cooler temperature
- Eat regular meals and do not go to bed hungry
- Avoid excessive consumption of liquids and alcohol in the evening
- Limit the consumption of all caffeinated products
- Avoid smoking during the night when you have trouble sleeping or quit smoking entirely

- Do not attempt to sleep while stressing about problems
- Position the clock so that you cannot see it
- Do not use your bed for anything other than sleep or sexual activity (e.g., do not read or watch television, etc. in bed)
- If you do not fall asleep within 30 minutes of going to bed, get out of bed and engage in a quiet activity (e.g., reading, watching television)
- Avoid napping during the day

These recommendations are seemingly straightforward, but can be difficult to follow. For example, people might engage in an activity for entertainment (e.g., reading or watching TV in bed), to be social (e.g., drinking alcohol), or because of other responsibilities (e.g., calming a crying baby or working toward a deadline), even if they know that it is not good for sleep. Even for those who are determined to change their behavior, some recommendations do not provide concrete and actionable information as to exactly what to do (e.g., How much liquid consumption at night is too much? What should you do if you need to get some sleep but are stressed out?). At the same time, sleep requirements are highly ***individualistic***. For example, some people may feel rested after having 5 hours of sleep while others may need 8 hours of sleep to have the same restful feeling. For this reason, sleep hygiene recommendations tend to remain as high-level suggestions. Although sleep hygiene is a good place to start thinking about the design space, sleep experts and HCI researchers can work together to explore ways to customize sleep hygiene recommendations and develop actionable goals.

Sleep consists of varying phases throughout the night, including two broad types—***Rapid Eye Movement*** (REM) sleep (when most people dream) and ***Non Rapid Eye Movement*** (NREM) sleep. The NREM sleep phase is further broken down into N1, N2, and N3; N3 is also referred to as the ***deep sleep*** phase, which is the most important contributor for people waking up feeling refreshed and restored. Bad sleep environments (e.g., an uncomfortable mattress or pillow, noise, light, too high or low temperature level), stress and anxiety, and consuming food and drinks before going to bed can have a negative impact on deep sleep, which may cause sleep deprivation—even for people who seem to be sleeping for an adequate period of time. People need varying amounts of sleep, but most sleep professionals, as well as the National Sleep Foundation (USA), recommend that adults get eight hours of sleep each night to feel rested and have full cognitive functioning during the day, with younger individuals tending to need more[9]. The 2013 Sleep in America Poll[10], however, indicates that on average, adults in the U.S. sleep 6.5 hours a night on weekdays, which can result in sleep deprivation. The problem has only gotten worse. People's reported average sleep hours for both weekdays and weekends have been continuously decreasing since the Sleep in America Poll was first undertaken in 2001. This trend indicates that regularly getting 8 hours of quality sleep has become a luxury for many people these days.

Sleep doctors work with patients to diagnose and treat a number of different sleep disorders. Sleep disorders include not being able to sleep (insomnia), difficulties in breathing during sleep (sleep apnea), sleepwalking or talking (parasomnia), falling asleep uncontrollably (narcolepsy), and circadian rhythm disruptions (delayed and advanced sleep phase syndromes) (Colten & Altevogt, 2006). Several other conditions or disorders—such as depression, obesity, or chronic pain—can negatively impact sleep. Patients suspected of having a sleep disorder spend a night sleeping at a clinical sleep center while doctors and technicians monitor the patients' sleep using night vision cameras and a variety of measurements, including EEGs, EKGs, pulse oximetry, and respiration. This highly studied night is then used in conjunction with manual sleep diaries completed by patients in their own homes to diagnose a variety of sleep disorders and recommend treatment.

3.3.3. Sleep Technology in HCI Research

Sleep is a relatively new topic in HCI research. In 2011, my colleagues and I conducted an in-depth formative study to identify the design space of sleep technologies (Choe et al., 2011). We found a broad interest in technologies for sleep, with a majority of people expressing interest in tracking sleep data over time. Simplicity, unobtrusiveness, and privacy were identified as

[9] National Sleep Foundation, http://www.sleepfoundation.org.

[10] 2013 Sleep in America Poll. http://sleepfoundation.org/sites/default/files/RPT495a.pdf.

crucial qualities of sleep technologies. Over the past few years, a growing number of sleep-related research has been published in the HCI and Ubiquitous Computing literature. This research has been primarily concerned with providing sleep hygiene recommendations (Bauer et al., 2012), tracking sleep (Chen et al., 2013; Lawson et al., 2013; Min et al., 2014; Shirazi et al., 2013) and environmental disruptors (Kay et al., 2012), and sharing sleep data with others (Kim et al., 2008; Shirazi et al., 2013). Most of these systems are designed for people without a diagnosed sleep problem; rather the systems strive to increase people's awareness of sleep behaviors. I covered the details of these sleep-monitoring technologies in *Section 3.2*.

Although a number of self-monitoring technologies already exists, much of the existing literature in HCI has not systematically looked at either people's needs in regard to sleep or their current self-monitoring practices. Also, people use these technologies in the bedroom, where they are often sleepy and groggy; yet, few researchers discuss the resulting design implications for such a setting. In addition, how self-reflection is supported by these self-monitoring technologies has not been fully investigated. Finally, we do not yet know what forms of feedback would best serve people's needs because few studies compare the different forms of feedback that deliver the same information.

3.4. Chapter 3 Summary

In this chapter, I distinguished the term self-monitoring technology from personal informatics and Quantified Self. I also described a number of self-monitoring technologies for health including commercial products and HCI research projects. Lastly, I provided background on sleep hygiene, types of sleep, different sleep disorders, and sleep-related technology. Although a number of self-monitoring technologies already exists, I identified gaps in the existing sleep-monitoring research in HCI. This dissertation work endeavors to identify and address these gaps in self-monitoring technology research.

Chapter 4
Understanding Quantified-Selfers' Practices in Collecting and Exploring Personal Data

Chapter 4 forms the first part of the formative studies. In this chapter, I address the first research question, "How do people currently practice self-monitoring?" I present current practices of self-monitoring through the lens of Quantified-Selfers (Q-Selfers). Researchers have studied how people use self-monitoring technologies and discovered a long list of barriers including lack of time and motivation as well as difficulty in data integration and interpretation. Despite the barriers, an increasing number of Q-Selfers diligently track many kinds of data about themselves, and some of them share their best practices and mistakes through Meetup talks, blogging, and conferences. I aim to gain insights from these "extreme users," who have used existing technologies and built their own workarounds to overcome different barriers. I uncover Q-Selfers' motivations to self-monitoring, tools for data collection and exploration, insights they gained, common pitfalls to self-monitoring, and workarounds to overcome the challenges. I discuss how these findings can have broad implications in designing and developing self-monitoring technologies.

4.1. Introduction

Q-Selfers offer us a useful perspective from which to re-examine the current design of self-monitoring technologies and ways to improve them. Q-Selfers encompass a broad spectrum of people ranging from those who use pen and paper to those who build their own tracking applications. Because they can be categorized as a somewhat *extreme user group*, their stories, including the successful ones, might not be generalizable or applicable to the broader population. However, as other researchers pointed out (Troshynski, Lee, & Dourish, 2008), the perspective of those who represent "extremes" gives us distinct insights because they have used existing technologies and spent numerous hours building their own workarounds when faced with problems.

We had many questions about this particular group. What motivates Q-Selfers to keep tracking data, despite numerous barriers? What tools do they use to collect and explore data? What insights do they gain from tracking? What are the outcomes of tracking? What challenges do they face and how do they overcome these? We explored these questions through a qualitative and quantitative analysis of 52 video recordings of QS Meetup talks stored on the QS blog (quantifiedself.com). Each talk illustrates a distinctive self-monitoring approach that could benefit human-computer interaction, health informatics, and information visualization researchers whose work is within the domain of self-monitoring and personal analytics.

In what follows, I begin by explaining the study methods, dataset, and profiles of Q-Selfers. I then detail themes that arise from our qualitative and quantitative analysis as I answer the Three Prime Questions posed to the QS Meetup speakers—(1) what they did, (2) how they did it, and (3) what they learned. While addressing these questions, I highlight several common pitfalls Q-Selfers experience. Lastly, I identify future research efforts that could help make progress toward addressing these issues.

4.2. Methods

At the QS Meetups, people talk about their firsthand experiences with self-monitoring methods and tools using a "Show & Tell" format. Talks on scientific theories, demos of tools and apps, and philosophical speculation are discouraged unless they are grounded in actual attempts at self-monitoring and self-experimentation. The uniqueness of the QS Show & Tell comes from the fact that they follow the specific guideline, provided beforehand, to organize the talk to answer the ***Three Prime Questions***:

- What did you do?
- How did you do it?
- What did you learn?

The consistent structure of the talks makes them a valuable dataset. In answering ***what they did***, speakers talk about initial problems and motivations to do self-monitoring and track items. In answering ***how they did it***, they talk about tools and methods they used and the visualizations they created from personal data. In answering ***what they learned***, they talk about insights gained and the outcomes of their tracking. The talks are usually 5 to10-minute-long followed by a question and answer period. The talks, including the Q&A, are often video-recorded and uploaded to the quantifiedself.com blog by Meetup organizers for sharing.

4.2.1. Dataset

As of April 8, 2013, 205 video posts had been uploaded to the QS blog since 2008. Of those, we analyzed 83 recent video posts, those uploaded since January 2012, to examine the most up-to-date landscape of QS practice (Figure 2).

Not all of the posts fit within our area of interest for this study. For example, despite the three prime questions guideline, some speakers presented a new tool that they developed without describing their actual self-tracking practice. Others presented academic research on other people's data. Some videos failed to capture the speakers' visual aids (i.e., slides) so it was difficult to understand the specific context. Thus, video posts had to meet the following two inclusion criteria in order to be added to our dataset: (1) the speaker should present their own QS practices; and (2) video posts should include personal data visualizations of some kind (e.g., table, graph) created with the speaker's own data. Of the 83 videos we reviewed, 52 videos met the inclusion criteria. The average length of these videos was 15 minutes, 53 seconds (including Q&A). We transcribed this entire corpus of videos to aid with analysis.

The speakers of QS Meetup talks were a self-selected group of people who volunteered to give a talk, and might not accurately represent the whole community of Q-Selfers, not to mention the general public. We suspect that they are more extreme in terms of technical ability and experience with self-monitoring than typical Q-Selfers. We also note that not all the QS Show & Tell videos are recorded and uploaded to the QS blog.

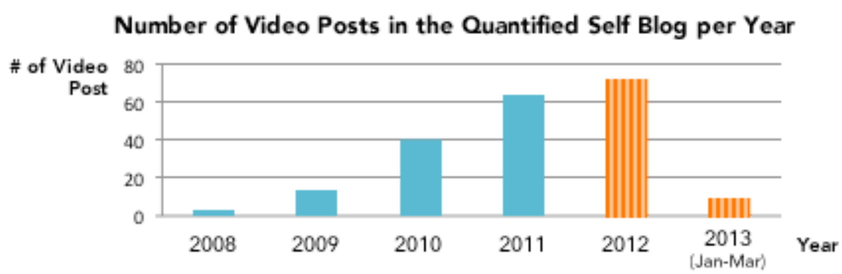


Figure 2. QS Video posts per year. Our dataset is colored in orange with vertical stripes.

4.2.2. Data Analysis Methods

We employed a variety of techniques to analyze our dataset. First, with an aim to understand and characterize Q-Selfers, we created a profile for each speaker by systematically capturing the following information: location, gender, job description, health condition, types of data collected, self-monitoring duration, data collection tool, data exploration tool, type of tool (user-generated vs. commercial), and data sharing aspects. Given that we had to rely on the information speakers disclosed during the talk, some information is missing for some speakers. Second, we conducted an affinity analysis (Beyer & Holtzblatt, 1997) as a group. After several passes, the video transcripts were broken into approximately 400 quotes, each of which contained one main idea. We inductively organized these into categories to identify key themes. We drew several bottom-up themes regarding people's motivations for self-monitoring, self-monitoring methods and tools, insights gained, outcomes of tracking, and common pitfalls of self-monitoring. Lastly, we captured 188 screenshots (and the slides when available) that included personal data visualizations. We analyzed the screenshots by categorizing the visualization type.

4.3. Results

In this section, I provide the profiles of Q-Selfers based on qualitative and quantitative coding analysis. I then report on findings categorized by the Three Prime Questions.

4.3.1. Profiles of the Quantified-Selfers

Location. Seventeen (33%) video posts were recorded from San Francisco / Mountain View / Silicon Valley area. This location is where the QS movement first started and still remains very active. Nine (17%) video posts were from Seattle followed by seven (13%) videos from London and New York respectively. Other locations included Toronto (6%), Pittsburgh (4%), Singapore (4%), Washington DC (4%), Boston (2%), San Diego (2%), and Portland (2%).

Gender. Forty-one (79%) speakers were male, while only 11 (21%) were female. Pew Research reported that in the general population, men and women are equally likely to report tracking their weight, diet, or exercise routine (Fox & Duggan, 2013).

Health Condition. Eighteen (35%) reported having some health conditions, such as sleep disorders, diabetes, panic attacks, cancer, obesity, or allergies. For them, their health conditions highly influenced what they tracked because they wanted to maintain a certain condition, find triggers, identify a medication's effect, or achieve some health goal.

Job Description. Twenty-one (40%) speakers were working at a startup. Eighteen (37%) speakers described themselves as a software engineer or programmer. Seven (13%) were working in data analytics and four (8%) were electrical engineers. Other job titles included creative director, psychologist, designer, product manager, graduate student, operations analyst, professor, and professional athlete.

Tracking Duration. The average duration of tracking was 25 months, where the obtained range was 5 days–20 years (*SD* = 44.0 months, *Median* = 8 months).

4.3.2. What Did You Do?

I organized the findings based on the Three Prime Questions. We begin by answering the first question—"what did you do?" I here describe the types of data Q-Selfers tracked and the motivations behind tracking. I also identify common pitfalls regarding data collection phase and Q-Selfers' approaches to alleviate some of the pitfalls.

4.3.2.1. Items Tracked

Activity (40% of Q-Selfers), food (31%), weight (29%), sleep (25%), and mood (13%) were the most popular items Q-Selfers reported tracking. In contrast to our results, bank statements, email history, and credit card bills were the top 3 items people reported in Li et al. (2010). This is possibly because Q-Selfers did not consider readily available data "self-monitoring," or if they did, they did not report on it during the Meetup talks. On average, Q-Selfers tracked 2.92 items (SD = 2.41) where the obtained range was 1–11. In all, they reported 57 unique items. The long-tail shape of Figure 3 indicates that Q-Selfers have diverse interests. Other items they reported tracking included cognitive performance, blood glucose, location, heart rate, symptoms, knowledge, stress, body fat, productivity, snoring, movies, posture, medicine, skin condition, home energy usage, clothes, and public transit usage. Some people tracked multiple items simultaneously with the intention of identifying correlations among the factors, while others tracked one or two items at a time but applied the self-monitoring practice on several topics over time as their interests change.

Figure 3. Number of people tracking a certain item.

4.3.2.2. Motivations to Practice Self-monitoring

We classified Q-Selfers' motivations to track into three main categories: (1) to improve health, (2) to improve other aspects of life, and (3) to find new life experiences. In Table 3, we break down these categories further and include tracking examples for each of the categories. Thirty-five (67%) speakers tracked one or more health-related items with an aim to improve aspects of health. Considering the number of people who had a health condition (35%), improving health was a prevalent motivation regardless of the presence of a health condition. Furthermore, they had very specific health-related goals—such as finding triggers for an allergy, finding out how exercise affects body mass and weight, finding the right drug dosage, or executing a treatment plan for treating panic attacks—rather than merely "to become healthy" or "to change health behaviors." Some people in this group claimed that they "treated" or "cured" a disease through self-monitoring. For those who experienced positive outcomes from self-monitoring, QS was an approach to better life, not just a data collection method.

Another group of Q-Selfers was interested in improving other aspects of life—predominantly work efficiency and cognitive performance. They used self-monitoring to measure their current use of time (with time tracking apps or calendar logging), cognitive performance (by taking an online cognitive test), or time spent on a computer (with productivity tracking software). People in this group—who were either software engineers or students—wanted to find ways to "optimize" their work and life and "maximize" learning.

Table 3. Quantified-Selfers' tracking motivations and examples for each category.

Motivations	Sub-categories	Tracking example
To improve health	To cure or manage a condition	Track blood glucose to hit the target range [P37]
	To achieve a goal	Track weight to get back to the ideal weight of 135 pounds [P39]
	To find triggers	Log triggers that cause atrial fibrillation [P55]
	To answer a specific question	Track niacin intake dosage and sleep to identify how much niacin to take for treating symptoms [P76]
	To identify relationships	Track exercise, weight, muscle mass, and body fat to see the relationships among the factors [P31]
	To execute a treatment plan	Log food, exercise, and panic as a recovery plan for panic attack [P35]
	To make better health decisions	Record ideas of things that thought were healthy and unhealthy to make better decisions [P18]
	To find balance	Log sleep, exercise, and time to get back from erratic lifestyle [P23, P42,

Motivations	Sub-categories	Tracking example
		P54]
To improve other aspects of life	To maximize work performance	Track time to know the current use of time and ways to be more efficient [P43, P63]
	To be mindful	Take a self-portrait shot everyday for 365 days to capture each day's state of mind [P26]
To find new life experiences	To satisfy curiosity and have fun	Log the frequency of "puns" to see how often these puns happened and what triggered them [P12]
	To explore new things	Track every street walked in Manhattan to explore as much of the city as possible [P34]
	To learn something interesting	Track heart rate for as long as possible and see what can be learned from it [P62]

The last category consisted of those who wanted to have new life experiences through self-monitoring. They often had no specific goals in mind when starting to track, but quickly discovered interesting patterns from data that led to data collection becoming habitual. For example, P62, who did not have a heart condition, collected heart rate data for 24 hours a day for over a year. He streamed his heart rate data to various channels online, which were updated every 30 minutes. He learned how his body responds to various routines and stressful events, which in return influenced his decision-making (e.g., avoiding heavy meals because his heart rate would go up 20%). The technical ability of people in this group combined with their creativity allowed them to explore new life experiences through self-monitoring.

4.3.2.3. Common Pitfall 1: Tracking Too Many Things

Q-Selfers described that they were often too ambitious at first and tried to track too many things, as P61 remarked: "*I can honestly say that I've made the classic newbie self-tracking mistake which is that I track everything*." Tracking too many things often led to either stop tracking entirely due to *tracking fatigue* or fail to do data analysis due to too much data in different formats.

Q-Selfers offered some suggestions on how to alleviate tracking fatigue. First, they suggested automating the tracking and data uploading if possible. P39 had been tracking her weight, food, and exercise for 6 years using several different tracking methods, such as manual entry with an Excel spreadsheet, pen and paper, Google docs, and most recently, automatically through a WiFi-scale. Her definition of successful tracking was to capture many data points for a long period of time. The use of a WiFi-scale, which automatically uploads data to a website, allowed her the longest and the most regular data acquisition (compared to other methods). Second, if automating is not possible, Q-Selfers suggested making tracking very simple and easy to do by (1) lowering data granularity (e.g., "*If you can't automate your tracking, make your tracking binary*" [P51]) or (2) making manual capture very easy. P35 built a manual counter app whose main design goal was to reduce user burden in capturing his panic symptoms and triggers: "*…recording is a one-tap process—have a drink at the bar, tap alcohol, go on a run, tap to start, tap to stop, simple*." At the expense of data granularity, they were able to lower the user burden associated with capturing, thereby capturing more data points overall. Lastly, Q-Selfers suggested making tracking a *rewarding experience*. P11 drew an interesting analogy to explain what tracking means to her: "*…when I was pretty young, I was really susceptible to being awarded Gold Stars, it makes me want to do the thing I've been awarded Gold Star for more. So the process of tracking was like awarding myself the Gold Star... So what I learned was track what you want to do more.*" P11 explained that focusing on the positives makes the tracking experience rewarding and less of a burden.

4.3.2.4. Common Pitfall 2: Not Tracking Triggers and Context

People who were new to QS made a common mistake of focusing too much on tracking symptoms or outcome measures but failing to capture the important triggers or context. This failure resulted in not having enough clues on how to improve outcome measures. P9 described, "*…I've been trying all this biometric tracking trying to be more consistent in my health than have more healthy habits. But the whole time, not just my health habits, but even my tracking habits were completely reliant on my emotional state. So here I was trying to track all these symptoms, and I was completely ignoring the cause*." After a few months of trial and error, P9 modified her tracking routine from capturing biomedical data to capturing negative emotion and the biometric data surrounding it. Likewise, P3—a student who diligently tracked every activity—was able to alter his initial question after a few months of tracking. He initially wanted to know *where his time was going* as specifically as possible. He kept track of time and planned everything ahead using a calendar. However, after tracking for four months, he realized he needed to step back from the day-to-day events and ask a different question: "*how to balance my life?*" Novice Q-Selfers found it difficult to know exactly what to track or what questions to ask in the beginning. In fact, the initial tracking phase helped Q-Selfers redefine what to track or what questions to ask. Q-Selfers thus endeavored to step back from time-to-time and reflect on whether they are tracking the right thing for the right reason. This finding is in line with Li et al.'s report on the transitions between the Maintenance phase and the Discovery phase during self-reflection (Li et al., 2011).

4.3.3. How Did You Do It?

In this section, we address the second prime question, "*how did you do it?*" We examine tools for data collection and exploration, reasons for building custom tools, and visualizations Q-Selfers created. We also describe the notion of *self-experimentation*, a prevalent practice among Q-Selfers to get concrete answers to their questions.

4.3.3.1. Data Collection and Exploration Tools

Q-Selfers reported using a variety of tools for self-monitoring, which we categorized into Data Collection Tools and Data Exploration Tools (see Table 4). Data exploration tools included data analysis and visualization tools. On average, Q-Selfers used 2.1 data collection tools (SD = 1.08) and 1.4 data exploration tools (SD = 0.63).

Data Collection Tools. Commercial hardware, such as a health monitoring device (e.g., Fitbit, ZEO, WIFI-scale, heart rate monitor), was the most popular tool (56%) followed by spreadsheets, such as Excel or Google Docs (40%). Eleven (21%) built custom software such as a snoring app, mood/stress tracking app, activity/location tracking app, or productivity tracking software. Ten (19%) used commercial software, such as standalone mobile apps for tracking sleep, productivity, or food. Two speakers reported building custom hardware, such as wearable sensors for tracking posture and smiles.

Data Exploration Tools. The most popular data exploration tool was a spreadsheet (e.g., Excel, Google spreadsheet) for running simple statistics and creating graphs (44%). Eighteen (35%) built custom software that required some programming such as using open-source JavaScript libraries to create a website or mobile apps with data visualization features. Fourteen (27%) relied on a commercial website (e.g., Fitbit, ZEO, Quantified Mind) where visualizations are auto-generated once data is manually entered or uploaded from commercial hardware. Six (12%) used commercial software that often interconnected with commercial hardware to aid with analytics and data storage (e.g., software that comes with blood glucose monitor). Only two speakers used statistical software, such as R. None mentioned using commercial data exploration software (e.g., Tableau).

In all, thirty (58%) used only commercial tools, twelve (23%) used only user-generated (custom) tools, and ten (19%) used a mix of commercial and user-generated tools for collecting and exploring data. Our analysis shows that not many tools support the whole spectrum of QS from data collection to data exploration. We also found that many Q-Selfers built custom tools, especially for data exploration purposes. Both findings indicate issues with data portability, regarding people having to export and import data from a tracking tool to an exploration tool. To further exacerbate the situation, some companies (e.g., Fitbit) charge fees for people to export their data, which makes it hard to combine data from different sources.

Table 4. Types of data collection and data exploration tools and usage frequency.

Data Collection Tool	% (#)
commercial hardware	56% (29)
spreadsheet	40% (21)
custom software	21% (11)
pen and paper	21% (11)
commercial software	19% (10)
commercial website	10% (5)
camera	6% (3)
open-source platform	6% (3)
custom hardware	4% (2)
other	10% (5)

Data Exploration Tool	% (#)
spreadsheet	44% (23)
custom software	35% (18)
commercial website	27% (14)
commercial software	12% (6)
open-source platform	8% (4)
statistical software	4% (2)
pen and paper	2% (1)

4.3.3.2. Visualization Types

Visualizations were the key means to gain insights from data. To analyze how Q-Selfers explored data, we captured 188 screenshots composed of 243 charts and types of graphical feedback. From these, we analyzed visualization types and frequency of usage. We identified 21 unique visualization types, which are shown in Figure 4 (word size reflects the frequency of usage).

Figure 4. Tag cloud showing the usage frequency of visualization types. Line chart, bar chart, and custom visualizations were the top 3 most commonly used ones.

Line charts were by far the most frequently used, followed by bar charts and custom visualizations such as an infographic-style website, calendar (Figure 5-a), physical light (Figure 5-b), map and photo grid on timeline (Figure 5-c), and a combined visualization composed of line charts, stacked bar charts, and tables (Figure 5-d). Figure 5-a is an example of "*appropriation*" where iCal, a personal calendar application, was used for tracking and visualization purposes. Figure 5-b is a rare example of real-time feedback—the blue lights blink whenever wearable EEG sensors detect smiling. Figure 5-c and Figure 5-d are highly customized and complex timeline visualizations, which helped creators understand how they spent time online and offline (Figure 5-c) and how to optimize performance (Figure 5-d).

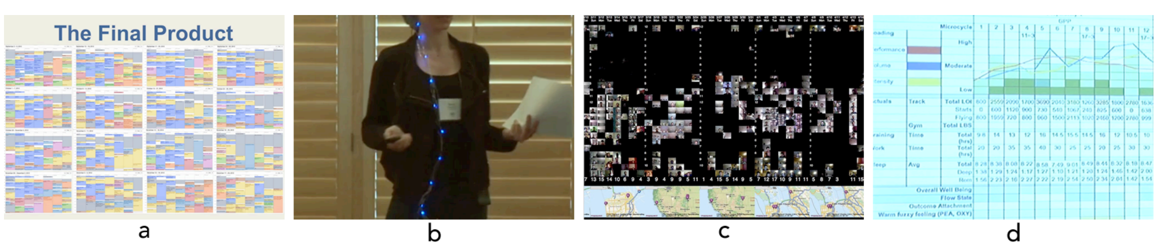


Figure 5. Examples of custom visualizations.

4.3.3.3. Reasons for Building Custom Tools

Although numerous commercial self-monitoring tools are available, many Q-Selfers built their own tools. We identified common reasons for building custom tools. First, few commercial tools supported the two key features that Q-Selfers preferred—(1) being able to track and explore data using a *single tool*, and (2) being able to perform *self-experimentation*. When Q-Selfers had the technical ability, they built a custom tool to meet their needs. For example, P15 had a snoring problem. He first looked for existing snoring apps, but they did not have the features he wanted. He envisioned an app that could do both snore tracking and analysis. He also wanted to test if certain things (e.g., snoring remedies, drug, alcohol) affected his snoring. Not finding what he wanted, he built an app called SnoreLab, which he released to a commercial app store. Second, Q-Selfers built a new tool when they wanted to do centralized tracking as P23 remarked, *"I found myself using Nike Plus for my exercise, my Foursquare for social check-ins and several different apps just for tracking my time. When I finally said, 'You know what? I'm fed up with this. I want to make my own tool that allows me to do this in a more cohesive manner.'"* Third, many Q-Selfers built custom websites for data presentation, which was typically done with publicly available visualization APIs such as d3 (*Data-Driven Documents*) and the Google Charts API (*Google Charts API*). Lastly, some Q-Selfers built a custom tool simply because no existing tool supported their needs. One example is software developed by P70 who tracked her inventory of clothes that would help her coordinating clothes and simplify her wardrobe.

4.3.3.4. Self-Experimentation

Q-Selfers wanted to draw definitive conclusions from their QS practice—such as identifying correlation (e.g., sleep and cognitive performance are not correlated) or even causation (e.g., weight tracking causes weight loss). To accomplish this goal, they needed to first generate hypotheses to test. Testing ideas came from careful observations of previous behavioral patterns (e.g., P39 strongly suspected a beer allergy), other speakers' Show & Tell talks (e.g., P33 was inspired by another Q-Selfer who tested whether eating butter increases cognitive performance), or individual needs (e.g., P76 wanted to find the right medication dosage).

Q-Selfers often described the process of seeking answers as *self-experimentation*. When used in an academic context, self-experimentation means participating in one's own experiments when recruiting other participants is not feasible. However, in QS, the goal of self-experimentation is not to find generalizable knowledge, but to find meaningful self-knowledge that matters to individuals. P33 emphasized, "*I discovered the importance of testing for myself because what works for some people does not work for me*."

4.3.3.5. Common Pitfall 3: Lack of Scientific Rigor

Q-Selfers conducted a wide variety of self-experimentations without having a control condition. A typical personal experiment resembles the experiment that P69 conducted—he observed patterns between his allergic reactions and spending on beer, so he suspected that beer might have caused his allergy: "*So ignoring the doctor, who said I was fine, I decided to do one final experiment. We had colleagues going away, and it was a very happy day, I drank lots of beers. By the following Sunday and the following Monday I was in the worst form that I had ever been, and I decided that was enough for me. This must be the trigger*." P69 did not control for other confounding factors, which threatens the internal validity of his finding. However, as long as Q-Selfers were happy with the outcomes of tracking, most of them did not seem to care about the lack of scientific rigor, as it is not the main goal of QS.

Although a minority, some Q-Selfers attempted to design more rigorous personal experiments. We identified three approaches that could possibly increase internal validity—(1) having a control condition, (2) triangulating with other methods, and (3) using the experience sampling method. We note that the terms we used in this paper such as "internal validity" and "triangulation" are our interpretation of the behavior, not the language used by the Q-Selfers.

Some Q-Selfers conducted self-experimentation with a control condition to reduce biases. As a doctor and researcher, P55 was well aware of skepticism his colleagues had about QS. He said, "*When I talk about these things, I feel like I'm talking by myself because most of all in medicine, and in science, they're going to roll their eyes. It's like science is being done in your garage, but I really think that there's some real potential*." Then he explained how he conducted a within-subjects design to identify what triggers his atrial fibrillation by comparing what he did right before the onset of the disease (hazard period) to the usual routine (control period). Then he calculated an Odds Ratio—a measure of association between an exposure and an outcome—and identified risk factors for his atrial fibrillation such as caffeine, air flight stress, more than 1 glass of wine, and public speaking in the previous 2 hours of the onset of the disease.

Triangulation—using two or more different methods to measure the same phenomenon—was commonly used to facilitate data validation through cross verification from multiple sources. P28 said, *"... if I compare my Zeo data to my Fitbit data I really only wake up when I flip over in bed, so it's actually very accurate for me."* However, P7 came up with a disturbing finding: *"I discovered that my glucose meters aren't that good. So, comparing the measurements from these two different meters, I only came out with R-squared of 0.46, which I would have hoped for a lot better agreement between the two meters."*

Some Q-Selfers employed the *experience sampling method (ESM)*. P59 used an app called 'The Mappiness' to track his stress level. He configured the app to prompt him at random times during the day. The number of prompts was also configurable. He acknowledged that ESM produces the gold standard of experience measurement.

Nevertheless, critics abound. Although some biases might be reduced from some of these attempts, critics claim that experimenters might be biased to produce the result they expect to see (Rosenthal, 1966). By definition, QS is designed and conducted by the experimenter, and thus, the issue regarding experimenter's bias remains open.

4.3.4. What Did You Learn?

The way Q-Selfers reported their learning was twofold. First, they reported insights gained from their data exploration. Second, they reported desirable and undesirable outcomes in a broader tracking context. After we report on these two types of learning, we discuss the key hurdle in gaining insights, which is data interpretation.

4.3.4.1. Gained Insights

Q-Selfers often summarized their findings by reporting *descriptive statistics*. For example, after tracking the usage of tabs in a web browser, P4 learned that in a two-month period he opened or closed 32,000 tabs, which averages to about 500 a day. This finding led him to think about his next project: extending his code to track every time he switches tabs as a proxy for how much attention he is giving to something. *Comparison measures* were common descriptive statistical methods that helped Q-Selfers quickly gain insights. When both control and intervention conditions were in place, they typically reported differences in means or Odds ratio. However, they could still compare within themselves without the explicit control condition by categorizing data points in several bins after data collection was complete and then comparing differences across the bins. One example is how P78 reported his sleep data—he compared sleep efficiency and time to fall asleep across restless nights and restful nights. Q-Selfers also compared themselves against the general population with similar demographics if they had access to the population data. P70, a Canadian engineer who tracked work time, learned that she worked more hours than average for Canadian workers.

Q-Selfers with a statistical background reported statistical test results. *Correlation* was the most commonly reported statistical test. For example, P18 learned that "*high idea days were correlated with conferences, sedentary events, Internet usages, high calories, and very little working out*," which he found problematic. People were surprised when they found low correlation between things that they thought were highly associated: "*So far, I had pretty much no correlation, so it's really interesting to think about how I have found no correlations between even the most meaningful things in my life and how I rate the day*" [P17]. Q-Selfers with no statistical background still reported perceived correlation in layperson's language: "*When my symptoms were good or when my body felt good, I happened to be in a good mood. When my symptoms were bad or my body felt bad, the mood or my mental state was bad*" [P61].

Stepping back from the data, Q-Selfers articulated high-level, qualitative take-away points. For example, P11 learned that "*the biggest contributors to my daily happiness are the small things*," and

P26 learned "*how tragic it is that we all age*" after taking self-portrait shots every day for a year. Some of them took an interpretive approach and declared that, "*it's not all about the numbers*" [P23]. After 6 months of tracking GPS data, P23 emphasized, "*Numbers are very important, but I think we can aspire for something higher, and I think it also is about the perspective that it allows you to gain*." The general agreement was that important things are found from long-term tracking although people are easily influenced by a day-to-day activity or single data point. The challenge, however, was to find meaningful measures that reflect long-term trends and to keep preserving the initial motivation to tracking even when the latest data point conveys discouraging information.

4.3.4.2. Tracking Outcomes

Q-Selfers discussed various outcomes of self-monitoring, most of which were desirable outcomes such that tracking helped them achieve their initial goals. Many people improved their health and created healthy habits, such as eating healthy, losing weight, and being physically active. Others identified triggers of symptoms and managed to avoid them. P35 realized that driving and drinking coffee were triggers for his panic attacks, and eliminated coffee altogether from his diet, which resulted in a decrease in frequency and severity of the attacks. On the contrary, P8 found a disease that he had not known before, which was ironically a positive outcome for him. P8 initially got into QS to improve his body and get back into shape, but he later discovered that he had Crohn's disease by noticing anomalies from the stool tests he ordered online and his genetic test data. Another positive outcome was the increased awareness of oneself and of the surrounding environment. Being mindful of these things helped people see themselves in a new way such that they were able to understand where ideas came from and how the ideas evolved [P1] or how to create a suitable learning environment for maximizing performance [P41]. Lastly, Q-Selfers reported a positive *reactivity effect* (i.e., change in frequency of the behavior often occurred in the desirable direction) in tracking emotion [P9] and posture [P53]. "*I realized that just by tracking my emotions, I was completely changing them*," said P9, who was able to stabilize her emotional state and prevent herself from experiencing negative emotions.

A few Q-Selfers experienced undesirable outcomes, such as frustration, tracking fatigue, or relapse. Being aware of and confronting negative emotions through tracking caused frustration. "*Because I'm aware of it, it makes it even worse because now I can tell that I'm more anxious than I should be. Before, I was oblivious of being anxious*," said P54, who tracked anxiety and stress. Tracking fatigue was another common outcome of tracking, especially for intensive trackers. After one month of intense tracking on public transit usage, P77 learned many surprising and unexpected findings—such as average commuting time, total cost of using the bus, cost per hour of travel time, and cost per mile. He would not have learned this had he not been tracking, but he said, "*By the end of it, I was really sick of doing it. I just got really fatigued*." However, as the result of a month-long tracking, P77 decided to buy a bike instead of taking the bus because taking the bus was more costly and time-consuming than he had expected. Stopping tracking was not harmful in this case. In fact, P77 used his findings to make a good decision. However, P35, who thought that his panic disorder was under control, stopped tracking and consequently started having panic attacks again. To deal with the relapse and to sustain his commitment to tracking, he built a custom tool to lower the user burden of capturing and looked for a recovery partner for accountability.

4.3.4.3. Open Challenge: Difficulty in Data Interpretation

Data interpretation was a key hurdle for many Q-Selfers. "*It's not that we lack the information, we're virtually drowning in it. The obstacle is that we don't have the proper tools to interpret the significance of our data*," said P61, a personal trainer who used to track 11 different things and cross-referenced them with sleep, mood, energy level and acuity. However, after he could not figure out how to extract meaningful information from the 2 years of data, he simplified his tracking strategy to track only two variables. We observed many people who simplified their tracking strategy after their first failed attempt because there was no easy way to analyze and interpret data. Visualizations were helpful in gaining insights, but again, the learning curve for data manipulation (i.e., data cleanup and formatting) and identifying and creating the most appropriate visualization for a given data type was very steep. Helping the general public effectively explore and easily understand their data using visualizations is an active research area for the Information Visualization research community.

4.4. Implications

We have identified Q-Selfers' common pitfalls and workarounds when they practice self-monitoring. Better self-monitoring tool designs could help any potential tracker avoid some of the pitfalls. Here we identify future research efforts that could help address these problems. We discuss how our findings can have broad implications in designing and developing self-monitoring technologies.

4.4.1. Provide Early Feedback to Help Identify What to Track

In deciding what to track, Q-Selfers encountered two common pitfalls—tracking too many things, which might cause tracking fatigue, and not tracking triggers and context, which might undermine attempts to gain insights later. When the tracker's motivation is high, it is not a problem to track many things, especially at the beginning of the tracking practice. Tracking multiple things could help people decide which items to keep and which items to stop tracking. What is important then is the self-monitoring tool's ability to automate data analysis, provide early feedback on the relationships among different factors, and to suggest eliminating variables that do not seem to correlate with anything. Q-Selfers usually put off data exploration (e.g., running correlations, visualizing data) until later because often, the process involved tedious tasks such as cleaning up data, formatting, and running statistical tests, which could be dramatically reduced by largely automating the process (Kandel et al., 2011). We envision a self-monitoring tool extracting meaningful information, initiating early check-ins, providing real-time visual/textual feedback, and showing comparisons across conditions or correlations among significant factors that can be easily understood.

4.4.2. Support Self-experimentation by Design

Q-Selfers conducted self-experimentation while compromising scientific rigor. Although innate limitations of self-experimentation (e.g., the experimenter's bias) are hard to avoid, we could help people conduct more rigorous self-experimentation by integrating the single-case research design format (Kazdin, 2011) into the self-monitoring technology design. The three requirements for single-case research design include continuous assessment, baseline assessment, and variability in data (Kazdin, 2011). Automated sensing allows easy, unobtrusive, and repeated capturing of a behavior, which could facilitate continuous assessment of the target behavior. Moreover, self-monitoring technology could become a platform where people can systematically configure a varying length of baseline and intervention period, tracking frequency, and independent/dependent variables. Quantified-mind (*Quantified Mind*), a cognitive performance testing website, conveys the similar idea of walking people through setting up self-experimentation. However, cognitive performance is measured by a set of cognitive game scores (hence the fixed dependent variables), which is known to have large practice effects (i.e., repeated testing increases the score).

4.4.3. Maximize the Benefits of Manual Tracking

Sensing and computer automation have many advantages in collecting personal data in terms of reducing mental workload and increasing data accuracy. However, these advantages do not come for free: this automated data collection could reduce awareness and self-reflection resulting from people's engagement with data collection. Q-Selfers expressed that they feel "intimacy with data" when they track data manually. It appears that people make sense of data not only when they explore data but also when they collect data. In addition, some types of data (e.g., subjective sleep quality and pain), by definition, can only be collected via manual tracking. For these reasons, several Q-Selfers built manual tracking tools that drastically lower the user burden, which helped them easily track data and increase awareness. Pushing this idea further, we envision striking a balance between fully automated sensing and manual self-report that can increase awareness, achieve better accuracy, and decrease mental workload.

4.4.4. Promote Self-reflection

Ironically, the name, Quantified Self is misleading in that it makes people think that Q-Selfers' goal is to quantify their behaviors. It is not. Collecting and quantifying data is just one aspect of

QS. The ultimate goal is to reflect upon one's data, extract meaningful insights, and make positive changes, which are the hardest part of QS. HCI research on supporting self-reflection on health monitoring data includes helping people create *unstructured*, *open-ended* diaries and *sharing* them with others in-person (Mamykina et al., 2008). QS Meetup is another example of engaging people—both speakers and the audience—in self-reflection through storytelling. These are good starting points, and we should further examine ways to support self-reflection on personal data with an aim to enhance *positive reactivity effects*.

4.5. Chapter 4 Summary

I analyzed QS Meetup talks and identified that Q-Selfers wanted to improve health, maximize work performance, and find new life experiences through self-monitoring. Although many Q-Selfers had positive outcomes from self-monitoring, some of them had difficulties throughout the process, such as tracking too many things which led to tracking fatigue, not tracking triggers and context which led to not gaining insights, and lacking scientific rigor which led to inconclusive results. My goal was to gain insights from Q-Selfers for designing better self-monitoring tools in general. Experienced trackers' workarounds might help people avoid some of the pitfalls, but better designs could promote broad adoption of self-monitoring technologies by fundamentally boosting their benefits. Specific areas for future research include exploring ways to provide early feedback, to support designing rigorous self-experimentation, to leverage the benefits of—while easing the burden of—manual tracking, and to promote self-reflection.

Chapter 5
Opportunities for Computing Technologies to Support Healthy Sleep Behaviors

In Chapter 4, I described people's tracking practices. Sleep was one of the popular activities that people tracked. In Chapter 5, I specifically focus on characterizing opportunities for tracking sleep to support healthy sleep behaviors. Chapter 5 forms the second part of the formative studies. In this chapter, I address the second research question, "What is the design space for sleep technologies?" To better understand how technology can play a role in improving sleep health, I conducted a triangulated formative study, which includes a literature review on existing sleep-related technologies, contextual interviews with domain experts, a large-scale survey of peoples' attitudes toward sleep-related technologies, and interviews with a subset of the survey respondents. In this chapter, I summarize methods and results, and discuss opportunities for technologies to help people develop and maintain good sleep habits.

5.1. Introduction

Eating a nutritious diet, exercising regularly, and getting adequate sleep are three important activities that people can do to support a healthy lifestyle. The first two have been the focus of many new technologies designed to promote good health. For example, mobile and sensing technologies have been used to encourage healthy eating and exercise habits, track progress over time, and help people set and meet health-related goals. However, while sleep has been the subject of rigorous scientific research in the medical community, the HCI community has placed considerably less attention on ways that technology can support sleep. Similar to how technology has been used to improve other aspects of health, I believe there is an interesting research agenda surrounding the exploration of technologies for promoting healthy sleep habits.

Along with my colleagues, I conducted a literature review and formative study aimed at uncovering opportunities for computing to support the important area of promoting healthy sleep. In what follows, I present methods and results of this formative study. I report on a number of design considerations, challenges, and opportunities for using computing to support healthy sleep behaviors, as well as a design framework for mapping the design space of technologies for sleep.

5.2. Methods

To help uncover opportunities to support sleep, we used a mixed-method approach in our investigation. We began by conducting interviews with experts at the sleep disorders center affiliated with the University of Washington's medical school (N = 4). These interviews helped inform the design of the remainder of our study elements, which included an online survey (N = 230) followed by in-depth, semi-structured interviews with a subset of the survey respondents (N = 16). In this section, I describe the procedure and analysis methods in detail.

5.2.1. Contextual Interviews with Sleep Experts

We interviewed four sleep experts. Our first interview was with the co-director of the sleep disorders center affiliated with our university, who has an M.D. in Neurology and is certified in Neurology and Sleep Medicine. During this interview, we learned about good sleep hygiene, different sleep disorders and their respective treatments, and existing sleep-related technologies. We were also given a brief tour of the sleep disorders center where overnight sleep studies are performed to diagnose different disorders. On a subsequent visit, we conducted a group interview with the co-director and three technicians. During this second visit, we were shown how a sleep study is conducted and reviewed anonymous data collected from patients during a sleep study. We collected artifacts that the center provides on good sleep health, forms used for sleep studies, and blank copies of the sleep diaries that patients complete at home to help doctors with the diagnosis of sleep disorders. The co-director then became a member of the design team and is a co-author on this paper, and thus that perspective is reflected throughout this work. The contextual interviews also informed the design of the online survey as well as our overall understanding of many aspects of sleep.

5.2.2. Online Survey

Based on our literature review and interviews with the sleep domain experts, we developed an online survey to help define requirements for technologies to support sleep. Questions focused on what people perceived to affect their sleep habits, their needs for supporting healthy sleep, and what they believed they would be willing to use in terms of technology. The survey also asked about experiences with sleep disorders and current practices for inducing sleep or waking up. There were a total of 34 questions with a mix of open-ended, Likert, and multiple-choice types.

Table 5. Demographic information for survey respondents.

Total Respondents = 230	
Gender	Female (57.8%), Male (41.7%)
Age Range	21 or under (18.7%), 22-30 (37.0%), 31-40 (21.7%), 41-50 (12.2%), 51-60 (7.4%), 61 or older (2.6%)
Highest Education Level Completed	Some high school (1.7%), High school (4.3%), Some college (20.8%), College degree (22.9%), Some graduate education (16.0%), Graduate or professional degree (31.2%), Training certificate (2.2%)

To recruit survey respondents, we posted paper flyers with the URL of the survey in public locations around the Seattle area, as well as posted a link to the survey on general purpose online discussion forums, websites for people with sleep disorders, and the researchers' social network pages. To encourage participation, each respondent was entered into a drawing to win one of three $50 USD Amazon.com gift cards. In total, 230 respondents completed the survey. Their basic demographics are provided in Table 5. There was a slight bias toward younger, college-educated females. Most respondents were from the United States.

5.2.3. Semi-structured Interviews

In the online survey, respondents were asked to provide their email address if they were interested in participating in a follow up interview. Among those who volunteered, we selected 16 interviewees (10 females, 6 males) based on diversity of demographics, interest in sleep technology, and experience with different sleep disorders. According to self-reported survey data, four participants had no sleep disorders, while 12 had experienced aspects of sleep disorders such as sleep apnea, insomnia, parasomnia, narcolepsy, and severe snoring. We conducted the semi-structured interviews in person or over the phone; interviews were audio recorded and transcribed. Each participant received a $15 USD Amazon.com gift card. The interview focused on the participants' sleep routines, factors affecting sleep, attitudes toward sleep technologies, and any sleep-related issues that may be addressed by technology.

5.2.4. Data Analysis Methods

To analyze qualitative data from the interview transcripts and open-ended survey responses, we segmented direct quotes that represented a single idea. Four researchers used affinity diagramming to categorize the segmented quotes into similar themes. Some themes were structured around the questions (i.e., sleep aids and waking methods) while other themes emerged from the data (i.e., sleep related health goals, factors affecting sleep). From the initial grouping, two researchers independently re-sorted the ideas, and then compared their results. They discussed ideas upon which they did not initially agree (approximately 20% of the ideas) and came to a consensus after a series of refinements. For the quantitative data collected from the survey, we present descriptive statistics.

5.3. Results

In this section, I present findings from the study with a focus on current sleep practices, factors affecting sleep, sleep-related health goals, and attitudes toward sleep-related technologies. Opinions from sleep experts are interwoven into each section to provide additional information in context.

5.3.1. Current Practices

Many survey respondents reported having inconsistent sleep schedules and sleeping for longer on the weekends (*M* = 8.97 hrs, *SD* = 3.58 hrs) than weekdays (*M* = 7.40 hrs, *SD* = 2.46 hrs). Moreover, 27% (*n* = 62) of the respondents reported that they did not have regular sleep habits where they slept for a continuous stretch; they instead slept in intervals using frequent naps. Nonetheless, many survey respondents were aware of the importance and benefits of establishing a healthy sleep routine and reported strategies that they had attempted to maintain a routine that they believed would make them feel less tired during the day.

5.3.2. Factors Affecting Sleep

Responses from the surveys and interviews suggested that numerous factors might have impacted an individual's sleep. Some factors were beyond their control whereas others were something that could have been corrected if they had adequate knowledge about sleep hygiene.

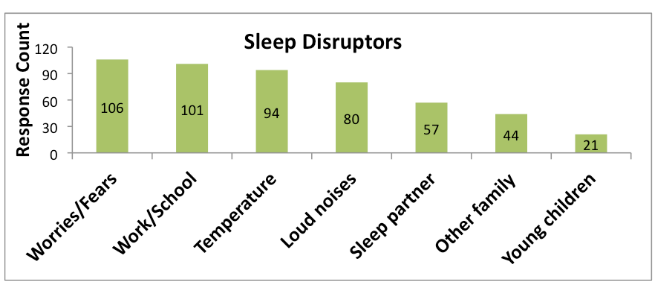


Figure 6. Frequent sleep disruptors from the survey. Participants could select more than one response.

5.3.2.1. Commitments and Stressors

The survey data showed that sleep and wake times were often determined by external factors, especially job and school, which often competed with healthy sleep habits. Work- and school-related commitments were by far the highest number of all the commitments that may have impacted people's sleep patterns. Full time workers' sleep schedules tended to be more consistent compared to those of part time employees and students. People who reported having erratic sleep schedules included caregivers for babies, new mothers, graduate students, professors, medical students, and the unemployed. As shown in Figure 6, stress, fears, and worries that stem from work, school, or their personal life were common sleep disruptors as they caused racing thoughts and thus made it difficult to fall asleep. Other sleep disruptors that were beyond the respondents' control included side effects from other medical conditions, pain, needing to use the bathroom, pets keeping them up or preventing restful sleep, hectic lifestyles, nightmares, and family situations such as:

Survey Respondent (F, 31-40): "*Right now my life is chaotic! I use* [sic] *to have very regular patterns of sleep & wake but due to personal issues (divorce, relocation) I do not have a regular routine any more and typically use Tylenol PM to get to sleep.*"

5.3.2.2. Environmental Factors

A number of respondents and participants noted that environmental conditions often made restful sleep difficult. This included too much light, street noise, loud neighbors, or the temperature of the room being too hot or too cold (Figure 6). According to the National Sleep Foundation, noise levels as low as 40 decibels or as high as 70 decibels can disrupt sleep, as well as the absence of a familiar noise, such as a fan or wind chime. Ideal temperature level is more individualistic although in general, temperatures above 75 degrees Fahrenheit or below 54 degrees Fahrenheit disrupt sleep. Other disruptors that are not listed in Figure 6 included guilty pleasures such as late night caffeine or alcohol, playing video games, or surfing the web.

5.3.2.3. Strategies for Aiding Sleep

Figure 7 shows survey responses on sleep aids. Interestingly, *temperature* was reported as one of the most popular sleep aids in addition to being a sleep disruptor. Uncomfortable sleep environments led people to devise strategies to control their environment. This included installing heavy curtains, using earplugs or blindfolds, covering up with a blanket, and using microwavable heat packs. Others attempted to relax or drown out unwanted noises with music, TV, fans, or white noise machines that play nature sounds. An interview participant explained:

Interview Participant (F, 31-40): "*I put the TV on the cartoon channel, and I have the volume on really low. It helps by being like white noise for me. I live on a fairly busy street and it helps to drown out the traffic noise from outside.*"

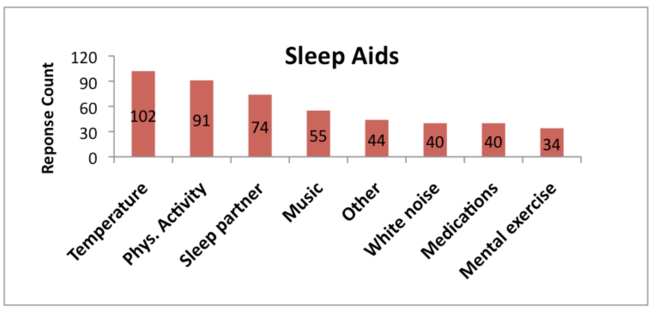


Figure 7. Strategies used by survey respondents to go to sleep. Respondents could select more than one response.

Physical activity was another popular sleep aid, which participants thought to be helpful for their sleep. Sleep doctors, however, suggest avoiding intense exercise within three hours before sleep because it stimulates the heart, brain, and muscles, which makes it harder to go to sleep. This implies that the type of exercise people do and the time of day should be carefully chosen so that it does not impede having a good night's sleep. Other strategies people often used to help them go to sleep included: music; meditations; mental exercises; drinking warm milk, tea, or alcohol; taking a warm bath; listening to rain falling on the roof; reading; watching TV; surfing the web; and sex.

While many people mentioned lighthearted activities and mental relaxation methods to prepare for sleep, there were other extreme lines of thought such as staying up until feeling tired or exhausted and working hard until right before going to bed. Several interview participants and survey respondents used medication or diet to help them fall asleep, although most admitted that this was not a sustainable solution and had varying success. The sleep experts also stressed that sleep medications can become addictive; they typically do not recommend the use of medication except in extreme circumstances. The survey and interview results suggest that people often had misconceptions about what could help them to sleep better, and that there is room for technology to help bridge the gap by providing accurate and customized sleep hygiene information.

5.3.2.4. Technology to Help Waking

Survey respondents were asked to choose one or more strategies that help them wake up. Most respondents used some sort of alarm clock, the majority of whom used the alarm clock on their cell phone (60.4%) or a standalone alarm clock (50.9%). They had strong opinions about the use of alarm clocks, which ranged from a need to have soothing music or nature sounds to which to wake up, to as loud of a sound as possible, to the elimination of the alarm clock entirely. Some felt their alarm clocks gave off too much light or were overly complicated to use while groggy. Most felt their alarm clock needed to be as simple to use as possible. Many also made extensive use of the "snooze" feature, which delays waking by shutting off the alarm and ringing it again a set number of minutes later. Others used multiple alarms to ensure that they would get out of bed. However, the noisy alarm was more or less seen as a necessary evil.

Interview Participant (F, 51-60): "*I try and turn off the alarm before it goes off. That noise will just ruin my whole day*."

Some participants had success with gentler methods of waking, such as the use of a "dawn simulator" that gradually brightened the room, for example,

Interview Participant (F, 22-30): [on success of dawn simulator] "*I think it was just something about waking up more gradually just felt more natural, like I've been rested as opposed to just like being jarred awake by a loud noise*."

As such, participants seemed to have an inner tension between wanting to wake up gently and yet a need to use one or more clamorous alarm clocks to awake on time. Interestingly, sleep experts typically do not recommend using an alarm clock because in an ideal world, people would not need them, and rather they would just wake up when they were rested. However, in some patients with idiopathic hypersomnia where rising in the morning can be particularly problematic, sleep experts recommend multiple alarm clocks, some placed away from the bed, to facilitate waking. The experts also suggested turning the face of the clock away from the bed once the alarm is set so people do not see the time while in bed, as it can increase anxiety.

5.3.3. Sleep-related Health Goals

Many survey respondents and interview participants wanted to become better educated on good sleep habits. This included improving the consistency of their sleep by going to bed and waking up at the same times everyday or becoming more of a "morning person" to synchronize with others around them.

Interview Participant (F, 31-40): "*I'd definitely like to go to sleep earlier. (…) My sleeping schedule is horrible and I need to fix it….I might go to bed at 2 in the morning but I usually don't fall asleep until maybe 3 or 4… it's not that I'm lazy, it's just how my sleep cycle is. I'd like to be able to change that so it's a little more in sync with most of the people around me*."

Others were interested in breaking bad habits, such as a dependence on sleep medications and letting work, homework, or other distractions (e.g., books, television, the Internet) interfere with their sleep routines.

Survey Respondent (F, 21 or under): "*One of my goals is to be able to sleep through the night without the use of medication, but that has never become a reality.*"

In some cases, participants who were already aware of their bad sleep habits suffered from a lack of motivation to do something about it. For example,

Interview Participant (F, 51-60): "*I have a tendency to read in bed, and if I'm in the middle of a good book, I'll read all night. I can't put it down... Or like last night there was a really great show about kamikazes. I was going to go to bed, but I started watching it… I was going to go to bed around 12, but I stayed up.*"

Establishing consistent and healthy sleep patterns was another important sleep-related goal, but this was especially challenging for people with insomnia. When insomnia patients see a sleep doctor, the doctor first monitors patients' sleep patterns by keeping track of their sleep data for about two weeks using either a sleep diary or an actigraphy device. From this data, doctors calculate the patients' sleep ability—how many hours someone can actually sleep—and prescribe times they should go to bed. This is part of *Cognitive Behavior Therapy* (CBT) that has a goal of creating a consistent sleep cycle for insomnia patients. Doctors recommend creating a nighttime routine because it contributes to having a consistent sleep cycle—consistent to-bed and wake up times. The routine would include creating a suitable sleep environment (e.g., dimming the lights), and engaging in an activity that helps people go to sleep. One interviewee illustrated the success of her pre-sleep rituals as a way to get prepared for bed and stated that if she did not do these rituals, she often had trouble sleeping.

Interview Participant (F, 51-60): "*I make my lunch, I pick out my uniform for the next day, I get everything ready by the door, I do my toiletries, I get my bed ready, it's quite a ritual, then I watch TV for awhile and do Sudoku.*"

5.3.4. Attitudes toward Technologies for Sleep

The survey respondents were asked about their attitudes toward using technologies for sleep. Overall, 62.6% (n = 144) of survey respondents answered "Yes" or "Maybe" to the question on whether they would be interested in using a technology to help them sleep. We looked further into the characteristics of these people who were open to the idea of using technologies for better sleep. A chi-square test of independence was performed to examine the relation between attitudes toward using technologies and experience with sleep disorders. The relation between these variables was significant, χ^2 (1, N = 230) = 9.80, p = .002. This suggests that respondents who are interested in using technologies for sleep are more likely to have experienced sleep-related problems than those who are not interested. Our data shows that those who were interested in sleep technologies had experienced sleep-related problems such as having trouble waking up (54.2%) or falling asleep (40.3%), insomnia (51.4%), and mood disorders such as depression or anxiety (50%).

Of those who responded "Yes" or "Maybe," we asked them to report on their level of interest of different features for sleep technologies. Some popular features were recording sleep data automatically, assessing the quality of sleep, recommending optimal sleep conditions, and tracking and reviewing sleep data over time. Other features respondents would like to see included: screening for the presence of sleep partner's snoring; daylight simulation; recording dreams; and helping them set and maintain regular routines. Many were particularly enthusiastic about ideas that would help them fall asleep faster, help them have quality sleep throughout the night, and help them feel rested when they wake up. Some were interested in determining their optimal sleep cycle:

Survey Respondent (F, 31-40): "*It would be nice to know how much sleep I SHOULD be getting and also why I don't crash the day after no sleep... knowing how long I have before that sort of thing catches up to me would be nice*."

Though not surprising, some of the main requirements included the ability to customize the alarm sound and bedroom light levels to cater to different preferences, augment existing technologies when possible (e.g., cell phone) rather than introducing new ones, and be cost effective. Several respondents recounted discarding poorly designed alarm clocks out of frustration.

Technology features that were the least popular included those that required manual input of data and those that shared data within a social network. When participants speculated about their attitudes toward wearable sensing technologies (e.g., *SleepPhase*[3]), they liked the idea of tracking sleep data, but were resistant to the thought of having to wear a special device every night.

Interview Participant (M, 22-30): "*I wouldn't like attachments to my body while I sleep. That would be uncomfortable. It would have to be something like you put it on your nightstand and records all that stuff. If you gotta attach something to your head or to your hands or something, I wouldn't want that*."

Many also emphasized that simplicity and unobtrusiveness were crucial. Although these are important qualities for many technologies, they are especially important for something used every day and at a time when people are typically sleepy or groggy, making them even less willing to deal with complex user interfaces than typical users.

Survey Respondent (F, 31-40): "*Minimal effort on my part. I wouldn't do something that takes a significant amount of time or thought, especially in the morning (I am not a morning person)*."

It is worthwhile to mention that 38.5% (n = 89) of survey respondents were not interested in using any technology for sleep. When we asked why, many remarked that they did not have anything they wanted to change about their current habits, did not see how technology could help them, or did not want "yet another technological intrusion" in their lives. Many people in this group reported that they were in control of their sleep patterns, and were able to fall asleep and wake up when they want to. Thus, if they were up late at night, it was usually by choice.

5.4. Discussion

Our results uncovered a number of design opportunities and insights for technologies to encourage and support healthy sleep behaviors. To help identify new areas of exploration, in this section, I describe a design framework for sleep technologies. I also discuss a number of design considerations and opportunities for technology designers who are interested in working in this space.

5.4.1. Design Framework

Based on the literature review and formative studies, we have mapped out a preliminary design framework for technologies to encourage and support healthy sleep behaviors. Our intent was to help technology designers, researchers, and clinicians in understanding the spectrum of possibilities and identify the current gaps in the design space. Our proposed design framework consists of six dimensions, within which are several elements. It is possible for a technology to have multiple elements within each dimension. The six dimensions are:

- **Goal**: What is the goal of the system relative to sleep habits? Elements include: diagnosis; treatment; monitoring; waking; and sleep inducing.
- **Features**: What are the primary features of the specified system? This includes features for: tracking sleep information; persuasion; education on aspects of healthy sleep habits; awareness; relaxation; social applications; and entertainment applications.
- **Source**: What is the source of the design or strategy used by the sleep technology? The sources include: sleep medicine community (e.g., clinically-validated sleep therapy, guidelines from the National Sleep Foundation, sleep experts, or doctors); peer-reviewed literature (e.g., general behavior change techniques from psychology, HCI, preliminary study); other literature (e.g., books about theory, design, sleep); popular media (e.g., news, magazine, blog post); folk wisdom; or none.
- **Technology Platform**: What technology form factor does the application use? This includes: wearable technologies; stand-alone appliances (e.g., an alarm clock); mobile applications; web applications; software running on a PC or laptop; or ubiquitous computing.
- **Stakeholders**: Who are the stakeholders of the specified application? This includes people with sleep disorders; without sleep disorders; indirect stakeholders (e.g., bedmate); sleep clinicians; and sleep researchers.

- **Input Mechanism**: How does the user interact with the application? This includes: manual input by user; automatic entry via sensors or some other mechanism; or none.

We have applied the framework to 10 existing sleep-related commercial products and research projects that represent a broad spectrum of goals (Table 6). To illustrate the framework, we characterized the popular Zeo Sleep Coach appliance as an example: Zeo has the goals of *monitoring* and *waking*; the features of *tracking* and *education*; the sources of *peer-reviewed literature* (Shambroom & Fabregas, 2012); the technology platforms of *stand-alone appliance*, *wearable*, and *web;* the target users of those *without disorders*; and input mechanisms of both *input by user* and *automatic*. Table 6 shows that there was still room for opportunity to design sleep technologies that have the goals of in-home *diagnosis* and *treatment*, the features of *persuasive* and *education*, and finally, the form factor of *ubiquitous* and the input mechanism of *automatic*.

Table 6. Design framework of technologies for supporting healthy sleep behaviors.

	Goal	Features	Source	Technology Platform	Stakeholders	Input Mechanism
Actigraph	Monitoring, Diagnosis	Tracking	Sleep medicine community	Wearable, PC/Laptop	With Disorder, Clinicians, Researchers	Input by user, Automatic
WatchPAT (WatchPAT)	Monitoring, Diagnosis	Tracking	Sleep medicine community	Wearable	With Disorder, Clinicians, Researchers	Automatic
BioBrite (BioBrite)	Waking	Relaxation	Sleep medicine community	Stand-alone	With / Without Disorder	Input by user
Sleep Sound Conditioner (Sleep Sound Conditioner)	Sleep Inducing	Relaxation	Peer-reviewed literature	Mobile	Without Disorder	Input by user
Zeo	Monitoring, Waking	Tracking, Awareness	Peer-reviewed literature	Stand-alone, Wearable, Web	Without Disorder	Input by user, Automatic
BuddyClock (Kim et al., 2008)	Waking	Persuasion, Social	Peer-reviewed literature	Stand-alone	Without Disorder	Input by user
Reverse Alarm Clock (Ozenc et al., 2007)	Monitoring, Waking	Tracking, Persuasion, Education, Social	Peer-reviewed literature	Stand-alone	Without Disorder	Input by user, Automatic
Clocky (Clocky)	Waking	Persuasion, Entertainment	Popular media	Stand-alone	Without Disorder	Input by user
Sleep Cycle (Sleep Cycle)	Monitoring, Waking	Tracking	Popular media	Mobile	Without Disorder	Input by user, Automatic
Standard Alarm Clock	Waking	None	Folk wisdom	Stand-alone, Mobile	Without Disorder	Input by user

The framework also reveals that not many of these technologies, including Zeo, have been validated nor recommended by the sleep medicine community although many have the source of peer-reviewed literature. We note that not many technologies have the luxury of undergoing a thorough clinical trial, which limits their ability to be recommended by the sleep medicine community. But at the same time, we do want to emphasize that clinical perspectives toward technology interventions may not always be straightforward. This is exemplified by the actigraphy device, which took 20 years of research before the American Academy of Sleep Medicine officially announced that the device "may be a cost-effective method for assessing specific sleep disorders" (Ancoli-Israel et al., 2003).

5.4.2. Considerations and Opportunities

In this section, I outline some additional considerations that we uncovered as a result of this research and how they impact design opportunities for technologies that support healthy sleep behaviors.

5.4.2.1. Tracking Sleep Trends over Time

Tracking sleep is an important feature for technologies that aim to help monitor or aid in diagnosis. For those without sleep disorders, tracking sleep is not necessarily required, but can help them increase their awareness and encourage healthy behavior change, similar to other health- and wellness-related applications (e.g., a pedometer). For those who have sleep problems, tracking sleep is helpful for both clinicians to diagnose problems and for individuals to reflect on their sleep habits. However, most existing products rely on wearable sensors or continuous manual input, which many participants stated were undesirable. While highly motivated people who need to diagnose serious sleep problems may be willing to wear these devices for a short duration, wearable sensing solutions are often uncomfortable and people may simply forget to wear (or charge) them, making them unsuitable for long-term sleep tracking. This presents an opportunity to develop unobtrusive solutions that allow users to monitor their sleep schedules without requiring them to wear anything.

The sleep experts noted that precise sleep measurements were not necessarily needed to have a meaningful picture of sleep behaviors and trends. For this reason, we believe that a reasonable compromise can be made between accuracy of sleep data and unobtrusiveness of sensing. A remote sleep sensing tool that has one or more sensors (e.g., pressure, passive infrared motion) may provide a sufficient estimate of objective sleep data, such as frequency and duration of sleep, number of awakenings throughout the night, and sleep latency (the time elapsed from lights out until sleep onset). In addition, technologies could monitor sleep environments and create good sleep conditions with simple sensors such as light spectrum, light intensity, audio, and temperature sensors. Such tools can be incorporated into a smart home control system to assess the environment and automatically adjust the room temperature, for example, to prepare for bedtime.

Because much of the sleep-related sensing is likely to take place in the bedroom, there are often constraints such as other occupants in bed (e.g., a sleep partner or a pet). Not only can this create confusion for the sensors, but it can also raise privacy concerns. Technologies to support sleep should protect users' and any sleep partners' privacy by making it clear whose data the system is collecting, when the data is being collected, where the data is being saved, and who has access to it.

5.4.2.2. Persuasive Technology for Healthy Sleep Behavior

Interview participants had a wide range of goals for improving their sleep, such as improving the consistency of their sleep patterns, becoming a morning person, and breaking bad habits that affect sleep. To help people accomplish these goals, persuasive technologies may help encourage healthy sleep behaviors. Persuasive techniques could be incorporated into games, social support systems, goal-setting features, or timely reminders, all of which can motivate people to improve their sleep.

Keeping a consistent to-bed and wake time is an example of a sleep-related health goal. Many survey respondents and interview participants had good intentions for going to bed at a reasonable time, but would often lose track of time or become distracted by books, computers, video games, or TV. In addition, people with insomnia or delayed sleep phase syndrome often cannot go to sleep when they want to. One way to help people deal with this problem is to have them set healthy sleep goals (i.e., consistent to-bed and wake times) (Locke & Latham, 2002) and increase the visibility of these goals through visualization of the data and timely reminders.

To facilitate goal commitment, a persuasive sleep application may guide users to commit to an achievable goal through setting an alarm every day, with the default setting being the suggested wake time. The user's sleep goal can also be identified based on a prescribed sleep schedule from a sleep specialist. The sleep goals can then be added to other applications (e.g., an online calendar system, smart home control system) to provide ambient notice when bedtime nears to help users get ready for sleep.

5.4.2.3. Tensions between Technology and Sleep

While we explored design opportunities for technology to support people's healthy sleep behaviors, a number of study participants expressed their concerns toward the instrumental examination of sleep. Technology that suggests that a person should sleep more or less, should not drink coffee or alcohol at night, should turn off the TV or computer, and should go to bed at a certain time may be met with resistance. Thus, it is important to consider how to design the technology to help people become more mindful of their sleep and behaviors that affect sleep while not feeling as if their lives have been invaded by these devices.

In fact, it is conceivable that a user's sleep may be hampered by the use of technologies in the bedroom, and a sleep technology could be another entry in the long list of technologies that are already being used in the bedroom that disrupt sleep. For this reason, designing sleep technologies to not impede people's sleep quality is an important consideration and a topic that we intend to explore further.

5.4.2.4. Cultural Differences in Sleep

Across different cultures, interpretations of sleep and what constitutes "normal" sleep vary. Our study participants were primarily within the U.S., and thus the study results and discussion reflect U.S.-centric perspectives on sleep. We note that a study of the perspectives of people in countries with a *siesta* culture where taking an afternoon nap is prominent may have drawn very different design ideas. Similarly, cultural influences on infant sleep vary. Parents' co-sleeping with infants/children is practiced in Asian cultures where interdependence and group harmony are valued. American parents, on the other hand, generally seek separateness through enforced solitary sleep of their infants and children. Design specifications of sleep monitoring may be much different depending on the target audience's family context and cultural background.

5.5. Chapter 5 Summary

I sought to uncover how computing technologies could be used to help people set and maintain healthy sleep behaviors. To understand this design space, my colleagues and I conducted a contextual inquiry, online survey, and interviews on the opportunities for and feasibility of technologies to support healthy sleep. Based on our results, we proposed a design framework for technologies that support healthy sleep behaviors and discussed considerations and opportunities that can be used to develop further ideas in this domain.

This formative work helped me build a solid foundation in the intersection of sleep and technology. In particular, I learned that many factors—both behavioral and environmental factors—could influence sleep quality and that people need to pay attention to these factors to find unrecognized problems. Furthermore, the survey and interview results suggest that people often have misconceptions about what can help them to sleep better, and that there is room for technology to help bridge the gap by providing accurate and customized sleep hygiene information. Many people showed interests in improving their sleep quality by adopting sleep technology such as sleep tracking technology. However, others were concerned about the general idea of instrumenting technology to improve sleep because it can easily become another technology that disrupts sleep. Privacy concerns were also raised due to the bedroom environment where sleep monitoring takes place.

Based on this research, we identified crucial qualities of technology for sleep—simplicity, unobtrusiveness, and privacy—which should be accounted for when designing technology for sleep.

Chapter 6
Design and Evaluation of the SleepTight System

In Chapter 4, I described people's current practices of self-monitoring, focusing on common pitfalls, workarounds, and open challenges. I also discussed implications for designing effective self-monitoring technology. In Chapter 5, I described background on sleep and opportunities for the design of sleep technology. In this chapter, I describe the design and evaluation of the SleepTight system to address some of the challenges identified earlier. I begin by proposing SleepTight's three design goals drawn from theories, literature review, and implications from formative studies. I then describe the SleepTight system design focusing on how each design goal influenced the design details. I report on evaluation study design and results from the study. I discuss lessons learned to design effective self-monitoring technology that emerged from SleepTight system design and evaluation study.

6.1. SleepTight Design and Implementation

Many factors, such as meals, exercise, caffeine, alcohol, tobacco, and medication, can significantly impact sleep quality. Monitoring and reflecting on these factors and their impact on sleep can be enlightening, but they are difficult to track automatically. On the other hand, manual tracking could provide much flexibility on what to capture and promote awareness of the target behavior. Thus, I examined ways to make it easier for people to manually capture these sleep-related factors along with sleep quality. As part of this research, I designed and developed SleepTight, a lightweight self-monitoring application widget that helps people capture and reflect on sleep behaviors.

SleepTight supports the three phases of self-monitoring technology process illustrated in Figure 8, which I built based on Li et al.'s stage-based model of personal informatics systems[11] (Li et al., 2010). The three phases of self-monitoring technology were composed of (1) configuration, (2) data capture, and (3) feedback. During the configuration phase, people customize items to be tracked. During the data capture phase, people track data easily and accurately. During the feedback phase, people identify insights from their tracking data. The difference between this model and the stage-based model is when the *reflection* occurs. I learned from the formative study of Quantified-Selfers that self-reflection occurs not only when people look at or explore the collected data (feedback), but also when they *capture* data, especially when data capture is done manually (Choe et al., 2014a). Because integrating access at the time of capture can help increase awareness and spontaneous reflection on the data (Kientz, 2011), data capture and feedback phases are opportune moments to augment self-awareness and self-reflection.

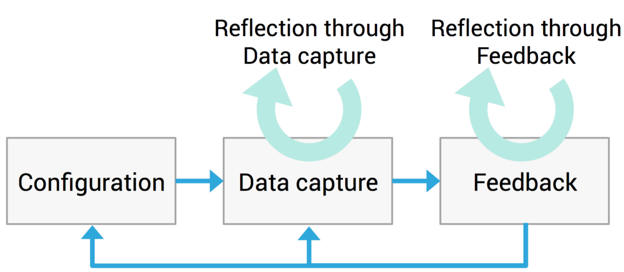


Figure 8. Three phases of self-monitoring technology process: (1) configuration; (2) data capture; and (3) feedback. When people need to add or reconfigure items to be tracked, they may go back to the configuration phase. Self-reflection happens throughout the data capture and feedback phases.

In the following, I describe the design goals that the SleepTight system aimed to support for each of these phases.

[11] The initial stage-based model of personal informatics systems composed of five stages—(1) preparation, (2) collection, (3) integration, (4) reflection, and (5) action. The stage-based model is linear and iterative.

6.1.1. Design Goals

The first design goal (G1) was to enable people to capture both ***target behaviors*** and ***triggers*** that are likely to influence the target behaviors. During the configuration phase, novice self-trackers make the common mistake of tracking only the target behaviors and not the potential contributing factors or context (Choe et al., 2014a). People make this mistake because they do not know what to track. Moreover, existing tools rarely support capturing both target behaviors and contributing factors. Thus, people would miss vital information on how to improve the target behaviors. To address this problem, our first design goal was to provide people with information about what to capture, including both target behaviors and contributing factors based on the sleep hygiene literature (Kryger, Roth, & Dement, 2010).

The second design goal (G2) was to lower the capture burden during the data capture phase thereby creating a consistent capturing habit. The captured data points must be accurate to enable effective self-reflection. Enhancing data accuracy in manual tracking is challenging because adherence to manual tracking is typically low. Studies of patients' diary compliance suggest that people may fail to complete manual journaling as instructed and that they may fake or backfill (i.e., batch completing) written entries so as to give the appearance of good compliance (Mazze et al., 1984; Stone et al., 2003). If people are not compliant with tracking protocols, then the captured data is subject to recall bias. At the same time, people may give up self-monitoring entirely if data capture imposes too much burden. Therefore, the design goal was to make the manual tracking very easy so that people can create a consistent capturing habit and collect many, accurate data points. In addition, to maintain high data accuracy, I aimed to prevent people from backfilling.

The last design goal (G3) was to provide feedback to help with self-reflection. Even a simple form of self-monitoring feedback constitutes self-reflection and contributes to behavior change (Kazdin, 1974). However, self-reflection on multiple data streams is challenging because it is complex and time-consuming. I aimed to provide feedback that allows consistent and frequent self-reflection to help people make sense of the relationships among different factors and find ways to improve their behaviors.

6.1.2. SleepTight Design

I implemented SleepTight as an Android widget and full application (app) that work closely together. I first describe SleepTight's widget design and ways in which it supports the design goals of capturing multiple factors, lowering capture burden, and providing feedback. I then detail the design of the app focusing on its three main tabs (menus)—Add Activity, Sleep Summary, and Comparison. I end this section with SleepTight's implementation details.

6.1.2.1. Leveraging App Widgets

Android's App Widgets (also known as "widgets") are miniature application views that can be embedded in other applications including the *lock screen* and *home screen* (Figure 9). Widgets are automatically updated and always running on the lock screen or home screen, and thus, people can quickly access application data without having to launch the full application.

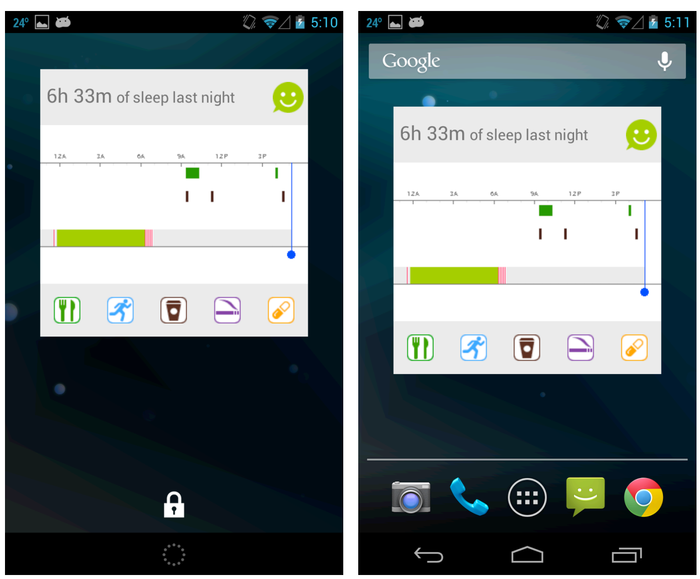


Figure 9. SleepTight implemented as an Android's app widget. SleepTight running on the Android's lock screen (left) and home screen (right).

I leveraged the App Widgets to support the design goals of lowering the capture burden (G2) and providing feedback (G3) in three ways. First, SleepTight's widget lowers capture burden because it allows people to capture data from the lock screen or home screen through a single tap. For example, to capture a caffeinated drink, people can tap on the caffeinated drink icon from the lock screen (i.e., no need to unlock the phone to capture), which would log the time stamp of when the person had the caffeinated drink (Figure 10–A). Second, SleepTight's widget enables easy access to the full application. Different parts of the widget direct people to different page on the full application. For example, tapping on the timeline region would invoke the Add Activity page (Figure 10–B). Tapping on the sleep summary line at the top would invoke the sleep summary page. If the previous night's sleep log has not been entered, tapping on the top button would invoke the page for completing the sleep diary (Figure 10–C). Lastly, SleepTight's widget serves as a glanceable display that provides visual feedback of people's activity and sleep logs. For example, once people log their sleep or activity, people can quickly access this information on their lock screen or home screen without launching the full application. The timeline on the widget shows the previous 18-hour log.

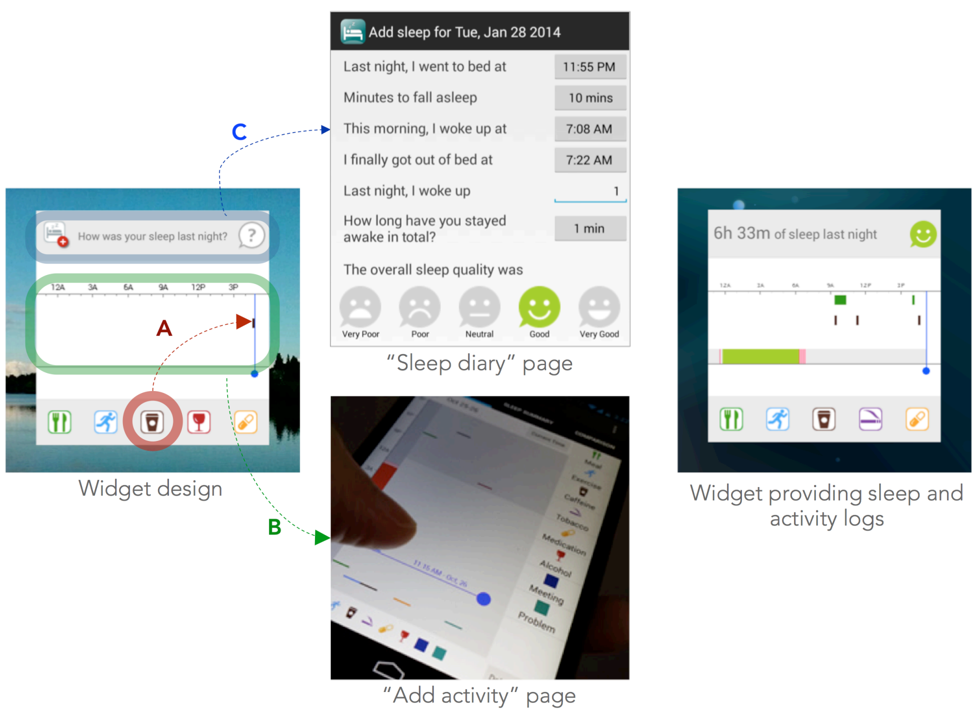


Figure 10. SleepTight's widget allows easy data capture from the lock screen or home screen (left), allows easy access to the full application pages (middle), and serves as a glanceable display (right).

6.1.2.2. Capturing Sleep Measures and Sleep-Related Factors

To learn how to improve the target behavior, SleepTight must support capturing both potential contributing factors and outcome measures. To support our first design goal of enabling people to capture both contributing factors and target behaviors (G1), SleepTight suggests what to capture. In the context of sleep, target behaviors are sleep measures (e.g., sleep quality, sleep duration, or sleep efficiency) and contributing factors are daytime and nighttime activities that are likely to trigger poor outcome behaviors. Consulting with a sleep researcher, I designed a sleep diary that consists of entries that are necessary to characterize individuals' sleep behaviors. In addition, I referred to an existing sleep diary, which incorporates six factors—meals, exercise, caffeine, alcohol, tobacco, and medication—that could highly affect sleep quality (Carney et al., 2012). Moreover, sources of sleep disturbances could be individualistic (e.g., meeting, stress, technology use). In the end, SleepTight provided the six sleep-related factors as default activities and allowed people to customize the activities to be tracked (i.e., delete unnecessary default activities and add custom activities) enabling them to track up to eight sleep-related factors. In what follows, I detail the design of the Sleep Diary page and Add Activity tab (Figure 11).

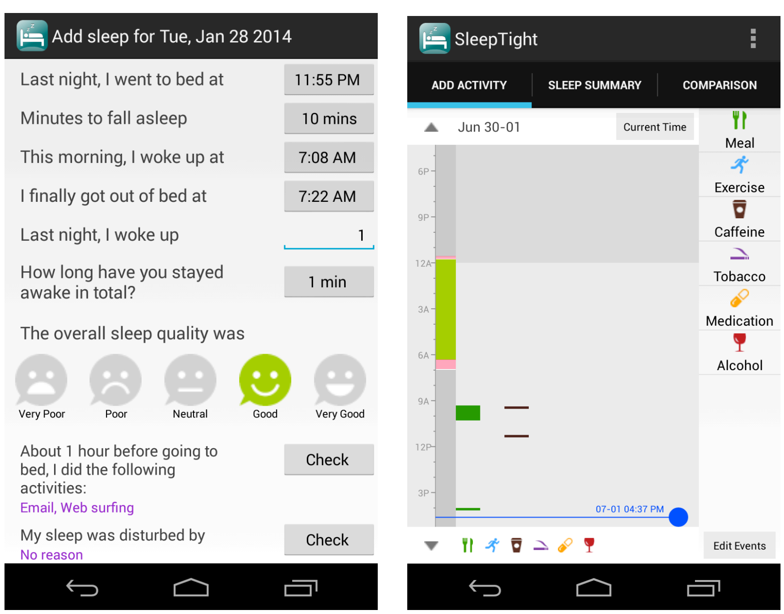


Figure 11. Sleep Diary page (left) and Add Activity tab (right). When a new Sleep Diary is entered, sleep duration, sleep quality, and sleep latency are drawn on the left bar of the Add Activity page.

Sleep Diary: SleepTight allows people to enter one sleep diary per day by clicking a link from the widget or from the app. While the general instruction is to enter the previous night's sleep diary right after a person wakes up, he or she can enter the sleep diary until midnight of the following day, thus having a 24-hour time window to complete the diary. I made this design choice to prevent backfilling and to create a consistent capturing habit (G2). Figure 11–left shows the page for entering a sleep diary consisting of 9 questions. The diary page includes required questions (subjective sleep quality, to-bed time, to-sleep time, wake-up time, out-bed time) and optional questions (number of awakenings, total duration of being awake during the sleep, nighttime activities, sleep disturbances). Subjective sleep quality is measured using a 5-point Likert-type scale, ranging from very poor to very good. I color-coded the subjective sleep quality with visuals—"red frowny" face for "very poor" and "green smiley face" for "very good." Nighttime activities (activities people conducted one hour before going to bed) and sleep disturbances are captured from a saved list of frequent items, which people can create beforehand or at the time of entry.

"Add Activity" tab: The Add Activity page is the landing page of the full SleepTight app, which can be accessed from clicking the widget's timeline (Figure 11–right). In the Add Activity page, people can record sleep-related factors and view the recorded data. The most frequent action people need to do with SleepTight is to capture sleep-related factors. For example, people usually eat meals three times a day and drink a few caffeinated beverages. If capturing a single factor is complex or if it requires multiple steps to perform the customization, people are unlikely to make use of such features. In Chapter 4, I reported on Quantified-Selfers' recommendations on how to reduce the capture burden. Incorporating these recommendations, I attempted to reduce the capture burden by lowering the data granularity and minimizing the number of steps required to track. People can record a factor with a single tap—the simplest interaction people can do with a mobile phone. To record an activity that happened in the past, they can drag the time bar (shown in blue in Figure 11–right) to the time when the activity occurred and tap the activity icon. To record the duration of an activity, they can tap-and-hold an activity icon from the list on the right to invoke the duration input dialog.

6.1.2.3. Providing Feedback about People's Sleep Behaviors

SleepTight provides two types of feedback about people's sleep behaviors—Sleep Summary and Comparison. These pages were designed to help people reflect on their sleep behaviors and sleep-related factors (G3).

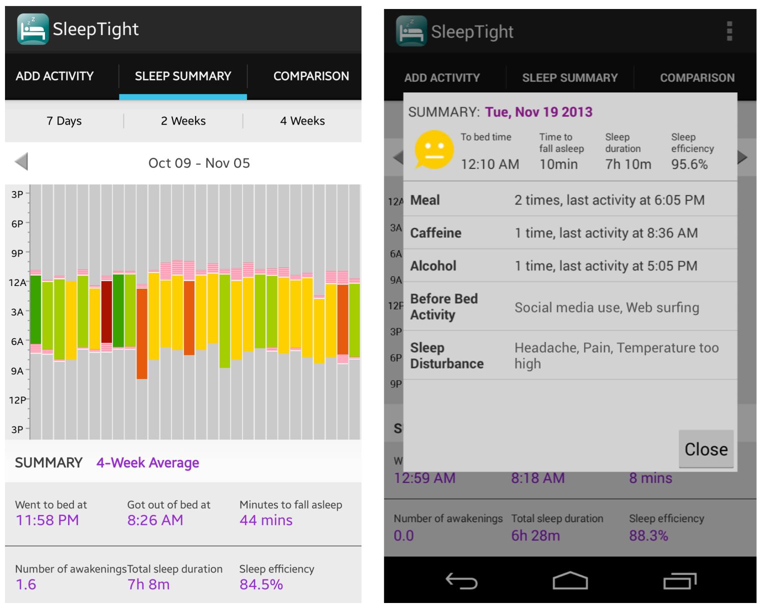


Figure 12. Sleep Summary tab (left) and Daily Summary pop-up window (right).

"Sleep Summary" tab: The Sleep Summary page visualizes the sleep pattern in terms of sleep duration and quality and provides a descriptive summary of sleep measures for a given time frame, such as the past 1 week, 2 weeks, or 4 weeks (Figure 12–left). The y-axis represents time of day (24-hour duration) and each bar on the x-axis represents a single day. The solid rectangle represents the sleep duration and its color represents sleep quality, enabling people to easily see overall sleep trends, consistency, and quality. The pink-hashed lines at the top or bottom of the solid rectangles represent the time people were lying in bed but did not actually sleep. Excessive pink-hashed lines could indicate sleep problems such as insomnia. The sleep summary page is depicted from four timestamps (to-bed time, to-sleep time, wake-up time, out-bed time) and subjective sleep quality captured from the sleep diary. SleepTight calculates the sleep efficiency (the percentage of time spent in bed that is asleep) from these timestamps. To aid in self-reflection, a detailed daily summary view (Figure 12–right) pops up when clicking one of the sleep bars from the 1-week view. The pop-up window consists of sleep summary, activity logs, nighttime activities, and sleep disturbances. The underlying assumption was that activity categories, their last timestamps, and nighttime activities affect the following sleep cycle.

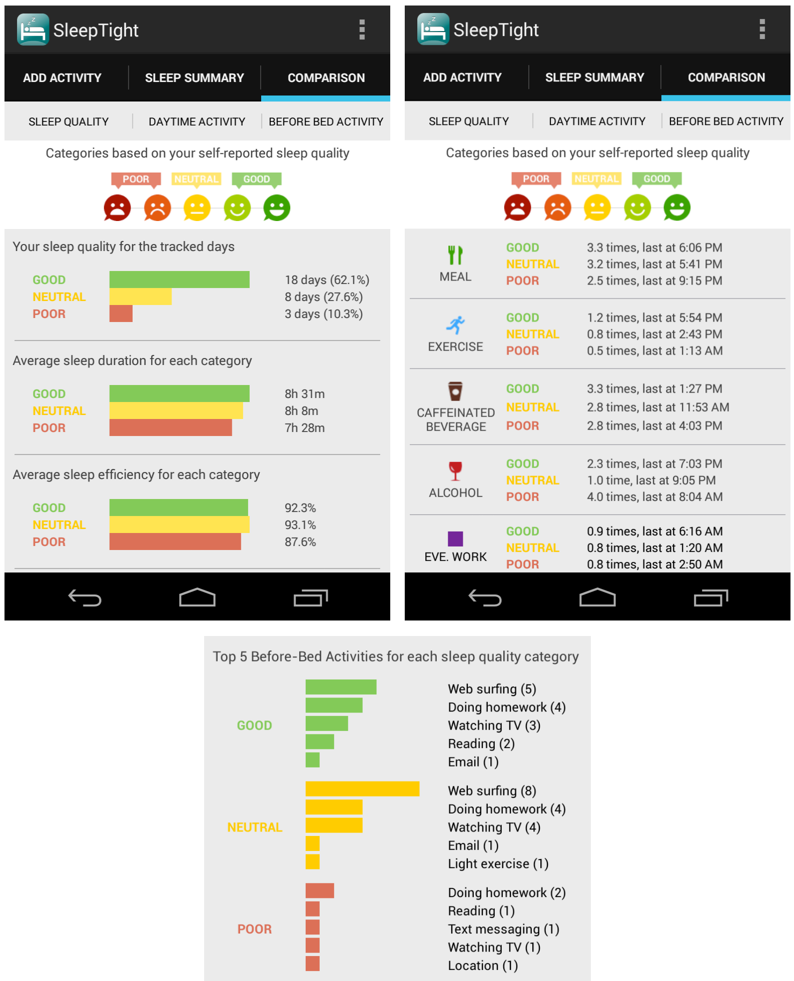


Figure 13. Comparison tab: sleep quality comparison (top-left), daytime activity comparison (top-right), and before bed activity comparison (bottom).

"Comparison" tab: From the formative study on Quantified-Selfers, I learned that a within-subject comparison was one of the most useful data exploration techniques among self-trackers (Choe et al., 2014a). I designed the Comparison tab to help people learn which activities contributed to different sleep qualities. The Comparison page groups the sleep behaviors and sleep-related factors by sleep quality. The first part of the comparison page shows the number of days for each sleep quality category—good, neutral, and poor night's sleep based on subjective sleep quality (Figure 13–top left). The next part shows the average sleep duration and sleep efficiency for each sleep quality category. For example, Figure 13–top left shows that the average sleep duration was longer when the sleep quality was good (8 hours 31 minutes) than when the sleep quality was poor (7 hours 28 minutes). Figure 13–top right shows the frequency and last timestamp of an activity categorized by sleep quality. A sleep researcher I consulted said that the time of an activity—for example, the time of the last meal of the day, caffeine consumption, or exercise—could influence sleep quality. Thus, I compared the average of last timestamps between the three sleep quality categories. For example, comparing the caffeine consumption between the days with good sleep quality and poor sleep quality (Figure 13–top right), this person had an average of 3.3 vs. 2.8 caffeinated beverages, and the last caffeine was consumed at 1:27 PM vs. 4:03 PM.

Lastly, people can compare the top 5 frequent nighttime activities (activities a person conducted during an hour of going to bed) for the days with good sleep quality against the days with neutral or poor sleep quality (Figure 13–bottom). For example, the top 5 frequent nighttime activities that people conducted were: web surfing, doing homework, watching TV, reading, and email.

6.1.3. SleepTight Implementation

SleepTight was implemented as a client-server system (Figure 14–top). The client side was implemented as a native Android App along with the lock screen and home screen widgets, sending and retrieving data from the SleepTight server on the web. The SleepTight server was implemented with Ruby on Rails using a MySQL database. The data communication between the client and server was implemented with JSON. All the graphical views were implemented with the Java 2D API, which provides rendering methods.

As shown in Figure 14–bottom, the SleepTight database consists of 8 tables such as tables for recorded sleep information, recorded activity information, sleep disturbances, and nighttime activities. For example, the sleep_tracks table contains sleep information such as to-bed time, to-sleep time, wake-up time, and out-bed time. The activity_tracks table includes the information about each activity such as activity id, activity start time, and activity end time.

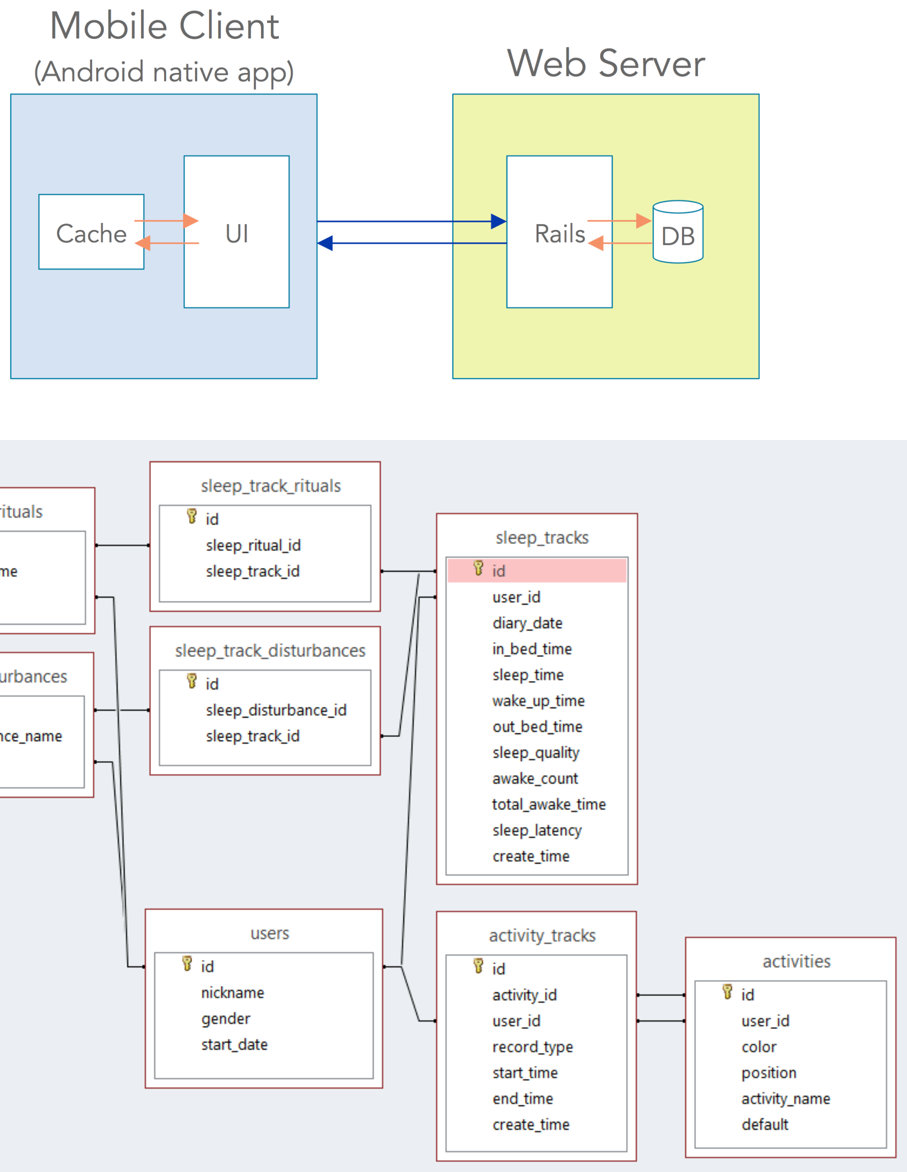


Figure 14. Schematic diagram of the client-server system (top) and database tables (bottom).

6.2. Deployment Study

I conducted a between-subjects deployment study to evaluate two versions of the SleepTight system—(1) Full System (which included the lock screen widget, home screen widget, and application) and (2) App-only System (no widgets, application only). The study included two in-person sessions (pre and post interviews) and 4 weeks of in-situ use of SleepTight. Participants were randomly assigned to one of the two conditions (Full System and App-only System), which allowed us to assess the effect of the lock screen and home screen widgets.

6.2.1. Participants

To recruit participants, I sent out recruitment emails to various mailing lists in our university. The email contained a link to the screening questionnaire. Among the 80 people who responded to the screening questionnaire, 22 participants met the following inclusion criteria (Table 7):

- Own an Android phone that runs the operating system version greater than or equal to 4.2.2 (this is the version that supports lock screen widgets)
- Have a data plan
- Do not have a diagnosed sleep disorder[12]
- Interested in learning about their sleep habits and sleep-related factors
- Interested in tracking sleep and sleep-related factors using SleepTight
- Not traveling between time zones during the 4-week study period

Among the remaining 22 participants, 41% were male (n = 9) and their ages ranged from 20 to 49 with an average age of 29.7 years old. Ten were employed full-time, five were employed part-time, six were full-time students, and one was self-employed. Our participants had varying levels of education, ranging from high school (n = 1); some college/Bachelor's degree (n = 8); some graduate work at Master's level/Master's degree (n = 9); and some graduate work at Doctoral level/Ph.D. degree (n = 4). While 91% (n = 20) of the participants had used home screen widgets, only 18% (n = 4) reported that they have experience using lock screen widgets. The slide/swipe lock was the most popular (55%, n = 12) followed by the pattern lock (n = 5) and pin/password lock (n = 4). Eighteen participants (82%) expressed that they have sleep goals. The goals were regarding waking up and going to bed at a certain time, having a consistent sleep cycle, getting more or less sleep, reducing excessive use of the snooze button, feeling rested when waking up, and having fewer interruptions during sleep. Participants' motivation to learn their current sleep behaviors was high (7.95 where 1 = not at all interested and 10 = very

[12] I excluded people with a diagnosed sleep disorder because they might be too familiar with the concept of a sleep diary and sleep monitoring, which might influence their use of SleepTight.

interested) and their confidence on current knowledge of their sleep behaviors was somewhat neutral (5.33 where 1 = not at all confident and 10 = very confident).

Table 7. Participant demographics.

ID	Sex	Age	Occupation	Sleep Partner	PSQI (/21) > 5 poor sleep	Previous Tracking Experience	Previous Widget Use[13]	Phone Lock Type
FS-1	F	30	Research scientist	Yes	3	fitbit	home; lock	pattern
FS-2	M	23	Student	No	11		home	swipe/slide
FS-3	F	32	Research associate/Student	Yes	4	fitbit	home	swipe/slide
FS-4	M	25	Software developer/Student	No	4	fitbit	home; lock	swipe/slide
FS-5	F	31	Creative director	Yes	4		home; lock	swipe/slide
FS-6	F	27	Animal husbandry technician	No	7		home	swipe/slide
FS-7	M	25	Student	No	7		home	swipe/slide
FS-8	M	29	Legal staff	No	4	ZEO; Jawbone	home	pattern
FS-9	M	28	Student	No	7		home	swipe/slide
FS-10	F	24	System analyst/Student	No	8	fitbit	home	pattern
FS-11	F	49	Knowledge management supervisor	No	5		home	swipe/slide
AS-1	F	20	Student	No	3		home	swipe/slide
AS-2	M	20	Student	Yes	6		home	PIN/password
AS-3	F	36	Academic coordinator/Student	Yes	8		don't know	swipe/slide
AS-4	F	28	Software engineer/Student	No	4	Bodymedia	home	swipe/slide
AS-5	M	34	Student	Yes	5		home; lock	PIN/password
AS-6	F	21	Student	Yes	2		home	-
AS-7	M	49	IT staff	Yes	5		home	PIN/password
AS-8	F	25	IT project manager/Student	No	8		home	pattern
AS-9	F	27	User research assistant	Yes	12		home	PIN/password
AS-10	M	49	Student	Yes	4		lock	swipe/slide
AS-11	F	22	Student	No	9		home	pattern

6.2.2. Study Procedure

The first in-lab session lasted about 90 minutes and consisted of a background survey, a standardized questionnaire on sleep quality (PSQI)[14], a semi-structured interview on factors impacting sleep, and the SleepTight installation and instructions.

[13] I asked participants if they had used widgets in the past. "Home" means experience with the home screen widget and "lock" means experience with the lock screen widget. One participant (AS-3) did not know what widgets were.

The background survey included questions regarding demographic information, smartphone use, sleep-related technology use, tracking technology use, and sleep-related goals. When participants were completing the surveys, I installed SleepTight on the participants' own mobile phone. Participants were randomly assigned to one of the two conditions as follows: half of the participants (n = 11) were assigned to the *Full System* condition (6 female; 5 male) and the other half (n = 11) were assigned to the *App-only System* condition (7 female; 4 male)[15]. After the SleepTight installation, I conducted a semi-structured interview to probe about participants' sleep habits, sleep rituals, and potential sleep-related factors. Lastly, I walked participants through a demonstration of SleepTight, helped them configure the settings, and instructed them on the use of SleepTight. I helped participants customize the main tracking activities, sleep disturbances, and activities before bedtime. Participants were also allowed to modify the settings as they use SleepTight.

For the following 4 weeks, participants were instructed to voluntarily use SleepTight. However, I did mention that participants will receive better quality feedback if they collect more data. I made it clear to the participants that the compensation is not tied to their actual usage of SleepTight. During the 4-week in-situ deployment study, I sent out weekly online questionnaires to ask if participants experienced any technical difficulties or learned any information as a result of using SleepTight.

After four weeks, participants returned to our lab for a debriefing interview and questionnaires. Questions during the exit interviews were grounded on participants' tracking logs. I probed about any outstanding behaviors and asked them to explain them. I also asked participants about their experience with SleepTight focusing on their typical usage pattern and gained information. Part of the key exit interview questions included the following items:

---

[14] The Pittsburgh Sleep Quality Index (PSQI) (Buysse et al., 1989) is a self-rated questionnaire, which assesses sleep quality and disturbances. A PSQI score > 5 provided a sensitive and specific measure of poor sleep quality, relative to clinical and laboratory measures.

[15] I will use "FS" to denote the participants assigned to the *Full System* condition, and "AS" to denote the participants assigned to the *App-only System* condition.

- Tell me about a typical SleepTight usage pattern.
- What did you learn—if any—from using SleepTight?
- What did you think about the granularity of the information captured?
- How do you feel about displaying your sleep data on the lockscreen / homescreen widget?
- How could SleepTight be improved?

I compensated participants $100 USD in gift cards in appreciation for their time. All but one participant (FS-5[16]) showed up for the exit interview. FS-5 continued using SleepTight until the end of the study, so I included her sleep and activity log data (that was sent to our database directly) in our analysis although I did not have her SleepTight usage log data.

During the exit interview, I learned that FS-4 and FS-7 removed the lock screen widget from their mobile phones. From the usage log file, I learned that FS-4 removed the lock screen and home screen widgets right after the first interview session (first day of the study) and used SleepTight as if he were assigned to the App-only System condition. Thus, I included FS-4 in the App-only System condition for the purpose of quantitative data analysis. Meanwhile, FS-7 removed the lock screen widget at the very beginning of the study and kept the home screen widget. Thus, I removed FS-7's data from the quantitative dataset[17].

6.2.3. Dataset and Analysis

The study produced a rich dataset. Data captured by participants was stored in our remote web server. I refer to this data as the ***tracking log***. In addition, I instrumented SleepTight to capture participants' usage data in a separate log file such that I could track when participants accessed the app and which page of the app was viewed. I refer to this data as the ***usage log***[18]. I used t-

---

[16] Participant naming scheme: FS or AS followed by a number (e.g., FS-1, AS-1) indicates a particular participant assigned to either of the condition.

[17] I included FS-4 and FS-7's interview data for the qualitative data analysis.

[18] AS-2 and AS-6 switched to a new phone during the study period, so I lost their usage log files (although I have their tracking log that was sent directly to the server). I also failed to load AS-4's usage log file due to a technical problem.

test and Mann–Whitney U test to analyze the tracking logs and usage logs to compare the overall usage between the two conditions.

Additionally, I audio-recorded and transcribed all initial and exit interviews. To analyze the transcripts, I used the general inductive approach, which is a way of condensing extensive and varied raw text data into a summary format and establishing clear links between the research objectives and the summary findings derived from the raw data. I read the transcripts several times to identify themes and categories regarding key questions—(1) how/whether SleepTight helped participants self-reflect on their sleep behaviors; and (2) how/whether SleepTight helped participants create consistent capturing habits. Participants' answers to these questions complemented the quantitative results from tracking and usage log analysis.

I also digitized the initial background survey, pre/post PSQI questionnaires, and weekly questionnaires. The demographic information is presented in Table 7. Lastly, the screenshots of participants' SleepTight pages and widgets gave us an overview of the types of feedback participants received during the study.

6.3. Results

From the SleepTight evaluation study, I aimed to answer RQ3, "How should we design manual self-monitoring technology for capturing and reflecting on sleep behaviors to enhance tracking adherence, data accuracy, data awareness, and self-reflection." To answer this question, I addressed the following detailed sub questions:

> **RQ3-1.** How does an easily accessible manual self-monitoring application that provides visual feedback enhance tracking adherence and data accuracy?
>
> **RQ3-2.** How does SleepTight's widget affect overall usage duration and access to information?
>
> **RQ3-3.** How does SleepTight's ability to customize tracking items affect individuals' tracking practice?

RQ3-4. How does SleepTight affect individuals' awareness and self-reflection?

RQ3-5. How does SleepTight's ability to prevent backfilling affect individuals' tracking practice?

To address RQ3-1, I compared the two versions of SleepTight (Full System and App-only System) in terms of participants' sleep diary adherence and data accuracy. To answer RQ3-2, I analyzed overall usage duration and data access frequency from the usage logs. To address RQ3-3, I analyzed what activities participants tracked besides default activities and how they were captured. To address RQ3-4, I identified how SleepTight's feedback helped participants reflect on their sleep behaviors and other related factors through the analysis of qualitative data. To address RQ3-5, I discussed the role of the 24-hour time limit in creating a consistent capturing habit based on analysis of qualitative data from exit interviews.

6.3.1. Overall Usage

In this section, I provide an overview of how participants used the two versions of SleepTight system. I report on participants' sleep diary adherence, time difference between when activities were conducted and captured, usage duration, information access frequency, and use of SleepTight's customization features.

6.3.1.1. Sleep Diary Adherence and Number of Captured Activities

To assess the efficacy of self-monitoring tool, researchers measure participants' adherence rate of self-monitoring protocol (e.g., Stone et al., 2003). I measured ***sleep diary adherence***, which is defined by the number of captured sleep diary over the course of 28 days. I reported on detailed data for each participant in Appendix A, "Overall Usage of SleepTight." Figure 15 shows the overall adherence for participants in the Full System condition and App-only System condition.

Overall, diary adherence for the Full System condition (M = 25.89, SD = 2.71) was higher than that of App-only System condition (M = 20.42, SD = 7.18), $t(14.85)$ = 2.42, p = .03. The average adherence rate was 92% for the Full System condition and 73% for the App-only System condition.

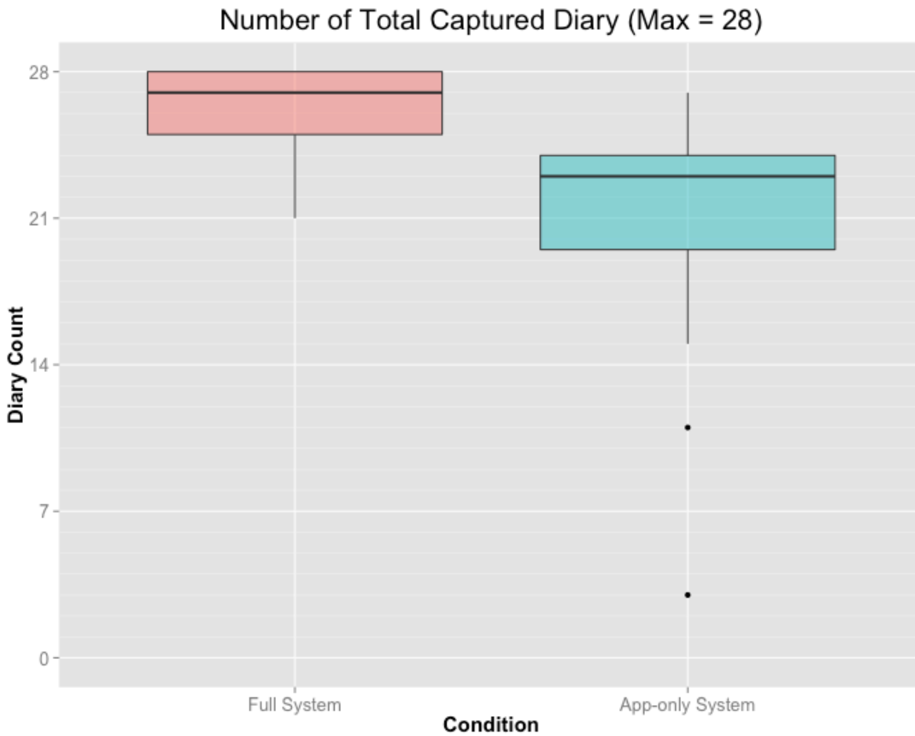


Figure 15. Number of total captured diary entries by condition. Participants in the Full System condition achieved higher tracking adherence than those in the App-only System condition (*p* = .03).

Analyzing the usage log revealed that among the diary entries captured by the participants in the Full System condition, 88% of the sleep entries was captured from either the home screen widget (77%) or the lock screen widget (11%) whereas the remaining 12% was captured from the Add Activity page from the app. This result indicates that the participants in the Full System condition heavily used the widgets to access the sleep diary page and that the widgets served as visual reminders prompting people to record the sleep diary on time.

Figure 16 shows the number of total tracked activities over the course of 28 days. On average, participants in the Full System condition tracked 152.11 activities (*SD* = 68.82) and 26.72 activities per activity category (*SD* = 13.89) while participants in the App-only System condition tracked 141.5 activities (*SD* = 78.00) and 20.32 activities per category (*SD* = 10.35). There was no significant difference between the two conditions, $t(18.41) = .33$, $p = .75$.

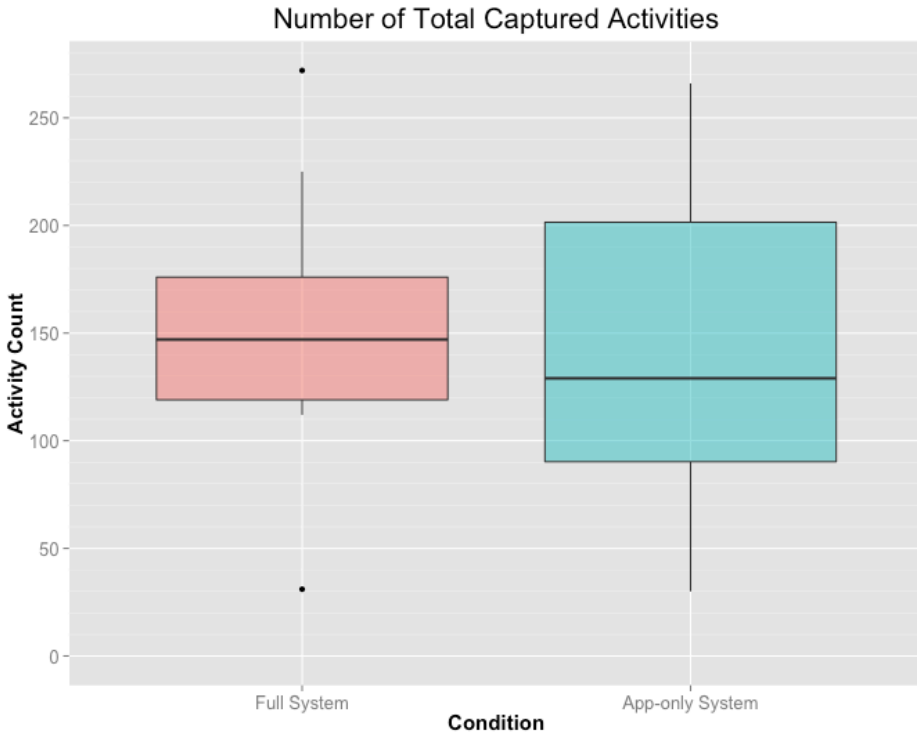


Figure 16. Number of total captured activities by condition. The difference between the two conditions was not significant (p = .75).

Among the activities captured by participants in the Full System condition, 91% of the captured activities was recorded from the Add Activity page from the app whereas the remaining 9% was recorded from either the home screen widget (7%) or the lock screen widget (2%). The fact that the majority of activities were captured from the Add Activity page indicates that participants did not use SleepTight as a real-time capturing tool and that they captured activities retrospectively.

During the exit interview, I noted that 6 participants (denoted by ‡ in Appendix A) in the App-only System condition created a shortcut icon by copying the SleepTight icon onto their home screen for easy access, thereby instrumenting one aspect of the SleepTight widget. AS-9 commented, "*I moved the icon onto one of my pages so that way when I'd be scrolling through my phone, it'd be like, don't forget to do this because I'd see it on there.*" Some participants in the App-only System condition (denoted by ∞ in Appendix A) set a separate reminder to log the sleep diary because it was hard to remember to log sleep before midnight. This indicates the need to provide a reminder or shortcut to a tracking tool to aid with timely data capture.

6.3.1.2. Usage Duration and Summary Page Access

Figure 17 shows total minutes used over the four-week study period. Participants' total usage time of the Full System condition (*Mdn* = 37.41) did not differ significantly from that of the App-only System condition (*Mdn* = 20.48), $W = 17$, $p = .074$, $r = -.43$. However, we note that Full System's total usage time does not include the time participants look at the lock screen and home screen widgets because it cannot be captured from the usage log.

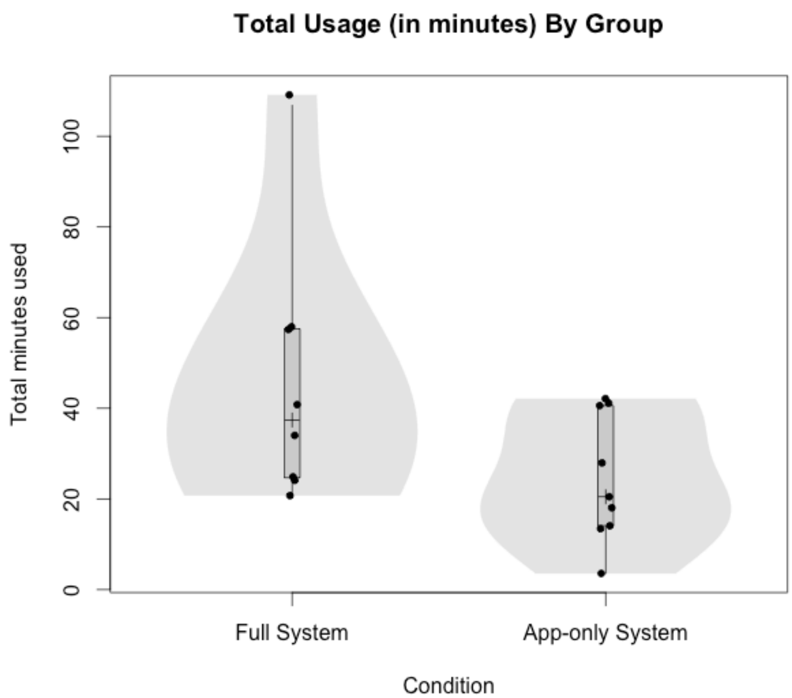


Figure 17. Total minutes used by participants in each condition. The difference in total usage between the two conditions was not significant ($p = .074$).

Next, I analyzed what features participants frequently accessed. Figure 18 shows participants' detailed usage of SleepTight's various features over the course of four-week study period. Colored lines correspond to active use of SleepTight where each color represents various events such as "add an activity" or "add sleep." Although both groups suffered from falloff, participants in the Full System condition showed more frequent use of various features than participants in the App-only System condition.

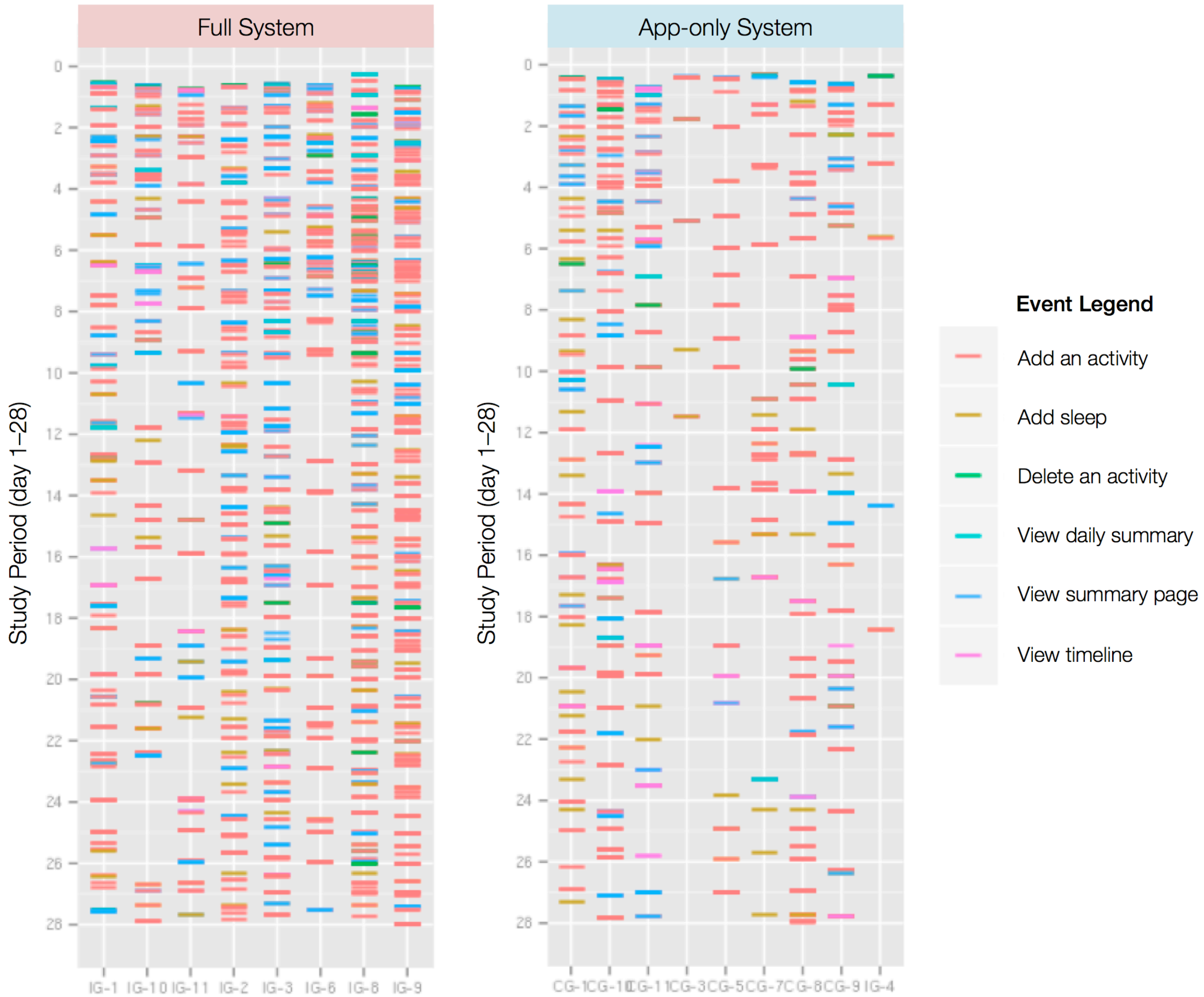


Figure 18. Chromograms of usage over the entire study period by condition. Colored lines correspond to active use of SleepTight on one of the events described in the legend.

Among the captured events, I analyzed the number of times the "Sleep Summary" page was viewed (Figure 19). Participants in the Full System condition (*Mdn* = 91.5) viewed the summary page more frequently than those in the App-only System condition (*Mdn* = 22), $W = 67$, $p = .002$, $r = -0.77$. The result indicates that the lock screen and home screen widgets helped participants view the sleep summary page often as the widgets served as a visual reminder and shortcut to the sleep summary page.

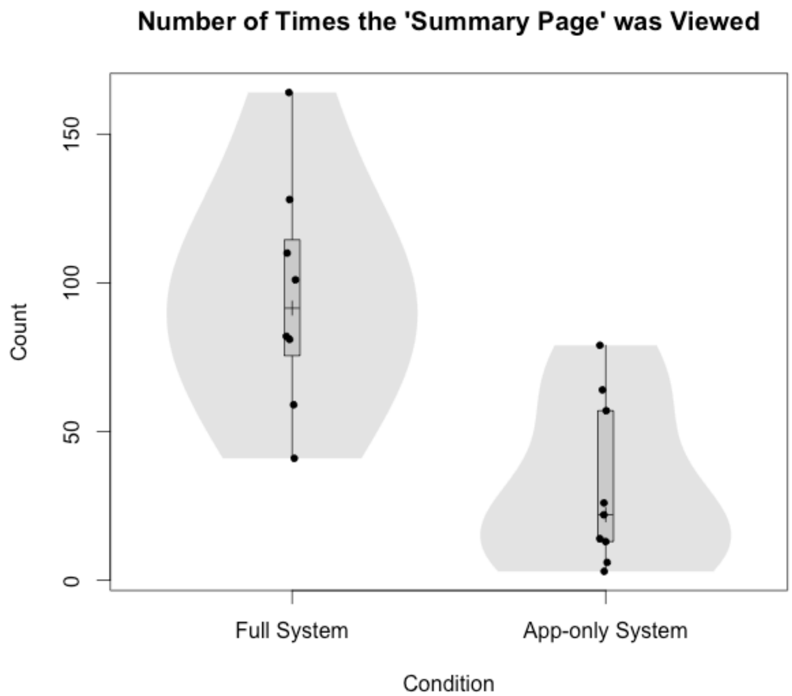


Figure 19. Number of times the sleep summary page was viewed. Participants in the Full System condition viewed the sleep summary page more frequently than those in the App-only System condition (p = .002).

6.3.1.3. Time Difference Between When Activities were Conducted and Captured

Participants could technically use SleepTight as a near real-time tracking tool. However, participants often captured activities retrospectively, thereby creating a time lag between when an activity was conducted and when the activity was actually captured. A big time lag could mean less accurate data due to the recall bias. To assess whether SleepTight helped participants capture accurate data, I analyzed the time difference between when activities were conducted and captured. Figure 20 shows the average time difference (in hours) for the Full System condition and App-only System condition. The time difference was smaller for the participants in the Full System condition (M = 7.06, SD = 3.33) than those in the App-only System condition (M = 11.66, SD = 5.00), $t(18.81) = -2.52$, p = .02. The result indicates that lock screen and home screen widgets helped participants capture activities close to their actual time.

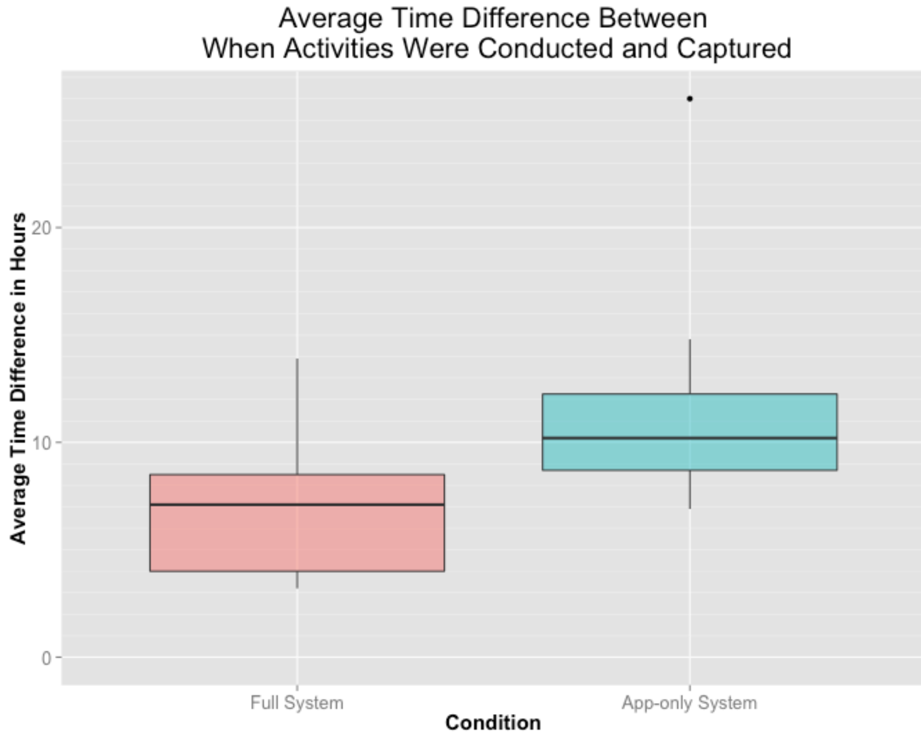


Figure 20. Average time difference between when activities were conducted and captured. Participants in the Full System condition had smaller time difference than those in the App-only System condition (*p* = .02).

During the exit interview, participants in the Full System condition mentioned that the widget served as a visual reminder to capture the sleep diary and other daytime activities. FS-2 remarked, "*having it* [widget] *here* [lock screen] *reminds me that I should be recording stuff.*" Similarly, FS-9 mentioned, "*I think the most utility I got out of this* [widget] *was that if I didn't see—If I noticed that it looked kind of sparse if it was like now, 5:00 PM and the lock screen visual thing was kind of sparse I think I'd be like, 'Oh I've got to enter my [sleep].' That's what that served for me was just realizing I hadn't filled in data.*" When participants had a short time to spare—for example, waiting for a meal or coffee that the participant just ordered, they would capture activities in near real-time.

Although widgets promoted participants in the Full System condition to record activities close to their actual time, on average, there was still a 7-hour time lag between when activities were conducted and captured. Analyzing the usage log showed that participants in both conditions tended to record daytime activities towards bedtime (Figure 21). During the exit interview, participants in both conditions confirmed that before bedtime was an opportune moment for data capture and reflection for many reasons—the memories of when things happened were still fresh in mind; capturing several activities in a row was convenient once they open the app; and real-time capturing could sometimes be awkward (e.g., having a meal with a friend) and interrupt the workflow. For these reasons, it was often at the end of the day when participants in both conditions captured a series of daytime activities and reflect on them.

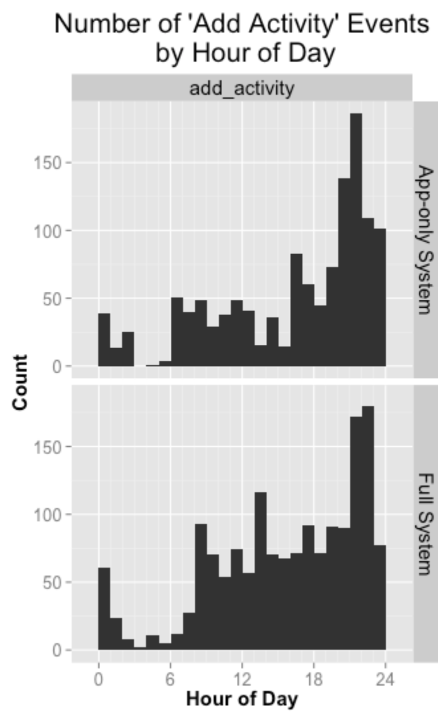


Figure 21. Number of "Add Activity" events by hour of day. The graphs show that participants in both conditions captured daytime activities toward bedtime.

6.3.1.4. Captured Activities

In addition to the six default activities SleepTight initially provided, participants added custom activities—such as nap, work (or homework), snacking, (eating) sugar, and TV watching—that they thought would influence their sleep quality. While participants were able to capture most of the custom activities using the current capture format, they had difficulty capturing activities that need to track their varying intensity such as *Stress*. FS-2 commented, *"I never actually did record stress because it was hard for me to tell when to start and when to stop. (…) It was harder for me* *to figure out the duration of the high stress point*." Some participants initially added a custom activity but soon stopped tracking it because it was futile. For example, FS-1 added *Meeting* category but soon stopped tracking it: "*Cause there were too many* [meetings]*, and I don't know that there would be any correlations anyways, 'cause if you have a meeting every day, it might not tell you a lot*" [FS-1].

Table 8. Activity categories and the number of participants who tracked each activity category. Each participant was able to capture up to 8 activities.

Activity Category	# of Participants Who Tracked the Activity
Meal*	22
Exercise*	20
Caffeine*	17
Alcohol*	15
Nap, Work/Homework	10
Snack/Sugar	9
Medication*	8
TV	6
Liquids, Game, Classes	3
Sex	2
Tobacco*, Meeting, Stress, Laptop, Driving, Climbing, Video/phone, Web surfing, Book, Music, Pet care, Shower, Tasks	1

* Default Activity

When capturing daytime activities, participants had an option to capture one timestamp by a single tap—referred to as "frequency" capture—or two time stamps (start time and end time) by press and hold—referred to as "duration" capture. Understandably, duration capture was more effortful than frequency capture. Among all the activities captured, 75% were frequency capture and 25% were duration capture. Table 9 shows how each activity was captured.

Table 9. Number of total tracked activities per activity category and how they were captured.

Activity Category	# of Tracked Activities	Frequency Capture (Count / %)		Duration Capture (Count / %)	
Meal	1205	983	82%	222	18%
Caffeine	509	507	100%	2	0%
Alcohol	214	210	98%	4	2%
Exercise	193	85	44%	108	56%
Medication	138	135	98%	3	2%
Tobacco	53	53	100%	0	0%
Custom	760	337	44%	423	56%
Total	3072	2310	75%	762	25%

Among the six default activities, participants preferred to capture *exercise* using *duration capture* more so than the other activities. FS-4 commented, "*So like for meals and snacks and coffee I didn't think the duration was as important but something like for exercise, it would be*." AS-2 also described, "*I only used the frequency for snacking and caffeine, because usually when I drink pop or a snack, it is a very quick thing. I mean, I snack in a minute or something. So I doubt those ones reflected more what I was actually doing*." Participants used duration capture for the activities that lasted for a long time, and frequency capture for the short duration activities. Although it was effortful, participants wanted to accurately capture their activities to receive accurate feedback[19]. However, when I prompted participants whether they felt the need to capture even more detailed data such as the type of food or exercise, majority of them said that it would be too much work.

6.3.2. Self-reflection with SleepTight

Capturing data is meaningful when it allows people to self-reflect on their behaviors. In this section, I describe what people learned during self-reflection with SleepTight based on qualitative analysis of weekly surveys and exit interview.

One dimension that arose from analyzing self-reflection descriptions was their ***level of certainty***—for example, whether a self-reflection description was framed as a *conclusive finding* or *hypothesis*. Conclusive findings were further categorized into a *neutral statement*; *confirmation of existing knowledge*; and *disproof of existing knowledge.* Another dimension was ***topic***—for example, whether a description was about *sleeping patterns*; *daytime activities*; *relationships between sleep and other factors*; or *tracking habits*. Table 10 shows the summary of categories and example quotes for each category.

[19] SleepTight's Comparison Tab accounts for the end time of an activity when it calculates the averages of last timestamps, so participants used duration capture for the activities that lasted a long time.

Table 10. Categories of self-reflection description and example quotes.

Level of Certainty	Topics	Example Quotes
Finding	Sleeping patterns	"That my time to go to bed is a little inconsistent [AS-5]." [Neutral] "I sleep a lot less than I thought [FS-9]." [Disproof] "I already knew I didn't sleep a lot, but this study really reinforced that [FS-4]." [Confirmation]
	Other activity patterns	"How little I eat and drink! [AS-8]" [Neutral] "That I don't drink as much alcohol as I thought I did [FS-3]" [Disproof]
	Relationships between sleep and other factors	"My sleep is poor when I am stressed out [AS-9]" [Neutral] "Caffeine doesn't have as much of an effect as I thought [AS-8]" [Disproof] "The app is starting to confirm my suspicions about when it's okay/not okay for me to drink caffeine [FS-9]." [Confirmation]
	Tracking habits	"I also reaffirmed that I am really bad a any sort of daily tracking activity [FS-7]." [Confirmation]
Hypothesis	Sleeping patterns	"I have learned that my sleep habits may not be as good as I thought they were [AS-1]."
	Relationships between sleep and other factors	"Drinking alcohol seems to lead to poor sleep. Exercise seems to lead to good sleep [FS-9]." "I have learned so far that watching TV before bed seems to affect my sleep negatively. Reading, talking, and doing simpler tasks seems to result in better sleep [AS-2]."

The majority of self-reflection descriptions were about *findings* on participants' *sleeping patterns*. I suspect that this result was due to the consistent capturing of sleep behaviors using the sleep diary. From the aggregated sleep data, participants were able to figure out their sleeping patterns such as the time they usually go to bed and wake up, and whether the sleep pattern is consistent or not. They were also able to compare within themselves the differences and similarities across weekends and weekdays and identify the ways in which the previous night's sleep affects the following night's sleep. Some participants made a value judgment on their sleeping patterns, for example, whether they got more sleep or less sleep than what they thought was appropriate.

SleepTight also increased participants' awareness of other activities besides sleep. For example, participants were more aware of their eating or drinking habits (e.g., "*I don't drink as much alcohol as I thought I did*"). They were also more aware of their nighttime routines because the sleep diary specifically asks about activities that they did before going to bed. Many participants did work-related activities until right before going to bed without knowing that it negatively impacts sleep. However, SleepTight helped them realize that they would need some time to unwind to have quality sleep as AS-10 mentioned: "*I sleep better when I have time to unwind before bed. If I go to bed directly from doing homework, my sleep is worse*."

Another type of self-reflection was about *findings* on *sleep and sleep-related factors*. Because SleepTight allows people to track multiple factors at a time, we expected participants to identify any relationships among the captured factors. Some participants made very specific observations, such as identifying the cut-off time for caffeine (e.g., "*I had a little caffeine one day at 7:30pm and I couldn't get to sleep until almost 2am, so I should probably avoid that in the future*"). But in general, most of self-reflection descriptions contained vague associations between sleep and other factors such as "*I sleep better when I have less sugar and eat more earlier* [*sic*] *in the day*."

Although few, some participants described with care what they had learned, acknowledging that there might be some flaws in their reflection. In those cases, they posed a hypothesis or showed hesitation instead of stating a conclusive finding, so I marked this type of self-reflection '*hypothesis*.' Such descriptions include "*Drinking too much alcohol seems to degrade my sleep quality* [AS-7]" and "*So far, it looks like I might be more tired Tuesdays and Wednesday?* [FS-1]" I observed this type of self-reflection from early and middle phase of the study (i.e., the first three weeks of SleepTight usage survey), but not from the late phase of the study.

Lastly, a few participants mentioned that they did not learn anything, especially during the first week. I suspect that this kind of remarks was due to a small number of data points. FS-7 said during the first week survey, "*Honestly,* [I] *haven't paid much attention to the data yet. I was holding off until I had a few more days logged before I bothered looking at it.*" I did not get this type of response during week 2–4 surveys and the exit interviews.

When I probed about the times when participants looked at the data and reflected upon it, it was often when they recorded activities and sleep. In particular, right before going to sleep turned out to be an opportune moment to do self-reflection as this time was often when they accessed SleepTight to track data. They often checked the visualizations SleepTight provided in the Add Activity tab and the Sleep Summary tab. The Sleep Summary tab was particularly helpful in finding aggregated sleep trends and enforcing tracking—participants said that if there was a missing bar on the Sleep Summary tab due to a missing diary, they felt bad about it. To our surprise, participants did not find the Comparison tab particularly helpful because it was "too data-centric." Participants were puzzled when the numbers from the Comparison tab provided counterintuitive findings. For example, one participant learned that the average time she had coffee was later time of the day when her sleep quality was good than when her sleep quality was poor.

In summary, participants were able to learn their sleeping patterns and other activity patterns with SleepTight. In doing so, feedback from the Add Activity tab and Sleep Summary tab was helpful. Participants were also able to identify relationships among the captured factors, but it was often due to their careful observation and self-awareness of their behavior rather than the feedback from the Comparison tab. Some participants would often jump into the causal relationships between different factors and sleep quality.

6.3.3. The Role of the 24-Hour Time Limit in Creating a Consistent Capturing Habit

The tracking data and log data suggest that the lock screen and home screen widgets helped participants remember to complete the sleep diary. While widgets served as visual reminders for Full System condition, SleepTight also enforced a 24-hour time window to complete the previous night's sleep diary for both conditions. To investigate how the time limit played a role in helping people form a consistent capturing habit, I asked participants about it during the exit interview.

I learned that the time limit forced participants to capture more data more accurately. Participants recognized the advantage of having the time limit as AS-2 remarked, "*I think if there was no time limitation, I might have probably forgotten at some point or something. But I knew there was a time limit. I knew when I woke up that day, or throughout the day if I didn't fill it out in the morning, I'm like, 'I have to do it before midnight.' So it was in the back of my head, 'cause I knew I had to do it before a certain time*." However, at the same time, participants were frustrated when they could not enter data even when they wanted to. For example, FS-1 described: "*There was one day I filled it out at like 11:58* [PM] *or something, and it was lagging. And I was like, 'You have to load! You have* *to load!'"* Recognizing that they cannot enter data past midnight, participants scheduled to log sleep data some time in the morning or during the day, which naturally became a habit. Also, this was when some participants in the App-only System condition started setting reminders to ensure that they do not forget to enter data (—no one in the Full System condition set a reminder.)

When routines broke (e.g., being sick, family visit, extremely busy schedule), having the time limit was not very helpful because people forgot about the time limit altogether with the fact that they should fill out the diary. For example, AS-9 diligently filled out the diary for the first few weeks: "*It* [time limit] *just helped me get in the habit of going in there to record the information. And I think it was when the moving stuff happened, and my schedule just kinda got all out of whack. And I was so distracted by everything else going on. That's when I forgot*."

A few participants—especially those who usually went to bed after midnight—did not like the time limit setup at all. AS-7 and AS-11 both said that the time limit probably caused loss of data because they were not able to enter sleep diary when they wanted to. We also learned that the time limit had little effect on creating a capturing habit for those who usually filled out the sleep diary in the morning time. Although few, this type of participant started to complete the diary in the morning from the very beginning of the study, and they never encountered any situation where they could not enter a diary. Some of these participants even forgot that there was such time limit for filling out the sleep diary.

A 24-hour time window seemed like a reasonable duration to be able to accurately reconstruct previous night's sleep schedule. FS-1 summed up the advantages of having the time limit by saying, "*It was pretty easy to construct that all during the day. But I think if I'd tried to reconstruct a week, it would've been disastrous. (…) And if I postponed logging sleep, I probably would've postponed logging the other things. And then those things would've been less accurate, and the sleep would've been less accurate. The whole thing would've been less useful*." I suspect that having to log sleep every day made the rest of the data (i.e., daytime activities) more accurate because participants might as well log everything else when they access SleepTight.

6.4. Discussion

In the following, I discuss lessons learned from this study and implications for self-monitoring technology design. I initially had three design goals: (1) enable people to capture both target behaviors and triggers that are likely to influence the target behaviors, (2) lower the capture burden and create a consistent capturing habit, and (3) provide feedback to help with self-reflection. I begin by revisiting these three design goals, and discuss design implications in relation to the design goals. I also discuss other implications for self-monitoring technology design—supporting customization; projecting personal data in a positive light; and identifying and capturing anomalies—which emerged from data analysis.

6.4.1. Capturing both Target Behaviors and Triggers

SleepTight allowed participants to capture target behaviors (e.g., sleep quality, duration, time to bed) through the daily sleep diary. SleepTight also allowed people to define and capture potential triggers that could influence the target behaviors (e.g., meal, exercise, caffeine, tobacco, other custom activities) through widgets and the Add Activity tab. However, identifying the target behaviors (outcome measures) and triggers are not always straightforward because what seems to be the outcome behavior could actually be a trigger and vice versa (e.g., lack of sleep results in an increased caffeine intake). Working closely with a domain expert (sleep clinician in our case) is crucial in configuring the initial tracking environment and determining the default activities to be tracked.

6.4.2. Lowering the Capture Burden and Creating a Consistent Capturing Habit

SleepTight's Full System version resulted in higher sleep diary adherence rate than App-only System. One way of interpreting this result is that widgets lowered the burden to access the sleep diary. Lock screen and home screen widgets had the direct link to the sleep diary page, which served as visual reminders. The study results showed the power of visual reminder on people's diary adherence rate.

Participants in both conditions captured many daytime activities. I suspect that participants in both conditions experienced the ceiling effect (there are a limited number of activities to capture in a day). I learned that once participants accessed SleepTight's Add Activity tab, they captured many activities in a row because the user interface to capture an activity was very easy (dragging a time bar followed by a single tapping.)

Previous research has already shown that setting a notification reminder could drastically increase the response rate (Bentley & Tollmar, 2013), so I purposefully did not implement this functionality in the SleepTight system. Instead, I implemented the 24-hour time limit for the sleep diary and learned from the exit interview that having the time limit could help participants create a consistent capturing habit. Less rigid time limit setup (e.g., the ability to customize the time limit depending on the individuals' sleep time) could address some participants' frustration. Because I implemented this feature in both conditions to test its feasibility, it warrants future research efforts to test the efficacy of the time limit setup through more rigorous study design.

In summary, widgets serving as visual reminders, simple steps to capture activities, and time limit setup all contributed to lower the capture burden and to create a consistent capturing habit. This result also suggests that providing easy access to a capturing tool is a key to lower the capture burden.

6.4.3. Providing Feedback to Help with Self-reflection

Participants reflected on their data while entering data and right after they entered data. Feedback provided on the widgets, the Add Activity tab, and the Sleep Summary tab were particularly helpful to improve participants' awareness of sleeping patterns. Aside from when they were entering data, however, participants rarely took time to look at the feedback and ponder upon it. Feedback provided on the Comparison tab (Figure 13) was hard to interpret and underused. Participants wanted to have a clear idea about how different factors affect their sleep quality, but SleepTight could not answer this question beyond providing descriptive statistics. I believe that SleepTight can be best used for the purpose of hypothesis generation rather than hypothesis testing. Also, it would require a more advanced statistical approach (e.g., linear regression) to model the relationships between multiple explanatory variables and dependent variable, which is interesting and important future work.

6.4.4. Supporting Customizability

Allowing people to add custom activities turned out to be an important feature because many participants added custom activities and removed default activities that were not applicable to them. However, participants were constrained by the SleepTight's capturing format (i.e., capturing the two timestamps of an activity). With this format, they could capture the frequency or duration of an activity, but not the intensity of an activity, which made it hard to track factors like stress or pain. Moreover, SleepTight did not support capturing unstructured data such as free text annotation. It warrants future research efforts to identify common data structures to allow more flexible data capturing.

Enforcing the 24-hour time limit to prevent backfilling could be improved by allowing people to customize the cut-off time according to individuals' sleep patterns. Instead of the midnight deadline, the cut-off time could be slightly later than a person's usual to-bed time while keeping the total available duration to 24 hours. Being able to customize the cut-off time would make people less frustrated while still be helpful to create a capturing habit.

6.4.5. Projecting Personal Data onto Widgets in a Positive Light

Self-monitoring tools are designed to capture personal things. As SleepTight supports capturing personal behaviors and projects the captured data onto lock screen and home screen widgets, some participants mentioned that they became too self-conscious. A recent study on people's phone unlocking behaviors showed that on average, people unlock their phones between 4.82–105.25 times per day (Truong, Shihipar, & Wigdor, 2014). Because SleepTight projects individuals' sleep, alcohol, caffeine, tobacco, and other behaviors on the widgets, it could cause added stress for participants, especially when the data shows negative information about oneself. Not wanting to see negative information (e.g., a bright red frowny face for negative sleep quality) every time a person unlocks the phone, some participants entered skewed data overestimating their behaviors. FS-2 commented, "*I felt like it was sort of broadcasting my sleep from the night before, so that might have affected like recording bad sleep, so I might have recorded it as just being neutral instead because I didn't want the bright red to be on my screen*." Self-monitoring technology runs a risk of making people become overly anxious or lie to the tools. Thus, feedback from the widget should be encouraging and yet correctly conveying the current state, which is particularly challenging when the data contains negative information about oneself.

Although few, some participants expressed privacy concerns over the lock screen widget because lock screens could be easily visible by bystanders or friends. FS-3 said that she did not want her friends to see the widget because they might judge her and say, "Oh, you get so much sleep," or "Oh, how come you couldn't sleep last night?" She felt that the widget made her obsessively think about sleep and she would rather have a neutral app like a big clock. Widgets are very effective precisely because they are very accessible. Because lock screen is a very valuable and limited space, people might have a strong preference for how they want to use that space. I acknowledge that installing a self-monitoring widget on the lock screen would not work for everyone because people have different levels of motivation, privacy concerns, and priority.

6.4.6. Identifying and Capturing Anomalies

We asked participants to collect nighttime activities with an assumption that what they do right before bed would have impact on people's sleep quality. However, it turned out that most participants do the same things—called "nighttime routines"—pretty much every night regardless of their sleep quality. Some of these activities included brushing their teeth, watching TV, talking to their spouse, or reading a book. What affected people's sleep more were the things that people did outside their routine—for example, having friends come over, travelling, or working late. Therefore, once people figure out what their nighttime routine is like, self-monitoring technology should be designed to help people collect anomalies—activities that they do outside their routine. Rare events are valuable data points.

6.5. Chapter 6 Summary

In this chapter, I presented the design and evaluation of the SleepTight system. To evaluate the efficacy of the lock screen and home screen widget for improving tracking adherence, data accuracy, and self-reflection, I conducted a between-subjects study comparing the Full System (lock screen, home screen, application) condition to App-only System condition. The evaluation study suggested that the widgets served as visual reminders, which helped people collect more data, more accurately. Participants valued the SleepTight's customization features. They added custom items they thought were impacting sleep quality although some activities or items were hard to track using the timestamp format. Participants were able to reflect upon their sleep behaviors and sleep-related activities and identified findings and hypotheses about their sleeping patterns, other activity patterns, and relationships among multiple factors. Participants showed mixed feelings about the 24-hour time limit—although it was helpful, some participants occasionally could not enter data even when they wanted to. Drawn from the evaluation study results, I discussed several implications for self-monitoring technology design including the need to (1) capture both target behaviors and potential triggers; (2) lower the capture burden and create a consistent capturing habit; (3) provide feedback to help with self-reflection; (4) support customization; (5) project personal data in a positive light; and (6) identify and capture anomalies.

Chapter 7
Persuasive Performance Feedback: The Effect of Framing on Self-Efficacy

In the previous chapters, I argued for the importance of designing effective self-monitoring feedback. The goal of self-monitoring is not simply to quantify one's behavior, but to improve it. Therefore, self-monitoring feedback needs to convey information to help people make health-enhancing, self-beneficial decisions. In this chapter, I address the last research question (RQ4), "how should we design ***persuasive performance feedback***" with an aim to nudge people toward positive health behaviors. To address this research question, I identified three types of framing that are applicable to present self-monitoring feedback and evaluated the their effects on individuals' self-efficacy. The framings were chosen based on previous research on the determinants of reactive effects of self-monitoring (Chapter 2.1) and Framing effects (Chapter 2.2).

7.1. Introduction

Consumer self-monitoring technologies for health have proliferated in recent years. Examples include pedometers for step count (*fitbit*), sleep tracking devices for sleep duration and quality (*fitbit*; *Jawbone UP*), electronic scales for weight and body fat percentage (*Aria*), and glucometers for blood glucose level (*The OneTouch UltraMini*). These self-monitoring technologies often provide real-time feedback on a user's current progress, which I call ***performance feedback***. Performance feedback provided in varied ways (e.g., text, visual, positive light, negative light) could foster changes in behavior under observation, which is referred to as ***reactivity*** (or ***reactive effect***) (Nelson & Hayes, 1981). Reactivity often manifests in the frequency of the target behavior changing in a desired direction. When properly combined with goal setting (Kazdin, 1974), real-time performance feedback is a powerful driver to increase reactivity for health behavior change.

The objective in this research was to identify ways to present performance feedback that will nudge people toward healthy behaviors. In creating influential, persuasive performance feedback, I was inspired by the well-known "Framing effects (Tversky & Kahneman, 1981)." The key idea is that the way information is framed (e.g., highlighting information in a positive light vs. negative light) influences people's behavior. A classic example is in the framing of the odds of a grueling operation: many would prefer an operation of where the outcome is "90 out of 100 are *alive* after five years" than one where "10 out of 100 are *dead* after five years." Although these two options contain the same information from an expected value perspective, people—even experts (i.e., doctors)—are systematically subject to Framing effects and more apt to prefer surgery when described by survival rate than death rate (Marteau, 1989).

Drawing from prior literature and existing self-monitoring technology designs, I identified three types of framing that can be applicable in presenting performance feedback in conjunction with a daily goal. First, I modified ***valence of performance***, a classic framing on positive versus negative outcomes as introduced in the example on surgical outcomes. Second, I modified ***presentation type*** comparing text-only feedback with text combined with visual feedback. Although most prior framing research examined framing using text descriptions, I investigated whether visual elements such as colors and figures can make the valence of performance even more salient than text-only valence descriptions. Third, I varied ***data unit***, which has been explored in the context of medical risk communication (e.g., communicating genetic abnormalities (Grimes & Snively, 1999)). I studied the effect of these framings using a hypothetical scenario of a person receiving his/her daily step count from a pedometer. I chose to use the step count scenario because pedometers are widely available consumer self-monitoring technologies, and thus people could easily understand the meaning of its feedback (i.e., step counts) without training.

In an effort to identify the kind of framing that can nudge people toward healthy behaviors, I conducted an online experiment in which I tested the effect of the three framings described above. In what follows, I describe the research questions, detail the study method and data analysis method, and report on the results. Based on our findings, I suggest design considerations for creating persuasive performance feedback.

7.2. Research Questions and Experiment Design

I explored whether the framing of feedback on an individual's performance affects his/her self-efficacy. I examined the effects of three types of framing: *valence, presentation type*, and *data unit*. I also suspected that these framings might have different effects at various levels of progress toward one's goals (*distance to the goal*), such as the beginning phase or the ending phase. I put these research ideas into the following sub research questions:

> **RQ4-1:** How do different types of performance feedback framing—(1) valence, (2) presentation type, and (3) data unit—influence an individual's self-efficacy?
>
> **RQ4-2:** Does the distance to the person's goal influence the Framing effect?

To examine these research questions, I designed a 2 (valence: achieved vs. remaining) x 2 (presentation type: text-only vs. text with visuals) x 2 (data unit: raw vs. percentage) x 2 (distance to the goal: low achievement (25%) vs. high achievement (75%)) mixed design with repeated measures. Valence (VALENCE), presentation type (PRESENTATION), and data unit (UNIT) were between-subjects factors and distance to the goal (DIST) was a within-subjects factor, thereby forming eight different conditions (Table 11).

7.3. Method

I conducted a mixed design study as an online experiment. I conducted several iterations with pilot participants before deriving the final questions, scenarios, and feedback designs that are presented here. I explored the two research questions in the context of receiving performance feedback on daily step counts from a pedometer where a daily goal was set to 10,000 steps. I chose the step count scenario with the daily goal of taking 10,000 steps because this scenario was relatively easy to understand and could be applicable to a wide audience. Although "10,000 steps a day" is not a magic number, it is easy to understand and applicable for most people to be active, considering the U.S. average daily step count is 5,100 (Bassett, Wyatt, Thompson, Peters, & Hill, 2010).

7.3.1. Survey Contents and Study Conditions

I created online surveys for the eight conditions. Each survey consisted of three sections—(1) interest in achieving 10,000 steps daily, (2) self-efficacy questions, and (3) demographic questions. To help participants understand the time and effort to achieve 10,000 steps daily, I asked, "Approximately how far do you think is 10,000 steps?" and revealed the answer (5 miles) on the next page. I also explained the time it typically takes to reach 10,000 steps—1 hour 40 minutes for moderate intensity (100 steps per minute), and 1 hour 17 minutes for vigorous intensity (130 steps per minute). I then asked the participants about their interest in taking 10,000 steps daily to maintain a desirable level of physical activity for health.

To situate participants in the context of receiving performance feedback, I provided the following hypothetical scenario:

> *Research has suggested taking 10,000 steps daily for maintaining a desirable level of physical activity for health. Suppose you purchased a pedometer (step counter) to monitor your step count, and set* ***a daily goal of 10,000 steps****. You need to wear it every day in your pocket or on your waist, and it gives you real-time feedback of the* **[*remaining* | *achieved*]** *steps toward*

your goal. [The wording (i.e., "remaining" or "achieved") was modified accordingly for each condition.]

Then I showed step count feedback (Table 11) as an image. I manipulated the feedback in the following manner:

- Valence of Performance: I varied the valence of performance by describing the performance using the "achieved" frame and the "remaining" frame.
- Presentation Type: I created text-only feedback and text with visual feedback. For the text with visual feedback conditions, we provided a progress bar colored in either green or magenta.
- Data Unit: I varied step count units by using raw number (steps) and percentage (%).

Table 11. Feedback manipulation for the eight conditions and the number of participants assigned to each condition for the low level of goal achievement (25%) case.

Valence of Performance	Presentation Type	Data Unit	Example feedback (2500 steps)	# of Participants Initially Assigned	# of Participants Included in the Analysis
Achieved	Text-only	Raw	2500 steps achieved	66	49
		Percentage	25% achieved	61	49
	Text with visual	Raw	2500 steps **achieved**	62	50
		Percentage	25% **achieved**	58	49
Remaining	Text-only	Raw	7500 steps remaining	65	50
		Percentage	75% remaining	68	48
	Text with visual	Raw	7500 steps **remaining**	69	58
		Percentage	75% **remaining**	62	47
			Total Number of Participants	511	400

According to the feedback manipulation, “2,500 steps achieved” in the achieved-frame conditions was equal progress to “7,500 steps remaining” in the remaining-frame conditions and to “25% achieved” in the percentage conditions. Also, the same feedback was provided with and without the visual (progress bar).

Each participant saw two feedback conditions, varying the *distance to the goal* at two levels—25% and 75%—in a randomized order. Feedback manipulation examples in Table 11 show low achievement of goal achievement (25%) case for the eight conditions. The scenario supposed that a participant is receiving the step count feedback on a *weekday at 4:30pm*, which is the time when much of the participants’ day had passed, but they could still have time to achieve their goal.

7.3.2. Measures

After showing each feedback, I measured participants’ self-efficacy by asking the following question adopted from Bandura (Bandura, 2006): “Rate how confident you are that you can achieve your daily goal as of now (4:30 PM, weekday).” Self-efficacy was measured on a 11-point Likert-like scale, where 0 = “Certain I cannot meet my goal” and 10 = “Certain I can meet my goal.” In addition, I conjectured that participants’ interest in taking 10,000 steps daily might be related to their base self-efficacy, so I measured the interest level on a 11-point Likert-like scale at the very beginning of the survey.

It was necessary for participants to understand the feedback so that they could answer the self-efficacy question based on correct understanding of the feedback. To assess whether participants correctly understood the feedback, I included a filtering question. I showed feedback illustrating “3,000 steps remaining” and asked a multiple choice comprehension question (Which of the following correctly describes the above feedback?) and provided three options—(1) Less than 50% of my daily goal remains, (2) More than 50% of my daily goal remains, and (3) None of the above. To filter out those who did not correctly understand the feedback, I placed this question before the self-efficacy question. Finally, I repeated the question, "Approximately how far do you think is 10,000 steps?" at the end of the survey to filter out those who did not pay attention to the wording of our survey.

7.4. Results

I recruited a convenience sample of 511 participants through word-of-mouth and researchers' social networks. I incentivized participation with the option to enter a drawing for one of four $25 gift cards. Participants were randomly assigned to one of the eight conditions. I removed data from 111 participants according to the following 4 exclusion criteria:

- Did not understand the feedback correctly (i.e., who got the filtering question wrong, 70 participants)
- Did not pay attention to the survey (i.e., who got the repeated 10,000 steps question wrong, 9 participants)
- iPhone/iPad user (due to a bug within the survey application, 7 participants)
- Outside of U.S. (due to the use of different distance metrics, 25 participants)
- Among the remaining 400 participants, 53% were male (n = 211) and 43% reported they have experience using a pedometer (n = 172). The participants' ages ranged from 19 to 68 with an average age of 32.7 years old.

I observed that participants' initial interest level (INTEREST) in taking 10,000 steps daily was significantly related to their self-efficacy of achieving the daily goal, $F(1, 391) = 48.64$, $p < .001$. Therefore, I used a mixed-design analysis of covariance (ANCOVA) controlling for the INTEREST as covariate.

I found a significant main effect of DIST on the self-efficacy scale, $F(1, 391) = 110.20$, $p < .001$. This result indicates that, at a set time (i.e., 4:30 pm in our scenario), people who were close to the goal (75% of the goal, $M = 7.65$) were more likely to report higher self-efficacy than those who were further from the goal (25% of the goal, *M* = 4.07). The result also indicates that I successfully manipulated DIST at two levels.

7.4.1. Effect of Valence Framing on Self-Efficacy

I found a significant main effect of VALENCE on self-efficacy scale after controlling for the effect of INTEREST, $F(1, 391) = 4.07$, $p = .04$. As shown in Figure 22, the result indicates that participants in the achieved-frame condition (*M* = 6.05, 95% CI[5.78, 6.33]) were more likely to report higher self-efficacy than those in the remaining-frame condition (*M* = 5.67, 95% CI[5.41, 5.93]).

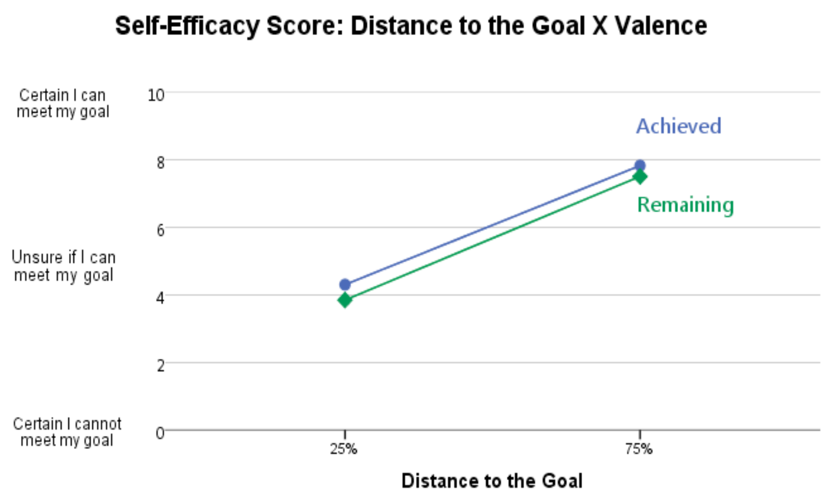


Figure 22. The effect of valence framing on self-efficacy score: participants' self-efficacy was higher when they were shown the achieved framing than remaining framing.

7.4.2. Effect of Presentation Type Framing on Self-Efficacy

I found a significant main effect of PRESENTATION on self-efficacy scale, $F(1, 391) = 7.43$, $p = .007$. As Figure 23 shows, participants in the text-only condition (*M* = 6.12, 95% CI[5.85, 6.39]) were more likely to report higher self-efficacy than those in the text with visual condition (*M* = 5.60, 95% CI[5.33, 5.86]).

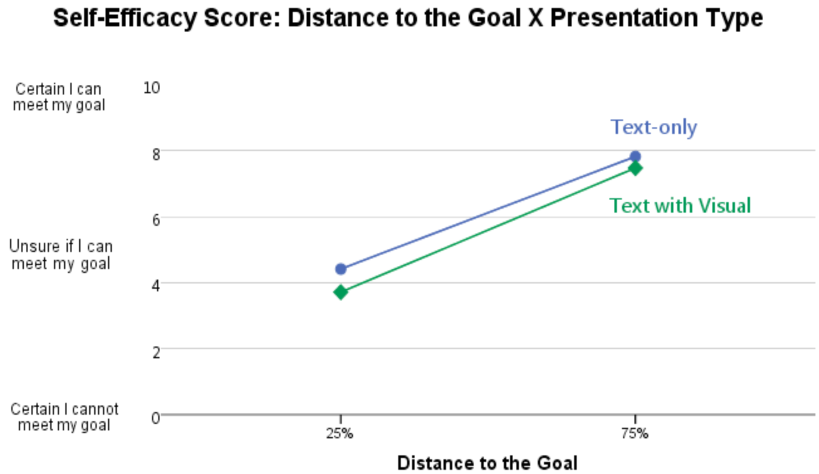


Figure 23. The main effect of presentation type on self-efficacy score: participants' self-efficacy was higher when they were shown the text-only feedback than text with visual feedback.

7.4.3. Effect of Data Unit Framing on Self-Efficacy

The main effect of UNIT was not significant, $F(1, 391) = 1.62$, $p = .20$. However, I found a significant interaction between UNIT and DIST, $F(1, 391) = 10.09$, $p = .002$. This indicates that a difference in data unit had different effects on the self-efficacy score at different levels of distance to the goal. To break down this interaction, simple contrasts were performed comparing each level of UNIT to one another across different level of DIST. As Figure 24 shows, at the lower level of goal achievement (25%), the feedback showing raw data led to a higher self-efficacy score ($M = 4.37$, 95% CI[4.02, 4.72]) than the feedback showing percentage data ($M = 3.76$, 95% CI[3.41, 4.11]), $F(1, 391) = 5.72$, $p = .02$. However, at the higher level of goal achievement (75%), UNIT had no effect, $F(1, 391) = .38$, $p = .54$.

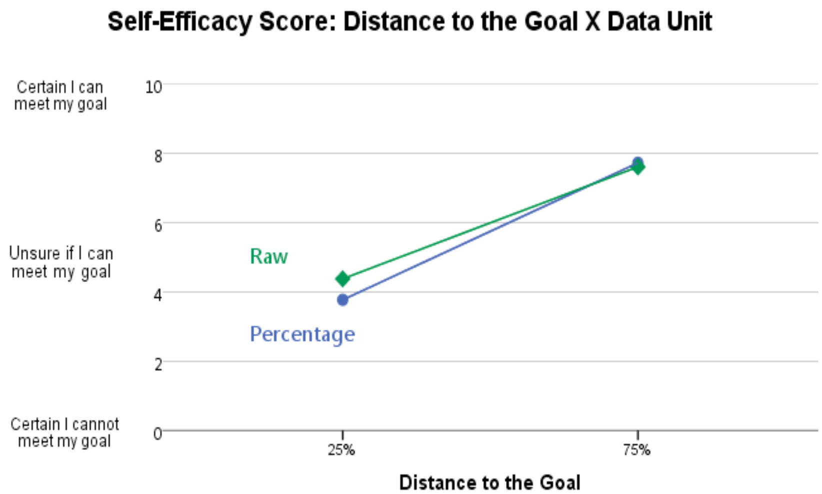


Figure 24. The interaction effect between data unit and self-efficacy score: participants' self-efficacy was higher at the 25% distance to the goal condition when they were shown feedback in a raw data format than in a percentage data format.

7.5. Discussion

The study results can guide the design of performance feedback. I identified several framings such as valence and presentation type that are better at enhancing people's self-efficacy than others, though some of these effect sizes were small.

I observed that use of the *achieved* framing could enhance self-efficacy throughout the various levels of goal achievement (25%–75%). This result aligns with previous attribute framing studies of health messages where a positive framing of an object or event usually leads to more favorable evaluations than a negative framing. Self-efficacy is concerned with a person's *beliefs* about one's capabilities of performing a behavior. Thus, I conclude that use of the achieved framing for performance feedback influences these perceptions—not necessarily the true capabilities of performing a task, but *feeling capable of performing a task* to achieve a goal. This result indicates that people are subject to framing not only when evaluating an external object but also when evaluating one's own capability. Validating the claim at more extreme cases toward the both ends of the goal (e.g., 5%, 95%) warrants future research.

I also observed that performance feedback conveyed using text and visuals (i.e., colored progress bar) did not enhance individuals' self-efficacy when compared to text-only feedback.

Ancker et al. noted that visuals that improve the accuracy of quantitative reasoning appear to be different from visuals that promote behavior change (Ancker et al., 2006). I suspect that the progress bar in the visual conditions supported the former rather than the latter, and thus enhancing people's quantitative understanding, rather than improving their self-efficacy. Now that this work has provided a better understanding of the role one type of visual plays in conveying performance feedback, the next step is to explore the influence of exaggerated, or even judgmental visuals (e.g., emoticons) on self-efficacy. Because judgmental visuals convey valence information more saliently, positively framed judgmental visuals, in particular, could lead to an improved self-efficacy.

To our surprise, data unit (i.e., raw versus percentage) mattered only at 25% (2,500 steps) of goal achievement but not at 75% (7,500 steps). I suspect that performance feedback shown in a raw data format (e.g., "2,500 steps" achieved) was perceived as a bigger achievement than the same information shown in a percentage format (e.g., "25%" achieved). However, this effect was not observed for the distance to the goal at 75% level. It appears that when performance achievement level approaches the goal, the perception gap resulting from data unit decreases.

The research reported in this paper has some limitations. As is often the case with a convenience sample, sampling bias could have affected our study results. Our sample is biased toward highly educated and technical people. Also, our participants seem to have high interest in physical activities, which is exemplified by their high interest level ($M = 7.14$) in taking 10,000 steps daily and previous high pedometer usage experience (43%). Previous research reports that Framing effects due to valence manipulation might not occur when the research topic has high intrinsic self-relevance to the research population (Krishnamurthy et al., 2001). The high-interest bias in the subject population makes it less likely that I would find a valence framing result, but nonetheless, I observed a significant main effect of valence framing. I note that 5% of participants ($n = 20$) indicated a self-efficacy scale of 10 (i.e., "Certain I can meet my goal") for both 25% and 75% distance to the goal conditions. For this group of people, framing would not matter much because their self-efficacy is so high they will achieve the goal regardless of feedback types. I included this data in the main analysis because this is valid data, and there will always be people with high self-efficacy. However, when I excluded this data and re-analyzed, we observed a more significant effect of valence framing, $F(1, 371) = 4.49$, $p = .035$. In any event, the high-interest bias merely supports our finding further.

Related contextual limitations of the research include using the hypothetical scenario and measuring self-efficacy rather than actual behavior. However, because I am still at the early stage of identifying persuasive framing for the design of effective performance feedback, I argue that conducting a field deployment study to measure behavioral outcomes is not the best first approach. Conducting an online experiment using a hypothetical scenario allowed me to recruit a large number of participants with relatively low cost and helped me understand the effect of different framings in a quick time frame. Design implications from this work will help designers and researchers create influential, persuasive performance feedback, which could be embedded in a self-monitoring technology for a long-term deployment study. Also, while I have focused on finding desirable performance feedback in the context of step count, it is possible to think of other performance feedback with different types of goals. For example, how framing is manifested differently depending on different types of goals in different contexts (e.g., "the higher the better" goal as in accumulated step counts, "the lower the better" goal as in smoking cessation, or "the ideal range" goal as in calorie intake) opens up many possibilities for future work.

7.6. Chapter 7 Summary

The objective in this research was to identify the type of framing that could best convey performance feedback to enhance individuals' self-efficacy. I accomplished this goal by conducting an online experiment with 400 participants who were given a hypothetical scenario of receiving a real-time performance feedback of daily step count. I found that valence and presentation type framings were highly related to individuals' self-efficacy. Specifically, an achieved framing led to higher self-efficacy than a remaining framing and a text-only framing led to higher self-efficacy than a text with visual (colored progress bar) framing. Furthermore, I found a significant interaction effect between data unit and distance to the goal, indicating data unit might influence people's perceived level of achievement especially at the early phase (25%) along the course of goal achievement. In designing performance feedback to enhance people's self-efficacy, I recommend using a positive framing with data unit that can increase the perception of one's performance capabilities. This work provides empirical guidance for creating influential, persuasive performance feedback, thereby helping people designing self-monitoring technologies to promote healthy behaviors.

Chapter 8
Contributions and Opportunities for Future Work

In this last chapter, I summarize previous chapters as well as state contributions and limitations. I end this dissertation with a discussion on opportunities for future work and a conclusion.

8.1. A Summary of Prior Chapters

In Chapter 1, I defined ***self-monitoring technology*** as "technology that facilitates capturing of the occurrences of target behavior and provide feedback to help people increase awareness and self-reflection." Self-monitoring technology must help people capture the behavior of interest and provide feedback to aid with self-reflection. In this light, I brought up two important challenges of designing self-monitoring technology—(1) lowering manual capture burden and (2) providing effective feedback. I explored this topic in the area of sleep monitoring because factors that are likely to affect sleep quality are plenty and hard to capture automatically.

In Chapter 2, I provided overviews of two theoretical backgrounds—(1) self-monitoring and (2) framing effects. In the first half of Chapter 2, I stated two purposes of "traditional" self-monitoring—***assessment*** function where data accuracy must be achieved and ***treatment*** function where reactive effects must be maximized. I argued that enhancing both data accuracy and reactive effects is essential when designing self-monitoring technology, which is often used for self-management purposes encompassing the assessment and treatment functions. To get insights from prior self-monitoring literature, I summarized factors that are known to affect data accuracy and reactive effects. In the last half of Chpater 2, I provided a summary of prospect theory, the core of which is the Framing effects. In particular, I explained that attribute framing—varying the valence of the attribute—could affect our evaluation of a target object or event characteristic. I argued that self-monitoring feedback should be designed in a way that nudges people toward positive health behaviors and that designers should apply the kinds of framing that could enhance people's self-efficacy.

In Chapter 3, I clarified the term, self-monitoring technology, and related it to personal informatics and Quantified Self. The decisive characteristic of self-monitoring technology is that it leverages individuals' voluntary, active involvement in data collection for enhancing self-reflection. I provided an overview of existing commercial products and HCI research projects that are in the realm of self-monitoring technology for health. For each product or project, I enumerated its target behavior, capture mechanism, and feedback mechanism. Interestingly, many of automated sensing tools (wearable sensing, embedded sensing) incorporated manual tracking to complement the limitations of automated sensing (e.g., to fix incorrect data, to capture the data that cannot be captured by automated sensing). In Chapter 3, I also introduced sleep hygiene recommendations, with which self-monitoring technology for sleep should comply.

In Chapter 4, I addressed the first research question—how do people currently practice self-monitoring. Understanding people's current self-monitoring practice helps designers learn challenges and opportunities in this area. I investigated this topic through the lens of Quantifies-Selfers. I conducted a qualitative and quantitative analysis of Quantified Self meetup videos. In the videos, experienced self-trackers uncovered their motivations to self-monitoring, tools they used to collect and explore data, insights they gained, common pitfalls, and workarounds to overcome the challenges. Based on these findings, I identified design opportunities for self-monitoring technology including exploring ways to provide early feedback, to support designing rigorous self-experimentation, to leverage the benefits of—while easing the burden of—manual tracking, and to promote self-reflection.

In Chapter 5, I addressed the second research question—what is the design space for sleep technologies. While I devoted Chapter 4 to identify design opportunities for self-monitoring technology, I dedicated Chapter 5 to identify opportunities for sleep-specific technology. To map the design space of sleep technology, I conducted a triangulated formative study, which included a literature review on existing sleep-related technologies, contextual interviews with domain experts, large-scale survey of peoples' attitudes toward sleep-related technologies, and interviews with a subset of the survey respondents. Based on the findings, I proposed sleep design framework, which consists of six dimensions: goal; feature; source; technology platform; stakeholders; and input mechanism. After mapping existing technology onto this design framework, I identified opportunities to design sleep technologies that have (1) in-home diagnosis and treatment goals, (2) persuasive and educational features, (3) a ubiquitous form factor, and (4) automatic input mechanisms. One salient feature of these technology ideas was the technology's ability to track sleep. However, privacy concerns were also raised due to the bedroom environment where sleep monitoring takes place.

In Chapter 6, I designed and evaluated SleepTight, a self-monitoring technology for collecting and reflecting on sleep behaviors. I had three design goals for SleepTight, which were drawn from prior work and literature review. The design goals were: (1) enable people to capture both target behaviors and triggers that are likely to influence the target behaviors; (2) lower the capture burden and create consistent capturing habit; and (3) provide feedback to help with self-reflection. I leveraged Android's lock screen and home screen widgets to support meeting these goals. In addition, SleepTight supported customizing the tracking items and prevented backfilling of sleep data. I conducted a between-subjects study comparing the Full System (lock screen, home screen, application) condition to App-only System condition to address the following four research questions: (1) How does an easily accessible manual self-monitoring application that provides visual feedback enhance tracking adherence and data accuracy?; (2) How does SleepTight's widget affect overall usage duration and access to information?; (3) How does SleepTight's ability to customize tracking items affect individuals' tracking practice?; (4) How does SleepTight affect individuals' awareness and self-reflection?; and (5) How does

SleepTight's ability to prevent backfilling affect individuals' tracking practice? The evaluation study suggested that the widgets served as visual reminders, which helped people collect more data, more accurately. People valued the SleepTight's customization features. They added custom items they thought were impacting sleep quality although some items were hard to track using the timestamp format. People were able to reflect upon their sleep behaviors and identified findings and hypotheses about their sleeping patterns, other activity patterns, and relationships among multiple factors. However, they showed mixed feelings about SleepTight's ability to prevent backfilling—although it did help people remember to log sleep, some people occasionally could not enter data even when they wanted to. Based on the design and evaluation of SleepTight, I provided several design implications for self-monitoring technology.

In Chapter 7, I further examined ways to provide self-monitoring feedback that could positively influence people. To achieve this goal, I leveraged framing effects, which I introduced in Chapter 2. I identified three types of framing that are applicable to present self-monitoring feedback. These three types include (1) valence of performance; (2) presentation type; and (3) data unit. Each of these framing types could be manipulated so that we can create pairs of feedback that convey the same information in different manner (e.g., varying the valence of performance by describing the performance using the achieved frame and the remaining frame). To address the research question of how these three types of performance feedback framing influence an individuals' self-efficacy, I conducted an online experiment. I found that valence and presentation type framings were highly related to individuals' self-efficacy. Specifically, an achieved framing led to higher self-efficacy than a remaining framing and a text-only framing led to higher self-efficacy than a text with visual (colored progress bar) framing. Furthermore, I found a significant interaction effect between data unit and distance to the goal, indicating data unit might influence people's perceived level of achievement especially at the early phase (25%) along the course of goal achievement.

8.2. A Summary of Contributions

In this thesis research, I made contributions to the interdisciplinary fields of HCI and Health Informatics. The contribution of this work can be summarized as follows: (1) design and implementation of self-monitoring technology that facilitates easy data capture and self-reflection; (2) empirical findings—such as design guidelines for self-monitoring technology design based on a deployment study, experimental study, survey, and interview; and (3) methodological guidelines for identifying persuasive performance feedback. In what follows, I elaborate on these contributions.

8.2.1. Design and Implementation

By designing and implementing the **SleepTight System**, I made an artifact contribution[20] to the field of HCI. I showed that a mobile, manual tracking application leveraging Android's lock screen and home screen widgets can facilitate easy data capture and self-reflection. Through a deployment study comparing the Full System (widgets and regular app) condition with the App-only System condition (regular app), I showed the efficacy of the **lock screen widget and home screen widget as effective data capture and reflection tool.** Those who used the Full System (app, lock screen and home screen widgets) captured more data, more accurately than those who used the App-only System, thereby proving the utility of the widgets. The widgets were suitable for self-monitoring technology because they were easily visible and accessible. The SleepTight system is one of the first systems to leverage lock screen and home screen for the purpose of self-monitoring technology. As many self-monitoring technologies rely on manual capture, other HCI researchers can leverage lock screen and home screen widgets to lower the capture burden as I have shown their efficacy through the SleepTight deployment study. In addition, lowering the capture burden has an immediate contribution to the field of health informatics as I have shown ways to enhance patients' self-report measures and reactivity.

[20] Artifacts in HCI are inventions, including new systems, tools, or design portrayals that reveal new opportunities or impel us to consider new possible opportunities. Artifact contributions are, by definition, dependent upon never-before-seen inventions that are instantiated as prototypes, sketches, mock-ups, demos, or other envisionments. See "Seven Research Contribution Types in Human-Computer Interaction (Wobbrock, n.d.).

8.2.2. Empirical Findings

Empirical findings based on systematic observation studies provide new data to reveal formerly unknown insights about human behavior and its relationship to technology (Wobbrock, n.d.). The studies could be qualitative (e.g., field study, interview) or quantitative (laboratory, experiment design). In my thesis work, I conducted various kinds of scientific studies to provide insights about designing self-monitoring technology. I elaborate on these findings.

Design guidelines for self-monitoring technology. Drawn from the SleepTight evaluation study results, I discussed several implications for self-monitoring technology design including the need to (1) capture both target behaviors and potential triggers; (2) lower the capture burden to encourage a consistent capturing habit; (3) provide feedback to help with self-reflection; (4) support customization; (5) project personal data in a positive light; and (6) identify and capture anomalies.

Common pitfalls of self-monitoring practice. Based on the video analysis of experienced self trackers, I identified the following three common pitfalls: (1) tracking too many things at the beginning, which leads to tracking fatigue or failure to do data analysis; (2) not tracking triggers and context, which leads to not having enough clues on how to improve outcome measures; (3) lacking scientific rigor when conducting self-experimentations. I also identified how the experienced self-trackers overcome these problems in Chapter 4, and some of their workarounds and recommendations were implemented in the design of SleepTight system.

A design framework that maps the design space of sleep technologies. Based on literature review and formative studies (survey, contextual inquiry, interviews), I proposed a design framework for sleep technologies, which consists of six dimensions: goal; feature; source; technology platform; stakeholders; and input mechanism. The design framework can be used for descriptive purposes (i.e., mapping a single sleep technology onto the framework to describe its design using the six dimensions), or for prescriptive purposes (i.e., mapping a class of existing sleep technologies onto the framework to identify missing holes and future design opportunities).

Design guidelines for creating influential, persuasive performance feedback. Performance feedback created based on self-monitoring data must be provided in a way to enhance individuals' self-efficacy, thereby helping people achieve their health goals and make health-enhancing choices. In Chapter 7, I provided empirical guidance for creating influential, persuasive performance feedback. When designing self-monitoring technologies to promote healthy behaviors, I recommend using a positive framing with data unit that can increase the perception of individuals' performance capabilities.

Some of these empirical findings directly influenced the design of SleepTight system while other empirical findings were lessons learned from the SleepTight deployment study. I encourage other HCI researchers and health informatics researchers to incorporate these empirical findings in the design of self-monitoring technology to design effective self-monitoring technology.

8.2.3. Methodological Guidelines

I made methodological research contributions to identify and create persuasive performance feedback. I designed and evaluated several persuasive performance feedback with an aim to identify the types of framing that can nudge people toward positive behaviors for health (Choe et al., 2013b). My prior work also applied the same concept to identify the types of framing that can lead people to make privacy preserving choices (Choe et al., 2013b). From these experiences, I propose the following methodological guidelines to design and identify the types of feedback that can be used to nudge people toward a certain behaviors:

(1) **Identify positive & negative valence framing**—examples include achieved steps vs. remaining steps, privacy-preserving score vs. privacy-invasive score.

(2) **Find visual representations that can convey the valence information**—leverage culturally prevalent visual attributes (e.g., colors) and semantics (e.g., valence, signs, symbols). These are highly culturally dependent, so when possible, leverage people's pre-conceptions built on their life experiences.

(3) **Conduct fast, iterative tests to identify more effective visual framing**—know what dependent variables to measure and conduct between-subjects studies to identify more effective visual framing.

(4) **Ethical considerations should precede deciding when and where to apply visual framing**— visual framing must be applied to enhance positive behaviors because people are unknowingly influenced by it.

HCI researchers and health communication researchers can use the methodological guidelines to create persuasive performance feedback to carry out their work in HCI and health informatics.

8.3. Limitations

In this section, I describe limitations from the studies I conducted. Each study method has its own limitations, and thus, researchers should choose a method (or mixed-methods approach) after understanding the tradeoffs of available options.

In Chapter 5, I used the video analysis method to understand Quantified Selfers' self-monitoring practices. The video recordings were publically available and the contents were well structured. Although analyzing publically available data saved me time for data collection, I could not ask follow-up questions on *how* they gained insights; I was bounded by how information was structured and presented by the speakers. Looking forward, this research has motivated me to reach out to the speakers and do a follow-up study of how Q-Selfers gain insights and how they construct visualizations to present the insights.

In Chapter 6, I described the SleepTight system and study design. The SleepTight system design was limited by the deliberate attempt to control for confounding factors. For example, although many studies suggested that goal setting is a powerful means to encourage behavior change (Locke & Latham, 2002), I purposefully did not employ the goal setting in the SleepTight system design because separating the effect of goal setting from the other features would have been difficult. Due to the same reason, I did not explore the role of time-based reminder (or notification) because time-based reminder has already been studied and proven to be effective (Bentley et al., 2013). Instead, I implemented the 24-hour time limit and visual reminders (i.e., widgets) and showed that they are effective in helping people collect data on time. Combining these different types of reminders in one system would have resulted in higher tracking adherence, but it would also have complicated the interpretation of the study results. The SleepTight system design was also limited by the decision to lower the user burden and to simplify steps to capture a factor, thereby offering two simple ways to capture—(1) duration capture; and (2) frequency capture. However, intensity or severity capture was not supported by SleepTight, so people could not capture factors such as pain or stress level where tracking the factor's intensity or severity was more important than its start and end time. Lastly, the SleepTight deployment study design could have been improved by lengthening the study period, thereby measuring not only the tracking adherence, but also the long-term change in sleep behavior.

In Chapter 7, I conducted an experimental study with 400 survey respondents. However, sampling bias could have affected the survey results. Related contextual limitations included using the hypothetical scenario and measuring self-efficacy rather than measuring actual behavior. With the presence of these limitations, however, the online experiment using a hypothetical scenario allowed us to recruit a large participants in a quick time frame. Also, this type of study could be a great first step to compare competing designs before selecting one design for a deployment study.

8.4. Opportunities for Future Work

My future research agenda is directed toward the long-term goal of supporting personalized and preventive medicine through health information technology. Although I addressed several challenges of designing self-monitoring technology in this dissertation research, I also opened up many areas of research that need to be examined further. In the following, I provide opportunities for future work.

- **Understanding the meaning of self-reflection and insight:** We frequently use the terms "self-reflection" and "insight" in self-monitoring research and personal informatics, but there are no formal definitions of these terms. We need to understand what people mean by "self-reflection" and "insight" so that tools can properly support the self-reflection process and gain insights. Chapter 5 touched upon what insights people gained from self-monitoring. To learn the process of gaining insights and the role of visualizations in aiding the process, we can interview Quantified-Selfers and analyze the data using and building upon the existing taxonomy of insights (Yi et al., 2008; Yang et al., 2014).

- **Understanding why people stop tracking personal data:** Although many people stop using self-monitoring technology because they face serious challenges, few studies have looked at why people suspend the continued use of self-monitoring technology. As studying the experienced user group such as Quantified Self gave us numerous insights, studying the other end of the extreme—those who quit using self-monitoring technology—would also help us understand the reasons for quitting and their unmet needs.

- **Providing personalized goals:** There is much room to explore how different kinds of goals affect the actual outcome of behaviors. In particular, how to design effective personalized goals raises opportunities for future research. For example, in the context of sleep, sleep doctors prescribe to-bed time and wake up time. There are also general suggestions on cutoff times for caffeine consumption, exercise, and tobacco. However, these general suggestions (goals) might not be effective when they are too challenging or not applicable to individuals' lifestyle. Therefore, how to create personalized goals that are challenging enough, relevant, and effective warrants future research efforts.

- **Studying the long-term effect of SleepTight on target behavior outcomes:** To show the effect of self-monitoring technology (e.g., SleepTight) on long-term behavior change, we need to conduct a longitudinal study that lasts for several months to years. Looking at both long-term tracking adherence and behavioral outcomes through a longitudinal study would help us understand the relationship among tracking, habit formation, behavior change, and maintenance.

- **Comparing the effects of different reminders:** Reminders are powerful means to enhance tracking adherence and self-reflection, but they have their own limitations (e.g., people might ignore too many reminders). Therefore, identifying pros and cons of different types of reminders—such as time-based notification, 24-hour time limit, and visual reminder, and comparing the efficacy of different types of reminders would help us understand ways to effectively use them and incorporate them into self-monitoring technology design.

- **Personalized approach to behavior change:** I suspect that people have different tolerance level toward self-monitoring, which I learned during the SleepTight evaluation study. Although I deployed the same self-monitoring technology, one group of people said that it was so easy to track data while other group of people said that it was too much work. Identifying what individual characteristic results in this kind of different reactions toward self-monitoring (or other persuasive techniques) would help designers employ personalized approach to behavior change.

- **Exploring the design space of semi-automatic self-monitoring technology:** Although the use of sensors and computer automation have many advantages in collecting personal health data in terms of reducing mental workload and increasing data accuracy, this automated data collection process could reduce awareness resulting from people's engaging with data collection. However, manual, self-reported data has the advantage of increased awareness and self-reflection. Finding the most effective balance between fully automated sensing and manual self-report that can increase awareness, achieve reasonable accuracy, and decrease mental workload is an important research agenda in various domains, such as food tracking, sleep monitoring, and activity tracking.

- **Motivating people who are not yet ready to change their behavior or who are experiencing setbacks:** Many existing health technologies are mostly successful at sustaining behavior change for people who are already motivated, but not for people who are not yet motivated to make changes. How can we create emotional, epiphany moments that can spark initial change for those who are not yet motivated to make changes? Designing health technology and feedback that can support this particular group to enhance and maintain their motivation is a challenging research problem.

- **Supporting patient-led information capture and management:** Patients are careful observers of their own care and are at the center of the treatment process. Inpatients have high motivation to be actively involved in the care process and are able to identify errors and undesirable events. By providing a method to capture and access the information patients need, we can encourage patients and caregivers to contribute to prevent medical errors and undesirable events, and to become proactive in seeking the best care.

- **Enabling the sharing of personal data between patients and clinicians:** Self-monitoring data captured by patients outside the clinic is valuable piece of information, which clinicians can use for better diagnosis. However, currently, there is no good way to share the information between patients and clinicians. I want to enable the sharing of patient-captured data between patients and clinicians without overwhelming the clinicians' workload.

- **Designing a platform for self-experimentation and personal data visualization:** With the enormous expansion of consumer health management tools, people collect large amounts of personal data but lack sufficient tools to visualize, explore, and understand this data. We need a tool to help people articulate hypotheses, configure data capture settings, and find meaningful insights from personal data they collect. This kind of tool could help people gain individualized insights.

8.5. Concluding Remarks

As I conclude my dissertation research, one of the most important findings I learned—although not too surprising—is that people like to get credit for what they have done. Thus, we should leverage this kind of motivation when designing health information technology. Sometimes, self-monitoring is a way of getting credit, when people conduct positive, healthy, encouraging behaviors. Other times, it is a way of facing a hard truth when they conduct negative, unhealthy, discouraging behaviors. The challenge is to keep people engaged even during the bad times as well as good times, while not compromising the integrity of the tracking data.

As exemplified in this dissertation research, easing the capture burden and designing feedback are closely intertwined in the self-monitoring tool design. Easing the capture burden could enhance the quality of feedback in return, creating a desirable feedback loop of increased awareness, self-reflection, and continued tracking (Choe et al., 2014b). Thus, self-monitoring technology designers should consider ways to support the whole spectrum of self-monitoring process—including configuration, data capture, and feedback phases.

Appendix A
Study Material for SleepTight Deployment Study

Appendix A-1. Factors Questionnaire (Pre/Post).

Participant ID

SleepTight Study
Attachment C-1: Factors Questionnaire
University of Washington

QUESTIONNAIRE 1 [PRE/POST]

1. **Based on your experience,** please list any plausible factors that you think might influence your sleep. Then indicate whether the factor influenced your sleep **either positively or negatively**. Try to be specific in describing each factor (e.g., "drinking a cup of water before going to bed" rather than "drinking water"). Also, indicate the degree of confidence you have in the effect of each factor on your sleep.

Description of the factor	Positively or Negatively?	Level of Confidence				
		Not at all confident	Slightly confident	Somewhat confident	Moderately confident	Extremely confident
	Positively ○ Negatively ○					
	Positively ○ Negatively ○					
	Positively ○ Negatively ○					
	Positively ○ Negatively ○					
	Positively ○ Negatively ○					
	Positively ○ Negatively ○					

Appendix A-2. Demographic Questionnaire.

QUESTIONNAIRE 2 [PRE]

1. Demographic Information

1. What is your **gender**?
   - ○ Male
   - ○ Female
   - ○ Other:____________

2. What is your **age**? _________ years old

3. Please select the **highest level of education** that you have completed. (check one)
   - ☐ High school diploma
   - ☐ Some college
   - ☐ Certificate (in what? __________________)
   - ☐ Bachelor's degree
   - ☐ Some graduate work at Master's level
   - ☐ Master's degree
   - ☐ Some graduate work at Doctoral level
   - ☐ Ph.D.
   - ☐ M.D.
   - ☐ Other professional degree (in what?_________________________)
   - ☐ Other: (_________________________)

4. Which best describes your **current employment status**? (check all that apply)
   - ○ Self-employed
   - ○ Employed full-time
   - ○ Employed part-time
   - ○ Homemaker
   - ○ Retired
   - ○ Student – Undergraduate
   - ○ Student – Master's
   - ○ Student – Doctoral
   - ○ Unemployed
   - ○ Other: (_________________________)

5. If you are employed, what is your **current occupation / job title**? (e.g., Registered Dietician, Administrative Assistant, Homemaker, etc.)

   ____

6. **With whom** do you live? (check all that apply)
   - ☐ I live alone
   - ☐ Spouse / Partner
   - ☐ Child(ren) (how many? _____ what are their ages?________)
   - ☐ Parent(s) (how many? _____)
   - ☐ Sibling(s) (how many? _____)
   - ☐ Roommate(s) (how many? _____)
   - ☐ Housemate(s) (how many? _____)
   - ☐ Other relative(s) (how many? _____)
   - ☐ Friend(s) (how many? _____)
   - ☐ Other(s) (please explain ________________ & how many? _____)

7. Do you have any **pets**? (check all that apply)
   - ☐ No
   - ☐ Yes - cat(s)
   - ☐ Yes – dog(s)
   - ☐ Yes – Other (what kind of pet(s)? ________________________)

2. Sleep Habits

1. Which of the following do you have? If you have the item, how often, if at all, do you use it?

	Do you have?		**If you have the item, how often, if at all, do you use?**				
Sleep-related app (if so, what kind?) ____________________	[no]	[yes]	[never]	[rarely]	[occasionally]	[frequently]	[always]
Alarm clock	[no]	[yes]	[never]	[rarely]	[occasionally]	[frequently]	[always]
Relaxation/biofeedback device/application (e.g., emWave, StressEraser)	[no]	[yes]	[never]	[rarely]	[occasionally]	[frequently]	[always]
Sleep sensing device:							
Zeo	[no]	[yes]	[never]	[rarely]	[occasionally]	[frequently]	[always]
Fitbit	[no]	[yes]	[never]	[rarely]	[occasionally]	[frequently]	[always]
Jawbone Up	[no]	[yes]	[never]	[rarely]	[occasionally]	[frequently]	[always]
Other:____________	[no]	[yes]	[never]	[rarely]	[occasionally]	[frequently]	[always]
White noise or similar sound machine	[no]	[yes]	[never]	[rarely]	[occasionally]	[frequently]	[always]

2. ′What, if any, other sleep-related items do you use in your bedroom at least on occasion which are not listed above?

   ____

3. If you have a smartphone or cellphone, approximately **how many hours** per day do you use them?
   - ○ Less than 30 minutes
   - ○ 30-60 minutes
   - ○ 1-2 hours
   - ○ 2-3 hours
   - ○ More than 3 hours

4. If you have a smartphone or cellphone, approximately **how many times** per day do you turn it on to access any of the features?
   - ○ Less than 5 times
   - ○ 5 to 10 times
   - ○ 10 to 20 times
   - ○ More than 20 times

5. If you have a smartphone or cellphone, **when** do you typically use the phone for the first time throughout the day?
   - ○ I don't have a pattern
   - ○ Right after I wake up (e.g., I use alarm clock features on cell phone)
   - ○ In the morning (before noon)
   - ○ In the afternoon (noon-5pm)
   - ○ In the evening (after 5 pm)

6. If you have a smartphone or cellphone, **when** do you typically use the phone for the last time throughout the day?
   - ○ I don't have a pattern
   - ○ Right before I go to bed
   - ○ In the evening (after 5 pm)
   - ○ In the afternoon (noon-5pm)
   - ○ In the morning (before noon)

7. What activities do you typically do before an hour of going to bed?

   ____

8. I have commitments (e.g., job, school, kid, pet, etc.) that may affect having a regular sleep schedule.
   ○ Yes
   (please describe:____)
   ○ No

9. During the past months, have you taken any action to learn about your sleep? (e.g., use sleep tracking devices such as Fitbit, Zeo, sleep diary, etc.)
   ○ Yes
   (please describe:____)
   ○ No

10. Do you have a particular goal regarding sleep?
    ○ No
    ○ Yes (Please describe in detail)
    Wake up time ____
    To-bed time ____
    Consistent sleep cycle ____
    Reduce caffeinated drinks ____
    More sleep / less sleep____
    Others-Please specify____

11. How motivated are you to learn about your sleep patterns right now? Rate your motivation on a scale of 1-10, where '1' is not at all motivated and '10' is very motivated.

[1]									[10]

Not at all — Neutral — Very

12. How confident are you about your current knowledge of your sleep patterns? Rate your confidence on a scale of 1-10, where '1' is not at all confident and '10' is very confident.

[1]									[10]

Not at all — Neutral — Very

13. In regard to my sleep,
    - ☐ I have no intention to take action in the next 6 months
    - ☐ I have intention to take action within the next 6 months
    - ☐ I have intention to take action within the next 30 days and some initial steps towards that action
    - ☐ Change in behavior occurred at some point in the past, but have not developed into habit
    - ☐ Overt behavior will never return, and there is complete confidence in coping without tear of relapse

    If you have taken any action or considered taking any action in the future, please explain in detail:
    ____
    ____
    ____

You have reached the end of the survey.
Please return completed questionnaire to the researcher.
Thank you!

Appendix A-3. Confidence Level Questionnaire.

QUESTIONNAIRE 3 [PRE/POST]

1. **Based on your experience,** indicate whether the following factors influence your sleep either positively or negatively. You may check “No effect” if you suspect that the factor does not influence your sleep. Then indicate the degree of confidence you have in the effect of each factor on your sleep. If a factor is not applicable, you may check “Not applicable” and skip the rest of the questions.

Description of the factor	Not applicable to me	Positively or Negatively?	Level of Confidence				
			Not at all confident	Slightly confident	Somewhat confident	Moderately confident	Extremely confident
Taking a nap (duration less than 1 hour)	☐	Positively ○ Negatively ○ No effect ○					
Taking a nap (duration more than 1 hour)	☐	Positively ○ Negatively ○ No effect ○					
Drinking caffeinated beverages before ___ am/pm	☐	Positively ○ Negatively ○ No effect ○					
Drinking caffeinated beverages after ___ am/pm	☐	Positively ○ Negatively ○ No effect ○					
Drinking alcoholic beverages before ___ am/pm	☐	Positively ○ Negatively ○ No effect ○					
Drinking alcoholic beverages after ___ am/pm	☐	Positively ○ Negatively ○ No effect ○					
Using tobacco products before ___ am/pm	☐	Positively ○ Negatively ○ No effect ○					

Appendix A-4. PSQI Questionnaire.

Subject's Initials________________ID#__________________Date_______________Time_____AM/PM

PITTSBURGH SLEEP QUALITY INDEX

INSTRUCTIONS:
The following questions relate to your usual sleep habits during the past month only. Your answers should indicate the most accurate reply for the majority of days and nights in the past month. Please answer all questions.

1. During the past month, what time have you usually gone to bed at night?

   BED TIME ___________

2. During the past month, how long (in minutes) has it usually taken you to fall asleep each night?

   NUMBER OF MINUTES ___________

3. During the past month, what time have you usually gotten up in the morning?

   GETTING UP TIME ___________

4. During the past month, how many hours of actual sleep did you get at night? (This may be different than the number of hours you spent in bed.)

   HOURS OF SLEEP PER NIGHT ___________

For each of the remaining questions, check the one best response. Please answer all questions.

5. During the past month, how often have you had trouble sleeping because you . . .

a) Cannot get to sleep within 30 minutes

Not during the past month_____	Less than once a week_____	Once or twice a week_____	Three or more times a week_____

b) Wake up in the middle of the night or early morning

Not during the past month_____	Less than once a week_____	Once or twice a week_____	Three or more times a week_____

c) Have to get up to use the bathroom

Not during the past month_____	Less than once a week_____	Once or twice a week_____	Three or more times a week_____

d) Cannot breathe comfortably

Not during the past month_____	Less than once a week_____	Once or twice a week_____	Three or more times a week_____

e) Cough or snore loudly

Not during the past month_____	Less than once a week_____	Once or twice a week_____	Three or more times a week_____

f) Feel too cold

Not during the past month_____	Less than once a week_____	Once or twice a week_____	Three or more times a week_____

g) Feel too hot

Not during the past month_____	Less than once a week_____	Once or twice a week_____	Three or more times a week_____

h) Had bad dreams

Not during the past month_____	Less than once a week_____	Once or twice a week_____	Three or more times a week_____

i) Have pain

Not during the past month_____	Less than once a week_____	Once or twice a week_____	Three or more times a week_____

j) Other reason(s), please describe____

____

How often during the past month have you had trouble sleeping because of this?

Not during the past month_____	Less than once a week_____	Once or twice a week_____	Three or more times a week_____

6. During the past month, how would you rate your sleep quality overall?

Very good ___________

Fairly good ___________

Fairly bad ___________

Very bad ___________

7. During the past month, how often have you taken medicine to help you sleep (prescribed or "over the counter")?

Not during the past month_____	Less than once a week_____	Once or twice a week_____	Three or more times a week_____

8. During the past month, how often have you had trouble staying awake while driving, eating meals, or engaging in social activity?

Not during the past month_____	Less than once a week_____	Once or twice a week_____	Three or more times a week_____

9. During the past month, how much of a problem has it been for you to keep up enough enthusiasm to get things done?

No problem at all	__________
Only a very slight problem	__________
Somewhat of a problem	__________
A very big problem	__________

10. Do you have a bed partner or room mate?

No bed partner or room mate	__________
Partner/room mate in other room	__________
Partner in same room, but not same bed	__________
Partner in same bed	__________

If you have a room mate or bed partner, ask him/her how often in the past month you have had . . .

a) Loud snoring

Not during the past month_____	Less than once a week_____	Once or twice a week_____	Three or more times a week_____

b) Long pauses between breaths while asleep

Not during the past month_____	Less than once a week_____	Once or twice a week_____	Three or more times a week_____

c) Legs twitching or jerking while you sleep

Not during the past month_____	Less than once a week_____	Once or twice a week_____	Three or more times a week_____

d) Episodes of disorientation or confusion during sleep

Not during the past month_____	Less than once a week_____	Once or twice a week_____	Three or more times a week_____

e) Other restlessness while you sleep; please describe__________________________________

____

Not during the past month_____	Less than once a week_____	Once or twice a week_____	Three or more times a week_____

Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ: Psychiatry Research, 28:193-213, 1989.

Appendix A-5. Pre-study Interview Guideline.

Initial Interview:

Hello and thank you for joining us today. We are conducting a study to learn how the mobile sleep application can help people identify what factors might be impacting their sleep. To better understand this process we would like to ask you a few questions about your current sleep quality and your sleep related habits. Would it be ok to begin recording now?

Just to confirm, it's okay that I audio-record our conversation, correct? *[Wait for response]*

You may refuse to answer any questions if you feel uncomfortable. You may stop this interview at any time.

[Sleep quality]

Do you sleep well?

- When was the last time that you remember getting a good night's sleep?
- Can you describe what you mean by "sleeping well" ?
- Can you describe your typical sleeping patterns? Weekend vs. weekdays, wake-up time, to-bed time, whether it's consistent, etc?
- Do you typically wake up during the sleep?
- Do you get out of bed right away right after you wake up? Or do you stay in bed?

During recruiting, you mentioned that you wanted to learn about your sleep habits and sleep related factors. Based on your experience,

- What do you feel prevent you from getting a good night's sleep? Could you use a specific episode to support your case?
- [Probing about various sleep disturbances] pets, kids, noise, light, temperature, full bladder, sleep partner, worry, nightmare, night sweat, caffeine, no reason
- What do you think helps you get a good night's sleep? Could you use a specific episode to support your case?

Do you have any sleep ritual?

What activities do you typically do within an hour of going to bed?

- [Probing about various activities]
- Exercise [What kind?]
- Reading, working, emailing, homework?
- Showering, washing, brushing, etc.

- What electronics, if any, do you use close to sleep time?
  - Use of computing device [What kind?]
  - TV use in bedroom close to sleep?
  - Laptop/smart phone use in bedroom close to sleep?
  - Lights used for in bedroom leading up to bedtime?

Do you use any sleep medication?

Is there anything about your sleep that you'd like to improve?

What, if anything, have you tried to improve your sleep in the past?

- [Probe on Journaling, what factors would you track etc.]

[Sleep related factors]

- Do you drink caffeinated beverage?
  [If yes,]
    - How often? What kind? When do you usually drink them?
    - How does it affect your sleep?
    - Do you think about sleep when you drink them?
    - Is there a certain time threshold that you don’t tend to drink them after that time because you suspect that it affects your sleep?
    - Is there a certain amount threshold that you don’t tend to drink more than that because you suspect that if affects your sleep?
- Do you drink alcoholic beverage?
  [If yes,]
    - How often? What kind? When do you usually drink them?
    - How does it affect your sleep?
    - Do you think about sleep when you drink them?
    - Is there a certain time threshold that you don’t tend to drink them after that time because you suspect that it affects your sleep?
    - Is there a certain amount threshold that you don’t tend to drink more than that because you suspect that if affects your sleep?
- Do you smoke?
  [If yes,]
    - How often? When do you usually smoke?
    - How does it affect your sleep?
    - Do you think about sleep when you smoke?
    - Is there a certain time threshold that you don’t tend to smoke after that time because you suspect that it affects your sleep?
    - Is there a certain amount threshold that you don’t tend to smoke more than that because you suspect that if affects your sleep?

[Sleep environment]

Can you describe your sleep environment such as the bed, the window is facing, door, neighbor, or other noise / light source, blinds and curtains, whether you leave the door open / pet situation?

- Can you describe your “ideal sleep environment? In terms of light, noise, temperature, bedding…etc?
- What sleep aids, if any, do you use?
    - For example, sleep/eye mask, ear plugs
    - Do you use white noise machine (or similar) while sleeping? (or falling asleep)
- Do you have others in bed? (sleep partner? kids? pets?) How do they influence your sleep? Positively, negatively?

[Tracking experience]

- Have you used any tracking tool (computing device, pen & paper) to track anything?
  [If yes,]
    - What motivated you to start tracking?
    - What did you track?

- What tools/applications did you use?
- For how long did you track?
- What did you learn?
- What were some barriers/enablers for continuous tracking?
- Did the tracking lead you to change any aspect of your behavior?

  [if stopped for some reason]

    - Why did you stop tracking?

Is there anything else you'd like to tell me that we have not already covered?
Thanks!

[Turn off the recorder]

Appendix A-6. Post-study Interview Guideline.

Final Interview [Intervention/Control]:

Hello and thank you for joining us today. We are going to be conducting a follow-up interview with you today. Would it be ok to begin recording now?

Just to confirm, it's okay that I audio-record our conversation, correct? *[Wait for response]*

You may refuse to answer any questions that you are uncomfortable answering. You may stop this interview at any time.

- How was your experience with SleepTight for the past four weeks?
  - Can you describe a typical usage patterns?
  - When did you fill out the diary/daytime activities?
  - How often did you fill out the diary/daytime activities?
  - How did you open the app?
- How typical have the past four weeks been?
  - [Probes: any illness, injuries, vacations, business trips, visitors, change to your typical schedule etc]
  - When did these occur?
  - How, if any, did they affect your sleep?
  - How, if any, did they affect your usage of the app?

What did you learn from using SleepTight?

  - Were you able to find any patterns among various factors?
  - How did SleepTight help you identify any factors that had influenced your sleep?
  - Was there any particular information that was helpful? Why?
  - Was there any particular information that was not very helpful? Why?
  - What do you want to learn more from collecting these data?
  - Did you trust the information presented by SleepTight?
  - Was there anything about the data that made you feel uncomfortable?

- What have you liked about your experience over the past four weeks?
- What have you disliked about your experience over the past four weeks?
- What else do you think SleepTight should keep track of?

What do you want to learn more from collecting these data?

- Were there any changes in your belief about what's affecting your sleep?
  - Were you able to identify any new factors that are influencing your sleep? What are those? How did you learn that?
  - Were you able to confirm any factors that you believed to influence your sleep? What are those? How did you learn that?
  - Were you able to disprove any factors that you believed to influence your sleep? What are those? How did you learn that?
- How, if at all, did SleepTight change your behavior?
  - For example, adjusting your nighttime routine, coffee drinking habits
  - In terms of your sleep patterns?

- In what situations do you think SleepTight might help you?
- How could SleepTight be improved?
- Did you talk about the study with anyone?

[Widget condition]

How did you use the widget?

- As an in-situ data capture tool?
- Information display?
- A gateway to the app?

What did you like/dislike about the widget?

Is there anything about the widget that you'd like to improve?

[Privacy]

How do you feel about displaying your sleep data on the lockscreen / homescreen widget?

- Where do you think the data you collected with SleepTight should be stored?
  - On the device itself?
  - On a remote computer server
  - On a password-protected website?

(Sleep Partner) Has your sleep partner been an issue for your sleep? Have you talked about your sleep with your sleep partner?

[Time window for entering sleep data]

How do you feel about having the time window (till midnight) to fill out the sleep diary?

What could help you create a habit to enter and review your data everyday?

[Burden vs. Granularity of the data]

- What did you think about the granularity of the information?
- Which one did you like better - Frequency capturing vs. Duration capturing? Why?
- Were there any cases when you needed annotation feature? If so, can you explain?

Is there anything else you'd like to tell me that we have not already covered?

Thanks!

Appendix B
Overall Usage of SleepTight

Overall Usage of SleepTight for participants in the Full System (FS) condition and App-only System (AS) condition. Full System condition showed higher diary adherence rate than App-only System condition.

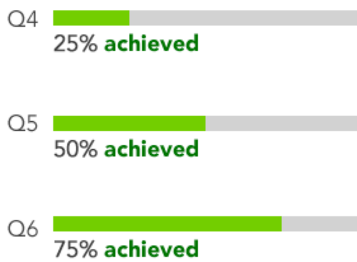
Full System Condition ID	Diary Count (A) (Max = 28)	Activity Count (B)	Category Count (C) (Max = 8)	Activity per Category (B)/(C)	App-only Condition ID	Diary Count (A) (Max = 28)	Activity Count (B)	Category Count (C) (Max = 8)	Activity per Category (B)/(C)
FS-2	28	119	3	39.7	AS-1	27	95	6	15.8
FS-3	28	176	6	29.3	AS-2‡	27	230	8	28.8
FS-8	28	225	8	28.1	AS-4∞	24	76	6	12.7
FS-9	28	272	5	54.4	AS-5‡	24	129	7	18.4
FS-1	27	147	6	24.5	AS-9‡	24	192	8	24.0
FS-10	26	137	8	17.1	AS-6‡	23	128	6	21.3
FS-11	25	112	5	22.4	AS-8‡	23	266	8	33.3
FS-5	22	31	5	6.2	AS-10‡	23	247	6	41.2
FS-6	21	150	8	18.8	AS-11∞	21	137	8	17.1
					AS-7	15	129	7	18.4
					FS-4*‡	11	39	5	7.8
					AS-3∞	3	30	6	5.0
Average (Adherence %)	25.89 (92%)	152.11	6.00	26.72	Average (Adherence %)	20.42 (73%)	141.50	6.75	20.32

* Initially assigned to the Full System condition, but removed the widgets so reassigned to the App-only System condition for the data analysis purpose
‡ Created a shortcut icon on the home screen
∞ Set a diary reminder

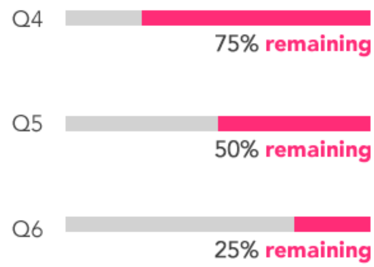
Appendix C
Persuasive Performance Framing Study Material

I created 8 versions of survey. Each survey showed one of the following eight feedback interventions:

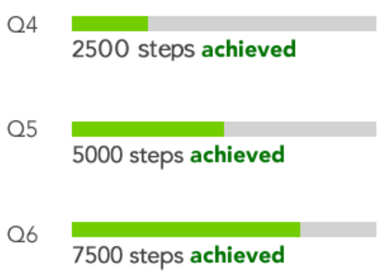
[Visual, positive, percentage] condition



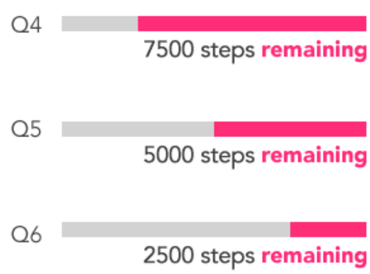
[Visual, negative, percentage] condition



[Visual, positive, raw data] condition



[Visual, negative, raw data] condition



[Text, positive, percentage] condition

Q4	25% achieved
Q5	50% achieved
Q6	75% achieved

[Text, negative, percentage] condition

Q4	25% remaining
Q5	50% remaining
Q6	75% remaining

[Text, positive, raw data] condition

Q4	2500 steps achieved
Q5	5000 steps achieved
Q6	7500 steps achieved

[Text, negative, raw data] condition

Q4	2500 steps remaining
Q5	5000 steps remaining
Q6	7500 steps remaining

References

Ancker, J. S., Senathirajah, Y., Kukafka, R., & Starren, J. B. (2006). Design Features of Graphs in Health Risk Communication: A Systematic Review. *Journal of the American Medical Informatics Association*, *13*(6), 608–618. doi:10.1197/jamia.M2115.

Ancoli-Israel, S., Cole, R., Alessi, C., Chambers, M., Moorcroft, W., & Pollack, C. (2003). The role of actigraphy in the study of sleep and circadian rhythms. *Sleep*, *26*(3), 342–392.

Anderson, I., Maitland, J., Sherwood, S., Barkhuus, L., Chalmers, M., Hall, M., Brown, B., & Muller, H. L. (2007). Shakra: Tracking and Sharing Daily Activity Levels with Unaugmented Mobile Phones. *Monet 12(2-3)*, 185–199. doi:10.1007/s11036-007-0011-7.

Aria. (n.d.). *Aria*. Retrieved May 25, 2014, from http://www.fitbit.com/aria.

Ayas, N. T., White, D. P., Manson, J. E., Stampfer, M. J., Speizer, F. E., Malhotra, A., & Hu, F. B. (2003). A prospective study of sleep duration and coronary heart disease in women. *Archives of Internal Medicine*, *163*(2), 205–209.

Azrin, N. H., & Powell, J. (1968). Behavioral engineering: The reduction of smoking behavior by a conditioning apparatus and procedure. *Journal of Applied Behavior Analysis*, *1*, 193–200.

Bandura, A. (1990). Perceived self-efficacy in the exercise of control over AIDS infection. *Evaluation and Program Planning*, *13*(1), 9–17. doi:10.1016/0149-7189(90)90004-G.

Bandura, A. (2006). Self-efficacy Beliefs of Adolescents - Google Books. *Self-Efficacy Beliefs of Adolescents*.

Bardram, J. E., Frost, M., Szántó, K., Faurholt-Jepsen, M., Vinberg, M., & Kessing, L. V. (2013). Designing mobile health technology for bipolar disorder: a field trial of the monarca system. Presented at the *CHI '13: Proceedings of the SIGCHI Conference on Human Factors in Computing*

Systems, ACM. doi:10.1145/2470654.2481364.

Bassett, D. R., Jr., Wyatt, H. R., Thompson, H., Peters, J. C., & Hill, J. O. (2010). Pedometer-Measured Physical Activity and Health Behaviors in U.S. Adults. *Medicine & Science in Sports & Exercise*, *42*(10), 1819–1825. doi:10.1249/MSS.0b013e3181dc2e54.

Bauer, J. S., Consolvo, S., Greenstein, B., Schooler, J., Wu, E., Watson, N. F., & Kientz, J. (2012). ShutEye: encouraging awareness of healthy sleep recommendations with a mobile, peripheral display. Presented at the *CHI '12: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM. doi:10.1145/2207676.2208600.

Bentley, F., & Tollmar, K. (2013). The power of mobile notifications to increase wellbeing logging behavior. Presented at the *CHI '13: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM. doi:10.1145/2470654.2466140.

Bentley, F., Tollmar, K., Stephenson, P., Levy, L., Jones, B., Robertson, S., Price, E., Catrambone, R., & Wilson, J. (2013). Health Mashups. *ACM Transactions on Computer-Human Interaction*, *20*(5), 1–27. doi:10.1145/2503823.

Beyer, H., & Holtzblatt, K. (1997). Contextual Design: Defining Customer-Centered Systems - Hugh Beyer.

BioBrite. http://www.biobrite.com.

Bonnet, M. H., & Arand, D. L. (1995). We are chronically sleep deprived. *Sleep*, *18*(10), 908–911.

Broden, M., Hall, R. V., & Mitts, B. (1971). The effect of self-recording on the classroom behavior of two eighth-grade students. *Journal of Applied Behavior Analysis*, *4*(3), 191–199.

Buysse, D. J., Reynolds, C. F., III, Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research. *Psychiatry Research*, *28*(2), 193–213. doi:10.1016/0165-1781(89)90047-4.

Carney C. E., Buysse D. J., Ancoli-Israel S., Edinger J. D., Krystal A. D., Lichstein K.L., & Morin C. M. (2012). The consensus sleep diary: standardizing prospective sleep self-monitoring. *Sleep, 35*(2), 287–302.

Chen, Z., Lin, M., Chen, F., Lane, N. D., Cardone, G., Wang, R., et al. (2013). Unobtrusive sleep monitoring using smartphones. Presented at the *PervasiveHealth '13: Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare*, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering.

Choe, E. K., Consolvo, S., Watson, N. F., & Kientz, J. A. (2011). Opportunities for computing technologies to support healthy sleep behaviors. Presented at the *CHI '11: Proceedings of the 2011 annual conference on Human factors in computing systems*, ACM.

doi:10.1145/1978942.1979395.

Choe, E. K., Jung, J., Lee, B., Fisher, K. (2013a). Nudging people away from privacy-invasive mobile apps through visual framing. Presented at *INTERACT '13,* 74–91.

Choe, E. K., Lee, B., Munson, S., Pratt, W., & Kientz, J. A. (2013b). Persuasive Performance Feedback: The Effect of Framing on Self-Efficacy. In *AMIA (American Medical Informatics Association) Annual Symposium Proceedings (Vol. 2013*), 825–833.

Choe, E. K., Lee, N. B., Lee, B., Pratt, W., & Kientz, J. A. (2014a). Understanding quantified-selfers' practices in collecting and exploring personal data. Presented at the *CHI '14: Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, ACM. doi:10.1145/2556288.2557372

Choe, E. K., Lee, B., & Kientz, J. A. (2014b). Personal visual analytics for self-monitoring. ACM *DIS 2014 Workshop* on A Personal Perspective on Visualization and Visual Analytics.

Choudhury, T., Consolvo, S., Harrison, B., Hightower, J., LaMarca, A., LeGrand, L., Rahimi, A., Rea, A., Bordello, G., Hemingway, B., Klasnja, P., Koscher, K., Landay, J. A., Lester, J., Wyatt, D., & Haehnel, D. (2008). The Mobile Sensing Platform: An Embedded Activity Recognition System. *Pervasive Computing, IEEE*, *7*(2), 32–41. doi:10.1109/MPRV.2008.39

Clocky. http://www.nandahome.com/clocky/aqua_3/.

Colten, H. R., & Altevogt, B. M. (Eds.). (2006). *Sleep disorders and sleep deprivation: an unmet public health problem* (1st ed.). Washington (DC): National Academies Press.

Consolvo, S., Everitt, K., Smith, I., & Landay, J. A. (2006). Design requirements for technologies that encourage physical activity. Presented at the *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, ACM, 457–466. doi:10.1145/1124772.1124840.

Consolvo, S., Klasnja, P., McDonald, D. W., Avrahami, D., Froehlich, J., LeGrand, L., Libby, R., Mosher, K., & Landay, J. A. (2008). Flowers or a robot army?: encouraging awareness & activity with personal, mobile displays. Presented at the *Ubicomp '08*. 54–63.

Data-Driven Documents. (n.d.). *Data-Driven Documents*. Retrieved May 22, 2014, from http://d3js.org.

Faubel, R., Lopez-Garcia, E., Guallar-Castillón, P., Balboa-Castillo, T., Gutiérrez-Fisac, J. L., Banegas, J. R., & Rodríguez-Artalejo, F. (2009). Sleep duration and health-related quality of life among older adults: a population-based cohort in Spain. *Sleep*, *32*(8), 1059–1068.

fitbit. (n.d.). *fitbit*. Retrieved May 21, 2014, from http://www.fitbit.com.

Fixsen, D. L., Phillips, E. L., & Wolf, M. M. (1972). Achievement Place: the reliability of self-reporting and peer-reporting and their effects on behavior. *Journal of Applied Behavior Analysis*, *5*(1), 19–30.

Fogg, B. J. (2003). *Persuasive technology*. Morgan Kaufmann.

Fox, S., & Duggan, M. (2013). Tracking for Health.

Frost, M., Doryab, A., Faurholt-Jepsen, M., Kessing, L. V., & Bardram, J. E. (2013). Supporting disease insight through data analysis: refinements of the monarca self-assessment system. Presented at the *UbiComp '13: Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, ACM. doi:10.1145/2493432.2493507.

Google Charts API. (n.d.). *Google Charts API*. Retrieved May 22, 2014, from https://developers.google.com/chart

Gottlieb, D. J., Punjabi, N. M., Newman, A. B., Resnick, H. E., Redline, S., Baldwin, C. M., & Nieto, F. J. (2005). Association of sleep time with diabetes mellitus and impaired glucose tolerance. *Archives of Internal Medicine*, *165*(8), 863–867. doi:10.1001/archinte.165.8.863.

Grimes, D. A., & Snively, G. R. (1999). Patients' Understanding of Medical Risks: Implications for Genetic Counseling. *Obstetrics & Gynecology*, *93*(6), 910.

Hayes, S. C., & Cavior, N. (1977). Multiple tracking and the reactivity of self-monitoring. *Behavior Therapy*, *8*, 819–831.

Hekler, E. B., Klasnja, P. V., Froehlich, J., & Buman, M. P. (2013). Mind the theoretical gap: interpreting, using, and developing behavioral theory in HCI research. Presented at the *CHI '13: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 3307–3316. doi:10.1145/2470654.2466452.

Herbert, E. W., & Baer, D. M. (1972). Training parents as behavior modifiers: self-recording of contingent attention. *Journal of Applied Behavior Analysis*, *5*(2), 139–149.

Jawbone UP. (n.d.). *Jawbone UP*. Retrieved May 21, 2014, from https://jawbone.com/up.

Kandel, S., Paepcke, A., Hellerstein, J., & Heer, J. (2011). Wrangler: interactive visual specification of data transformation scripts. Presented at the *CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM. doi:10.1145/1978942.1979444.

Kanfer, F. H., & Gaelick-Buys, L. (1991). *Self-management methods.* Pergamon Press.

Kay, M., Choe, E. K., Shepherd, J., Greenstein, B., Watson, N., Consolvo, S., & Kientz, J. A. (2012). Lullaby: a capture & access system for understanding the sleep environment. Presented at *the UbiComp '12: Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, ACM. doi:10.1145/2370216.2370253.

Kazdin, A. E. (1974). Reactive self-monitoring: the effects of response desirability, goal setting, and feedback. *Journal of Consulting and Clinical Psychology*, *42*(5), 704–716.

Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings (2nd ed.).* Oxford University Press.

Kientz, J. A. (2011). Embedded capture and access: encouraging recording and reviewing of data in the caregiving domain. *Personal and Ubiquitous Computing*, *16*(2), 209–221. doi:10.1007/s00779-011-0380-6.

Kim, S., Kientz, J. A., Patel, S. N., & Abowd, G. D. (2008). Are you sleeping?: sharing portrayed sleeping status within a social network. Presented at the *CSCW '08: Proceedings of the 2008 ACM conference on Computer supported cooperative work*, ACM, 619–628. doi:10.1145/1460563.1460660.

Klasnja, P., & Pratt, W. (2011). Healthcare in the pocket: Mapping the space of mobile-phone health interventions. *Journal of Biomedical Informatics*, *45*, 184–198. doi:10.1016/j.jbi.2011.08.017.

Klasnja, P., Consolvo, S., & Pratt, W. (2011). How to evaluate technologies for health behavior change in HCI research. Presented at the *CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 3063–3072.

Kopp, J. (1988). Self-monitoring: A literature review of research and practice. *Social Work Research & Abstracts*.

Korotitsch, W. J., & Nelson-Gray, R. O. (1999). An overview of self-monitoring research in assessment and treatment. *Psychological Assessment*, *11*(4), 415.

Krishnamurthy, P., Carter, P., & Blair, E. (2001). Attribute Framing and Goal Framing Effects in Health Decisions. *Organizational Behavior and Human Decision Processes*, *85*(2), 382–399. doi:10.1006/obhd.2001.2962.

Kryger, M. H., Roth, T., & Dement, W. C. (2010). *Principles and Practice of Sleep Medicine*. Elsevier Health Sciences.

Lawson, S., Jamison-Powell, S., Garbett, A., Linehan, C., Kucharczyk, E., Verbaan, S., et al. (2013). Validating a mobile phone application for the everyday, unobtrusive, objective measurement of sleep. Presented at the *CHI '13: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* ACM. doi:10.1145/2470654.2481345.

Leger, D. (1994). The cost of sleep-related accidents: a report for the National Commission on Sleep Disorders Research. *Sleep*, *17*(1), 84–93.

Levin, I. P., & Gaeth, G. J. (1988). How consumers are affected by the framing of attribute information before and after consuming the product. *Journal of Consumer Research*, 374–378.

Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes*, *76*(2), 149–188.

Li, I. (2009). Designing personal informatics applications and tools that facilitate monitoring of behaviors. *UIST 2009*.

Li, I., Dey, A. K., & Forlizzi, J. (2011). Understanding my data, myself: supporting self-reflection with ubicomp technologies. Presented at the *UbiComp '11: Proceedings of the 13th international conference on Ubiquitous computing*, ACM. doi:10.1145/2030112.2030166.

Li, I., Dey, A., & Forlizzi, J. (2010). A stage-based model of personal informatics systems. Presented at the *CHI '10: Proceedings of the 29th International Conference on Human Factors in Computing Systems*, 557–566.

Lin, J., Mamykina, L., Lindtner, S., Delajoux, G., & Strub, H. B. (2006). Fish'n'Steps: Encouraging Physical Activity with an Interactive Computer Game. *Ubicomp 2006*, 261–278.

Lipinski, D. P., Black, J. L., Nelson, R. O., & Ciminero, A. R. (1975). Influence of motivational variables on the reactivity and reliability of self-recording. *Journal of Consulting and Clinical Psychology*, *43*(5), 637–646.

Lipinski, D., & Nelson, R. (1974). The reactivity and unreliability of self-recording. *Journal of Consulting and Clinical Psychology*, *42*(1), 118–123. doi:10.1037/h0036059.

Lipkus, I. M., & Hollands, J. G. (1999). The visual communication of risk. *Journal of the National Cancer Institute. Monographs*, (25), 149–163.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation. A 35-year odyssey. *The American Psychologist*, *57*(9), 705–717.

Mace, F. C., & Kratochwill, T. R. (1985). Theories of reactivity in self-monitoring. *Behavior Modification*.

Mamykina, L., Mynatt, E., Davidson, P., & Greenblatt, D. (2008). MAHI: investigation of social scaffolding for reflective thinking in diabetes management. Presented at the *CHI '08: Proceedings of the 2008 annual conference on Human factors in computing systems*, ACM, 477–486. doi:10.1145/1357054.1357131.

Maquet, P. (2001). The Role of Sleep in Learning and Memory. *Science*, *294*(5544), 1048–1052. doi:10.1126/science.1062856.

Marteau, T. M. (1989). Framing of information: Its influence upon decisions of doctors and patients. *British Journal of Social Psychology*, *28*(1), 89–94. doi:10.1111/j.2044-

8309.1989.tb00849.x.

Mattila, E., Pärkkä, J., Hermersdorf, M., Kaasinen, J., Vainio, J., Samposalo, K., et al. (2008). Mobile diary for wellness management--results on usage and usability in two user studies. *IEEE Transactions on Information Technology in Biomedicine : a Publication of the IEEE Engineering in Medicine and Biology Society*, *12*(4), 501–512. doi:10.1109/TITB.2007.908237.

Mazze, R. S., Shamoon, H., Pasmantier, R., Lucido, D., Murphy, J., Hartmann, K., et al. (1984). Reliability of blood glucose monitoring by patients with diabetes mellitus. *The American Journal of Medicine*, *77*(2), 211–217. doi:10.1016/0002-9343(84)90693-4.

McFall, R. M. (1970). Effects of self-monitoring on normal smoking behavior. *Journal of Consulting and Clinical Psychology*, *35*(2), 135–142.

McFall, R. M. (1977). *Parameters of self-monitoring*. Behavioral self-management: Strategies.

Michie, S., Johnston, M., Francis, J., Hardeman, W., & Eccles, M. (2008). From theory to intervention: mapping theoretically derived behavioural determinants to behaviour change techniques. *Applied Psychology*, *57*(4), 660–680. doi:10.1111/j.1464-0597.2008.00341.x.

Min, J. K., Doryab, A., Wiese, J., Amini, S., & Zimmerman, J. (n.d.). Toss “N”Turn: smartphone as sleep and sleep quality detector. Presented at the *CHI '14: Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, ACM.

Morris, D., Saponas, T. S., Guillory, A., & Kelner, I. (2014). RecoFit: using a wearable sensor to find, recognize, and count repetitive exercises. Presented at the *CHI '14: Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, ACM.

Nelson, R. O. (1977). Assessment and therapeutic functions of self-monitoring. In M. Hersen, R. M. Eisler, & P. M. Miller (Eds.), *Progress in Behavior Modification* (Vol. 5, pp. 263–308). New York: Academic Press.

Nelson, R. O., Lipinski, D. P., & Boykin, R. A. (1978). The effects of self-recorders' training and the obtrusiveness of the self-recording device on accuracy and the reactivity of self-monitoring. *Behavior Therapy, 9*, 200–208.

Nelson, R. O., & Hayes, S. C. (1981). Theoretical explanations for reactivity in self-monitoring. *Behavior Modification*, *5*(1), 3–14. doi:10.1177/014544558151001.

Ozenc, K. F., Brommer, J. P., Jeong, B., Shih, N., Au, K., & Zimmerman, J. (2007). Reverse Alarm Clock. *DPPI '07*. 392–406.

Prochaska, J. O., DiClemente, C. C., & Norcross, J. C. (1992). In search of how people change. Applications to addictive behaviors. *The American Psychologist*, *47*(9), 1102–1114.

Quantified Mind. (n.d.). *Quantified Mind*. Retrieved May 22, 2014, from http://www.quantified-mind.com.

Romanczyk, R. G. (1974). Self-monitoring in the treatment of obesity: Parameters of reactivity. *Behavior Therapy*, *5*(4), 531–540.

Rosenthal, R. (1966). Experimenter effects in behavioral research.

Rothman, A. J., & Salovey, P. (1997). Shaping perceptions to motivate healthy behavior: the role of message framing. *Psychological Bulletin*, *121*(1), 3–19.

Shambroom, J. R., & Fabregas, S. E. (2012). Validation of an automated wireless system to monitor sleep in healthy adults. *Journal of Sleep Research, 21*, 221–230. doi: 10.1111/j.1365-2869.2011.00944.x.

Shirazi, A. S., Clawson, J., Hassanpour, Y., Tourian, M. J., Schmidt, A., Chi, E. H., Borazio, M., & Laerhoven, K. V. (2013). Already up? using mobile phones to track & share sleep behavior. *International Journal of Human-Computer Studies*, *71*(9), 878–888. doi:10.1016/j.ijhcs.2013.03.001.

Sieck, W. A., & McFall, R. M. (1976). Some determinants of self-monitoring effects. *Journal of Consulting and Clinical Psychology*, *44*(6), 958–965.

Sleep Cycle. (n.d.). *Sleep Cycle*. Retrieved May 21, 2014, from http://www.sleepcycle.com.

Sleep Sound Conditioner. (n.d.). http://www.amazon.com/Marpac-Dohm-DS-Speed-Sound-Conditioner/dp/B000KUHFGM.

Stepanski, E. J., & Wyatt, J. K. (2003). Use of sleep hygiene in the treatment of insomnia. *Sleep Medicine Reviews*, *7*(3), 215–225.

Stone, A. A., Shiffman, S., Schwartz, J. E., Broderick, J. E., & Hufford, M. R. (2003). Patient compliance with paper and electronic diaries. *Controlled Clinical Trials*, *24*(2), 182–199.

Strecher, V. J., McEvoy DeVellis, B., Becker, M. H., & Rosenstock, I. M. (1986). The Role of Self-Efficacy in Achieving Health Behavior Change. *Health Education & Behavior : the Official Publication of the Society for Public Health Education*, *13*(1), 73–92. doi:10.1177/109019818601300108.

Swan, M. (2009). Emerging patient-driven health care models: an examination of health social networks, consumer personalized medicine and quantified self-tracking. *International Journal of Environmental Research and Public Health*, *6*(2), 492–525.

The OneTouch UltraMini. (n.d.). *The OneTouch UltraMini*. Retrieved May 25, 2014, from http://www.onetouch.com/onetouch-ultramini.

Thoresen, C. E., & Mahoney, M. J. (1974). *Behavioral self-control*. Holt, Rinehart & Winston.

Troshynski, E., Lee, C., & Dourish, P. (2008). Accountabilities of presence: reframing location-based systems. Presented at the *CHI '08: Proceedings of the 2008 annual conference on Human factors in computing systems*, ACM, 487–496.

Truong, K. N., Shihipar, T., & Wigdor, D. J. (2014). Slide to X: unlocking the potential of smartphone unlocking. Presented at the *CHI '14: Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, ACM. doi:10.1145/2556288.2557044.

Tversky, A. A., & Kahneman, D. D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458.

Wagner, U., Gais, S., Haider, H., Verleger, R., & Born, J. (2004). Sleep inspires insight. *Nature*, *427*(6972), 352–355. doi:10.1038/nature02223.

WatchPAT. (n.d.). http://www.itamar-medical.com/WatchPAT%E2%84%A2/Medical_Professional.html.

Wobbrock, J.O. (n.d.). Seven research contribution types in human-computer interaction. Retrieved August 16, 2014, from https://faculty.washington.edu/wobbrock/teaching/types.pdf.

Yang, H., Li, Y., & Zhou, M. X. (2014). Understand users' comprehension and preferences for composing information visualizations. ACM Transactions on Computer-Human Interaction (TOCHI), 21(1), 6.

Yi, J. S., Kang, Y. A., Stasko, J. T., & Jacko, J. A. (2008). Understanding and characterizing insights: how do people gain insights using information visualization? In *Proceedings of the 2008 Workshop on Beyond time and errors: novel evaluation methods for Information Visualization* (p. 4). ACM.