SleepTight: Low-burden, Self-monitoring Technology for Capturing and Reflecting on Sleep Behaviors

Eun Kyoung Choe
Pennsylvania State University
echoe@ist.psu.edu

Bongshin Lee
Microsoft Research
bongshin@microsoft.com

Matthew Kay, Wanda Pratt, Julie A. Kientz
University of Washington
{mjskay, wpratt, jkientz}@uw.edu

ABSTRACT

Manual tracking of health behaviors affords many benefits, including increased awareness and engagement. However, the capture burden makes long-term manual tracking challenging. In this study on sleep tracking, we examine ways to reduce the capture burden of manual tracking while leveraging its benefits. We report on the design and evaluation of SleepTight, a low-burden, self-monitoring tool that leverages the Android's widgets both to reduce the capture burden and to improve access to information. Through a four-week deployment study ($N = 22$), we found that participants who used SleepTight with the widgets enabled had a higher sleep diary compliance rate (92%) than participants who used SleepTight without the widgets (73%). In addition, the widgets improved information access and encouraged self-reflection. We discuss how to leverage widgets to help people collect more data and improve access to information, and more broadly, how to design successful manual self-monitoring tools that support self-reflection.

Author Keywords

Sleep; health; self-monitoring; self-tracking; personal informatics; Quantified Self; manual tracking; self-reflection; self-awareness.

ACM Classification Keywords

H.5.2 [Information interfaces and presentation (e.g., HCI)]: User Interfaces.

INTRODUCTION

Self-monitoring requires an individual to deliberately record the occurrences of his or her target behavior [29]. When a person's behavior departs from his or her performance standard, a self-regulatory process is triggered [16], which contributes to behavior change that is mostly toward desirable, therapeutic directions [29]. For example, tracking what we eat is known to influence how we eat, thereby contributing to healthy eating behaviors [20]. Similarly, tracking sleep and reflecting on sleep behaviors lie at the core of instrumenting cognitive behavior therapy for insomnia (CBTI) [31]. This engagement with data collection enhances people's awareness and provides an opportune moment to reflect on their behaviors. In this sense, an obtrusive, high burden recording device—such as a manual tracking tool—can augment this positive, therapeutic behavior change [22].

Many recent efforts in ubiquitous computing have emphasized tools to automatically track health-related behaviors through wearable sensors (e.g., [12]) or smartphone-based monitoring (e.g., [28,34]). However, automating the process reduces people's awareness and engagement, which are often crucial in encouraging behavior change [24]. Manual tracking (or self-report) affords these benefits, but it imposes a high capture burden. This high burden and people's tendency to forget both compromise data quality due to missing or inaccurate data. These problems aggravate when tracking multiple behaviors at once. Thus, in practice, manually tracking multiple behaviors over the long-term remains challenging.

Our goal was to enhance the ***manual tracking*** method by designing a tool that supports people in easily capturing and reflecting on ***multiple behavioral factors***, while still preserving its advantages for behavior change. We explored this topic in the context of sleep tracking. Sleep is an interesting yet challenging application area for manual tracking because many contributing factors (e.g., meals, exercise, caffeine, alcohol, tobacco, and medication) are difficult to track automatically, and some factors, by definition, can only be tracked manually (e.g., subjective sleep quality). Moreover, contributing factors affect people differently [35], which makes it hard to define a fixed set of tracking factors that would work for everyone. Therefore, we examined ways to support easy yet flexible manual capture of target behaviors (i.e., sleep duration, sleep quality, to-bed time, and wake-up time) and contributing factors (i.e., factors that would affect the sleep measures). We used Android's lock screen and home screen widgets to reduce the capture burden and improve access to information. This new tool we designed and developed, called SleepTight (Figure 1), is a low-burden, self-monitoring application to help people capture and reflect on sleep behaviors. We evaluated SleepTight through a four-week field deployment

UbiComp '15, September 7–11, 2015, Osaka, Japan.

ACM 978-1-4503-3574-4/15/09…$15.00.
http://dx.doi.org/xx.xxxx/xxxxxxx.xxxxxxx

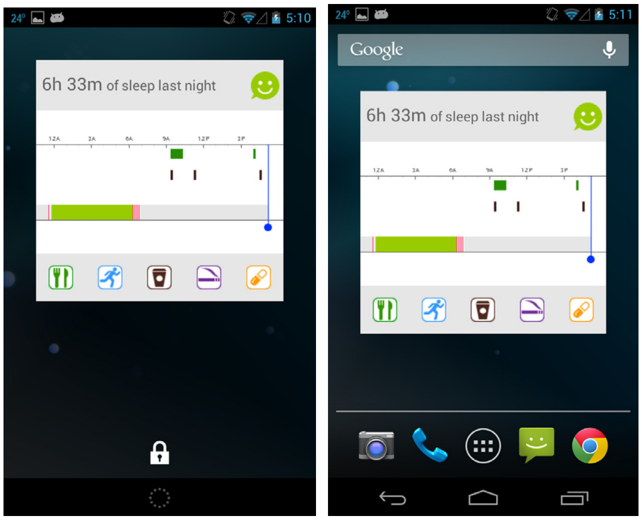

Figure 1. SleepTight implemented as an Android app widget. SleepTight running on the Android's lock screen (left) and home screen (right).

study with 22 participants accompanied by pre and post interviews and weekly questionnaires. We found that participants who used SleepTight with the widgets enabled had a higher sleep diary compliance rate (92%) than participants who used SleepTight without the widgets (73%). In addition, participants in the widget condition accessed SleepTight's various features more frequently than participants in the non-widget condition.

In what follows, we summarize related work and describe both design goals and detailed design instrumentation for SleepTight. We then describe the study design, report on our findings from the field deployment study, and discuss considerations and opportunities for low burden manual tracking technology design. The contributions of this work are twofold: (1) the design and development of SleepTight, a novel example of a manual tracking tool for capturing sleep behaviors as well as contributing factors, and (2) the empirical study of SleepTight that helps to understand facilitators for designing effective manual tracking tools that capture data and encourage self-reflection.

RELATED WORK

In this section, we cover related work in the areas of (1) self-monitoring, (2) data capture mechanisms in self-monitoring, and (3) self-monitoring for sleep.

Self-monitoring

Self-monitoring emerged as an area of research within behavioral psychology. It has been studied in the clinical and research settings since the 1970s for its treatment and therapeutic effects. More recently, self-monitoring has been widely embedded in the design of sensing and mobile applications. In 2014, major IT companies announced fitness and health tracking platforms—e.g., Apple's HealthKit and Google Fit—for capturing, storing, and retrieving data about health and fitness activities. Their major push is a result of a dramatic increase in the usage of health and fitness apps [13]. Moreover, many standalone fitness and health tracking devices such as Fitbit, Jawbone Up, and Microsoft Band have gained ground in the past few years—one market research shows that one in ten Americans over the age of 18 owns an activity tracker [15]. Enthusiastic self-trackers attend Quantified Self Meetups or annual Quantified Self Conferences to share their best practices and lessons learned [32]. On the research side, HCI researchers have been designing and evaluating self-monitoring technology in several domains including physical activity [27], sleep [17], stress [14], and food [21]. Researchers have identified several barriers toward the adoption of self-monitoring technology [9,24]. These challenges include lack of scientific rigor, missing important contextual information, and trying to track more data than necessary which leads to tracking fatigue and thus incomplete datasets for effective analysis. In this work, we have been particularly interested in ways to support people to track sleep behaviors and important contextual information while keeping the tracking burden low.

Data Capture Mechanisms in Self-monitoring

To conduct self-monitoring, a person needs to choose a target behavior and capture mechanism to track the target behavior. Capture mechanisms encompass a broad spectrum of tools ranging from pen and paper to a more sophisticated technology. The choice of the tool often involves tradeoffs. To capture sleep measures, for example, some people use manual tracking tools (e.g., paper or electronic sleep diaries) for documenting self-report measures. The advantage of a manual tracking method is the flexibility of choosing a target behavior and increased self-awareness due to direct engagement with data collection. However, people are prone to forgetfulness and delayed recording could compromise data accuracy. On the other extreme, people use sensing—either wearable [12] or embedded [1]—to automatically track sleep. These sensing tools have the potential to reduce mental workload and increase data accuracy, but can be cumbersome to wear—in the case of wearable sensing—and can reduce people's awareness of the data collected [24].

Recognizing the benefits of manual tracking, many self-monitoring technologies incorporate manual tracking in their system to complement sensing. For example, Fitbit allows people to manually add many things that cannot be automatically captured from its accelerometer, including food, activities (e.g., swimming), water consumed, and heart rate. Similarly, Ubifit [11] and Wellness Diary [25] 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) [33]. Our work builds upon these systems by specifically minimizing the number of steps required to manually capture events.

Sleep-monitoring

Choe et al. conducted an in-depth formative study to identify the design space of sleep technologies [8]. They 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 crucial qualities of sleep technologies.

Over the past few years, much sleep-related research has been published in the HCI and Ubiquitous Computing literature. This research is primarily concerned with providing sleep hygiene recommendations [2], tracking sleep [7,23, 26,33] and environmental disruptors [17], tracking sleep apnea [30], and sharing sleep data with others [19,33]. Although we see a growing number of sleep-monitoring technologies, most of these systems track target behaviors only. We see opportunities in designing a sleep tracking system that could support capturing both ***target behaviors*** (sleep measures) and ***contributing factors*** that are likely to influence the target behaviors so that people can understand the relationships between the target behaviors and contributing factors. A few examples that support capturing multiple data streams include Lullaby—which captures target behaviors as well as environmental sleep disruptors [17]—and the Mobile Health Mashups—which shows significant correlations across multiple data streams of automatically sensed data (e.g., location, weather, calendar), sensor inputs (e.g., step count, sleep, weight), and manual logging (e.g., food, mood, pain) [4]. Similar to Lullaby and Mobile Health Mashups, our goal was to support people in capturing multiple data types. In contrast to Lullaby's tracking of environmental sleep disruptors, we supported capturing people's behavioral factors initially determined by sleep clinicians. In addition, we aimed to design a very lightweight manual capture tool that provides the flexibility to add custom activities, which contrasts with Mobile Health Mashups' design in which people are bounded by a fixed set of factors that they can capture.

SLEEPTIGHT

In this section, we describe SleepTight's three design goals, which we drew from prior self-monitoring literature, sleep literature, and input from our collaborating sleep researcher. We then describe SleepTight's design and implementation details to support the design goals.

Design Goals

The first design goal (G1) was to enable people to capture both ***target behaviors*** and ***contributing factors*** that are likely to influence the target behaviors. Prior self-tracking literature found that novice self-trackers make the common mistake of tracking only the target behaviors and not the potential contributing factors or context [9]. 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 may miss vital information on how to improve the target behaviors. To address this problem, we decided to provide people with information about what to capture, including both target behaviors and contributing factors based on the sleep hygiene literature [35].

The second design goal (G2) was to ***reduce the capture burden*** and create a consistent capturing habit. The captured data points must be accurate enough 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 often fail to complete manual journaling as instructed and that they generate fake or backfilled written entries—which are likely to be inaccurate because of recall bias—to give the appearance of good compliance [36]. If a tracking task imposes too much burden, people will give up self-monitoring entirely. Therefore, we sought ways to make the manual tracking very easy and to discourage inaccurate backfilling.

The last design goal (G3) was to ***provide feedback to promote self-reflection***. Even a simple form of self-monitoring feedback constitutes self-reflection and contributes to behavior change [18]. However, self-reflection on multiple data streams is complex and time-consuming. We aimed to provide feedback that encourages frequent self-reflection and helps people make sense of the relationships among various factors to find ways to improve their behaviors. Supporting this goal was twofold: (1) we designed insightful, easy to understand feedback, and (2) we improved access to the feedback to promote frequent self-reflection.

SleepTight Design

We implemented two versions of SleepTight: (1) ***Full-system*** (which included the lock screen widget, home screen widget, and app) and (2) ***App-only system*** (no widgets, app-only). We hypothesized that the Full-system's widgets would support G2 and G3 better than the app-only system while both versions would support G1. In this section, we first describe SleepTight's widget design and how it supports our design goals. We then describe how the app supports data capturing and self-reflection, and end this section with SleepTight's implementation details.

Leveraging App Widgets

Android widgets are miniature application views that can be embedded in other applications including the *lock screen* and the *home screen* (Figure 1). Widgets are automatically updated and always shown on the lock screen or home screen. Thus, people can quickly access application data without launching the full app or even unlocking the phone.

We leveraged the widgets to support the design goals of reducing the capture burden (G2) and providing feedback for self-reflection (G3). SleepTight's widgets reduce the capture burden by providing an ability to capture ***current contributing factors*** through a single tap—the simplest interaction people can do with a mobile phone. For example, tapping on the caffeinated drink icon from the lock screen widget (i.e., no need to unlock the phone to capture data) captures the time stamp of clicking the caffeinated

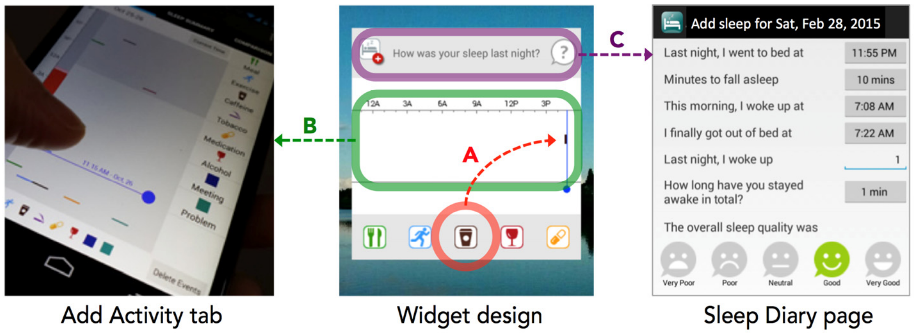

Figure 2. SleepTight's widget allows easy data capture from the lock screen or home screen (A), allows easy access to the full app (B, C), and serves as a glanceable display.

icon (Figure 2–A). In addition, SleepTight's widget provides easy access to the full app. For example, tapping on the timeline region invokes the Add Activity tab (Figure 2–B). Tapping on the sleep summary region at the top invokes the Daily Sleep Diary page (Figure 2–C, more on the next section) or the Sleep Summary tab if people have already completed the sleep diary. Lastly, SleepTight's widgets serve as a glanceable display, which provide visual feedback on people's activity and sleep logs. People can quickly see their captured data on the timeline presented on the lock screen or home screen widgets, which is designed to promote frequent reflection. We note that people using the *App-only system* could still capture sleep and other contributing factors by going through the following steps: (1) unlock the phone, (2) go to the "Apps" page (e.g., home screen), (3) look for the SleepTight icon, (4) open the app, (5) go to Add Activity page, and (6) click one of the factor icons or the sleep icon from the left column (Figure 3–A).

Capturing Target Behaviors and Contributing Factors

To support our first design goal of enabling people to capture both target behaviors and contributing factors (G1), we consulted with a sleep researcher to determine important entries that can influence individuals' sleep behaviors. SleepTight supports capturing target behaviors (e.g., sleep quality, sleep duration, or sleep efficiency) and contributing factors (e.g., daytime and nighttime activities that are likely to impact sleep). SleepTight's default list of potential contributing factors consists of six items known by the sleep medicine community to impact sleep—meals, exercise, caffeine, alcohol, tobacco, and medication [6]. Moreover, sources of sleep disturbances could be individualistic, so SleepTight allows people to customize the list of tracked activities by removing activities or adding their own new activities they believe might contribute to sleep quality. People can track up to eight contributing factors, but only five factors are shown on the widget due to the widget size. In addition, people can reorder the contributing factors, of which the top five factors from the list are shown in the widget (Figure 2–B). For example, if a person wants to track his tobacco use but does not want others to know, he can hide the tobacco icon from the widget but still track it from the full app (Figure 3–A).

The **Daily Sleep Diary** page is accessible by clicking a link from the widget (Figure 2–C) or from the app (Figure 3–A). A diary entry for the previous night's sleep is enabled between 12:00 AM and midnight of the following day, thus having a 24-hour time window to complete the diary. We made this design choice to prevent backfilling and to create a consistent capturing habit (G2). The diary page includes required questions (subjective sleep quality, to-bed time, to-sleep time, wake-up time, and out-bed time) and optional questions (number of awakenings, total duration of being

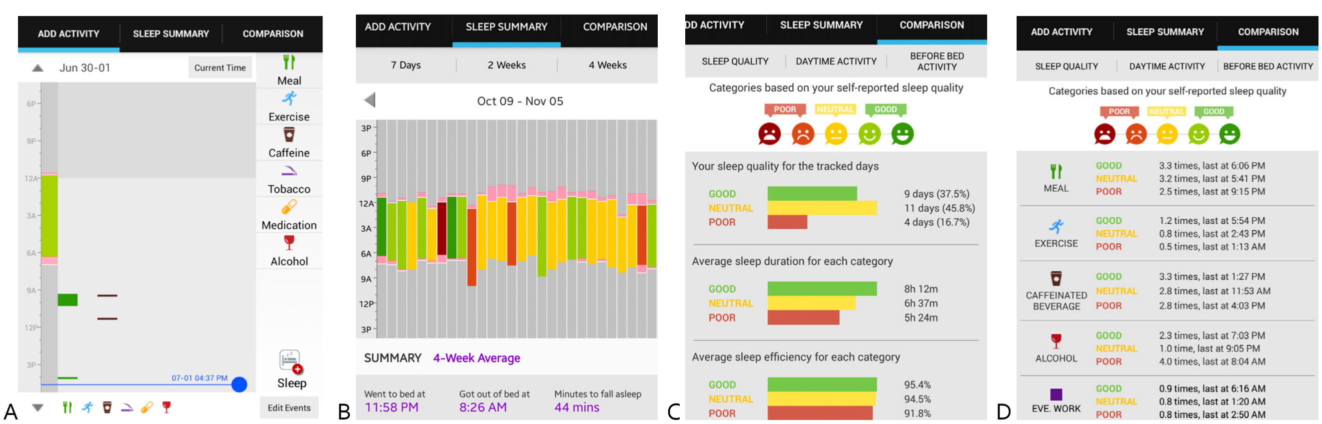

Figure 3. SleepTight's app pages: Add Activity tab and vertical timeline (A), Sleep Summary tab—4-week view (B), Comparison tab—sleep behaviors group by sleep quality (C), and Comparison tab—contributing factors group by sleep quality (D).

awake during the sleep, nighttime activities, and sleep disturbances). Subjective sleep quality is measured using a 5-point Likert-type scale, ranging from very poor to very good. We color-coded the subjective sleep quality with visuals—“red frowny face” for “very poor” and “green smiley face” for “very good.”

The **Add Activity tab** is the landing page of the full SleepTight app and is also accessible from clicking the widget’s timeline (Figure 2–B). From the Add Activity tab, people can record ***current or past contributing factors***—either their duration or frequency—and view the recorded data. For example, assume that a person had three meals and two caffeinated beverages but did not capture these factors at the moment. He or she can capture these factors retrospectively by opening the Add Activity tab, dragging the time bar (shown as a blue line with a handle in Figure 3–A) to the time when the activity occurred, and tapping the activity icon from the list on the right. To record the duration of an activity, he or she can touch-and-hold an activity icon to invoke the duration input dialog.

Providing Feedback about People’s Sleep Behaviors

SleepTight provides two types of post-hoc feedback on aggregated sleep behaviors—(1) Sleep Summary and (2) Comparison. We designed this feedback to help people reflect on their sleep behaviors and contributing factors (G3).

The **Sleep Summary tab** visualizes sleep patterns in terms of sleep duration and quality, and provides a descriptive summary of sleep measures for a given time frame, such as the past week, two weeks, or four weeks (Figure 3–B). The y-axis represents time of day (24-hour duration) and the x-axis represents date; each gray bar represents a single day. The solid rectangle represents the actual 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 tab is depicted from four timestamps—to-bed time, to-sleep time, wake-up time, and 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, SleepTight provides a detailed daily summary view, which consists of daily sleep summary, activity logs, nighttime activities, and sleep disturbances.

The **Comparison tab** allows a within-subjects comparison, which is one of the most popular data exploration techniques among self-trackers [10]. The Comparison tab helps people learn which activities contributed to different sleep qualities by grouping sleep behaviors (e.g., sleep duration, sleep efficiency) and contributing factors by sleep quality. The first part of the Comparison tab shows the number of days for each subjective sleep quality category—good, neutral, and poor sleep (Figure 3–C). The next part shows the average sleep duration and sleep efficiency for each sleep quality. For example, Figure 3–C shows that the average sleep duration was longer when the sleep quality was good (8 hrs, 12 mins) than when the sleep quality was poor (5 hrs, 24 mins). Figure 3–D shows the average frequency and last timestamp of contributing factors categorized by sleep quality. For example, comparing the caffeine consumption between the days with good sleep quality and poor sleep quality (Figure 3–D), this person had an average of 3.3 *versus* 2.8 caffeinated beverages, and the last caffeine was consumed at 1:27 PM *versus* 4:03 PM. Lastly, people can compare the top five frequent nighttime activities (activities a person conducted during an hour before sleep) for the days with good sleep quality against the days with neutral or poor sleep quality.

SleepTight Implementation

We implemented SleepTight as a client-server system. The client side was implemented as a native Android App, sending and retrieving data to and from the SleepTight server on the web. We implemented the SleepTight server with Ruby on Rails using a MySQL database. We used JSON for the data communication between the client and server, the Java 2D API, which provides rendering methods, for all the graphical views.

DEPLOYMENT STUDY

In this section, we detail our study design, participants, study procedure, data collected, and analysis method. Our university’s institutional review board approved this study.

Study Design

We designed a between-subjects study to evaluate the effects of widgets by comparing the two versions of the SleepTight system—(1) ***Full-system*** (which included the lock screen widget, home screen widget, and app) and (2) ***App-only system*** (no widgets, app-only). The study included two in-person sessions (pre and post interviews) and four weeks of in-situ use of SleepTight. We randomly assigned participants to either of the two conditions, which allowed us to assess the effect of the lock screen and home screen widgets. In Table 1, we summarize how each design goal was implemented in each version of SleepTight system. We note that the first goal (G1) was implemented in both systems to keep the number of possible capturing items equal (which allowed us to evaluate the capture burden). However, we also used the interview to examine whether supporting this goal was beneficial for the participants.

Participants

We recruited participants via convenience sampling. We sent out recruitment emails to various mailing lists. The email contained a link to the screening questionnaire. From the 80 people who responded, 22 participants met the following inclusion criteria: (1) own an Android phone that runs the operating system version greater than or equal to 4.2.2 (the version that supports lock screen widgets); (2) have a data plan; (3) do not have a diagnosed sleep disor-

Table 1. SleepTight's implementation of design goals.

Design Goals	Full-system (With widgets)	App-only system (No widgets)	
Capture outcome behaviors and contributing factors (G1)		Equally support	
Reduce capture burden (G2)	Number of steps to capture	Fewer steps	More steps
	Access to the capture tool	High access	Low access
Provide feedback to promote self-reflection (G3)	Feedback	More feedback (app+widgets)	Less feed-back (app)
	Access to the feedback	High access	Low access

der[1]; (4) interested in tracking and learning about their sleep behaviors and sleep contributing factors; and (5) not traveling between time zones during the 4-week study period. We recruited people who could use their own phone to ensure the most natural usage in the wild and to limit any effects of participants needing to spend time getting used to a new phone or having to carry two devices.

Among the 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). Although 91% (n = 20) of the participants had used home screen widgets, only 18% (n = 4) reported that they have experience using lock screen widgets. Eighteen participants (82%) expressed that they have sleep goals. Their goals included 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. Among the six participants who had experience using sleep tracking apps or devices to track sleep, only one person used it everyday.

Study Procedure

The first in-lab session lasted about 90 minutes, which consisted of a background survey, semi-structured interview on factors impacting sleep, SleepTight installation, and brief instructions on SleepTight's basic features.

While participants were completing the surveys, we installed SleepTight on their mobile phone. We installed the *Full-system* (FS) to half of the participants (n = 11, 6 female) and *App-only system* (AS) to the other half (n = 11, 7 female)[2]. After the installation, we conducted a semi-structured interview to probe participants' sleep habits, sleep rituals, and potential contributing factors. Lastly, we walked participants through a demonstration of SleepTight, helped them configure the settings, and instructed them on the use of SleepTight. We allowed participants to modify the settings as they used SleepTight.

For the following four weeks, participants were instructed to voluntarily use SleepTight. We did mention to all participants that they would receive better quality feedback if they collect more data. We made it clear to the participants that the compensation is not tied to their actual usage of SleepTight. During the deployment study period, we sent out weekly online questionnaires (compliance rate was 96.6%) 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 based on participants' tracking logs. We probed about any unusual behaviors and asked them to explain them. We also asked participants about their experience with SleepTight focusing on their typical usage pattern and gained information. We offered $100 USD in a gift card to compensate participants in appreciation for their time. This amount is consistent with compensation for similar studies conducted in our area.

Dataset and Analysis

The study produced a rich dataset. Tracking data captured by participants—referred to as the ***tracking log***—was stored in our remote web server. In addition, we instrumented SleepTight to log participants' usage data—referred to as the ***usage log***—in a separate log file, which we downloaded during the exit interview session[3]. We used the t-test and Mann–Whitney U test to analyze both the tracking logs and usage logs to compare the overall usage between the Full-system and App-only system conditions.

Additionally, we audio-recorded and transcribed both initial and exit interviews, and segmented weekly questionnaires. To analyze the qualitative data, we used a general inductive approach [37], 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. The lead author read through the transcripts several times to identify themes and categories regarding (1) SleepTight's effect on people's self-reflection on their sleep behaviors; and (2) people's tracking routines. The qualitative analysis complemented

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

[2] We will use "FS" to denote the participants assigned to the *Full-system* condition, and "AS" to those assigned to the *App-only system* condition. FS or AS followed by a number (e.g., FS-1, AS-1) indicates a specific participant assigned to either of the condition.

[3] Due to a technical difficulty and unexpected event—such as participants switching to a new phone during the study period, we were able to retrieve 17 out of 22 participants' usage log (8 FS and 9 AS participants).

the quantitative results from tracking and usage log analysis. Lastly, we took screenshots of participants' SleepTight pages and widgets, which gave us an overview of the types of feedback participants received during the study.

RESULTS

We organized the results according to the following topics: (1) data capture behavior, (2) information access, and (3) self-reflection with SleepTight.

Data Capture Behavior

To assess the efficacy of a self-monitoring tool, researchers measure participants' adherence rate (i.e., the number of diary entries during the study period) (e.g., [36]). Therefore, we measured ***sleep diary adherence***—defined by the number of sleep diary entries over the course of 28 days—and the ***total number of captured contributing factors***.

Sleep Diary Adherence and Number of Captured Factors

Diary adherence for the FS condition (M = 25.89, SD = 2.71) was significantly higher than that of AS condition (M = 20.42, SD = 7.18), $t(14.85) = 2.42$, $p = .03$. The average adherence rate was 92% for the FS condition and 73% for the AS condition (Figure 4).

Analyzing the usage log revealed that among the diary entries captured by FS participants, 88% of the sleep entries were captured from either the home screen widget (77%) or the lock screen widget (11%) whereas the remaining 12% were captured from the Add Activity tab from the app. Participants in the FS condition heavily used the widgets to access the sleep diary page to capture target behaviors; the widgets were shortcuts to the data capture page and/or served as visual reminders prompting people to record the sleep diary.

In terms of the number of total captured contributing factors over the course of 28 days, we did not find a significant difference; on average, participants in the FS condition tracked 152.11 factors (SD = 68.82) and 26.72 factors per category (SD = 13.89) while participants in the AS condition tracked 141.5 factors (SD = 78.00) and 20.32 factors per category (SD = 10.35), $t(18.41) = .33$, $p = .75$.

Among the contributing factors captured by FS participants, 91% of them were recorded from the Add Activity tab from the app whereas the remaining 9% was recorded from either the home screen widget (7%) or the lock screen widget (2%). This result indicates that participants did not use SleepTight as a real-time capturing tool and that they captured the factors in a retrospective manner.

Although participants could technically use SleepTight as a near real-time tracking tool, they often captured the contributing factors retrospectively, thereby creating a ***time lag*** between the time of an activity and the time of a capture. A big time lag could mean less accurate data due to recall bias [5]. To assess the widgets' effect on captured data accuracy, we analyzed the time lag difference between the two conditions (Figure 5). The time lag was significantly smaller for the participants in the FS condition (M = 7.06 hours, SD = 3.33) than those in the AS condition (M = 11.66 hours, SD = 5.00), $t(18.81) = -2.52$, $p = .02$. This result indicates that, on average, the participants in the FS condition captured an event significantly closer to its actual time than participants in the AS condition. Thus, the accuracy of the activity data could be greater for the FS condition as well.

During the exit interview, participants in the FS 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*."

Although widgets helped FS participants record daytime contributing factors closer to their actual time than AS participants, on average, 7-hour time lag still existed between when activities were conducted and captured. Our analysis of the usage logs showed that participants in both conditions tended to record daytime activities towards bedtime (Figure 6). During the exit interview, participants in both conditions confirmed that before bedtime was an opportune moment for data capture for many reasons—the memories of when things happened were still fresh in mind; capturing several factors in a row was convenient once they open the

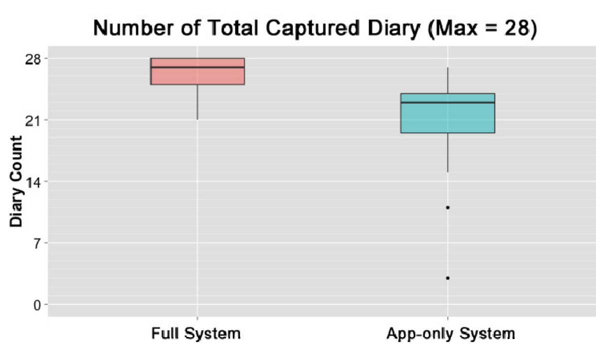

Figure 4. FS participants captured more diary entries than AS participants ($p = .03$), $N = 22$.

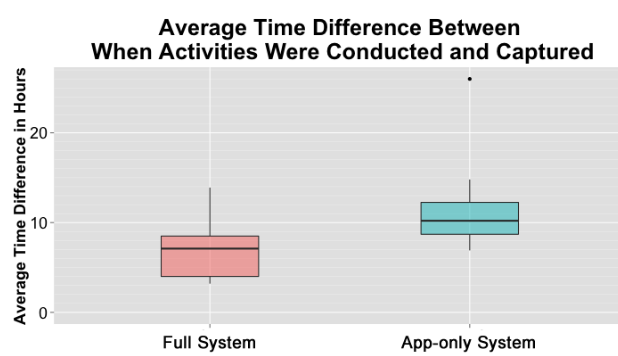

Figure 5. We observed shorter time lag in the Full-system condition than App-only system condition ($p = .02$), $N = 22$.

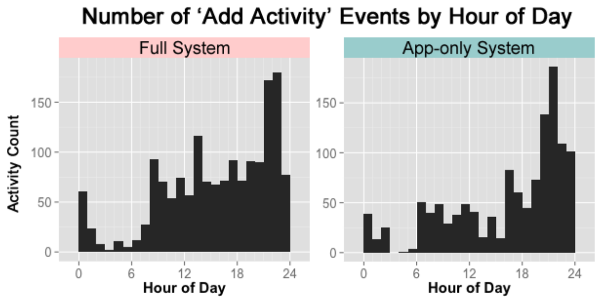

Figure 6. Participants in both conditions showed a tendency to record daytime activities towards bedtime, *N* = 22.

app; and real-time capturing could be socially awkward (e.g., interrupting a meal with a friend). For these reasons, participants had a tendency to delay the data capture, which explains the low usage of the lock screen widget for capturing contributing factors.

Information Access

To assess the widgets' effect on information access, we examined total minutes used over the study period. Participants' total usage time of the FS condition (Mdn = 37.41) was greater than but did not differ significantly from that of the AS condition (Mdn = 20.48), W = 17, p = .074, r = -.43. We, however, 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 measured from the usage log. Given that 88% of the sleep entries were captured from the widgets, the usage time of the FS condition would be substantially greater than what we reported.

We next analyzed what features participants frequently accessed. Figure 7 shows participants' detailed usage of SleepTight's various features during the 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, visual inspection suggests that FS participants accessed various features of SleepTight more frequently than AS participants.

Among the captured events, we analyzed the number of times the "Sleep Summary" page (Figure 3–B) was viewed. Participants in the FS condition (Mdn = 91.5) viewed the summary page more frequently than those in the AS condition (Mdn = 22), W = 67, p = .002, r = -0.77. This result indicates that the lock screen and home screen widgets reminded participants to view the sleep summary page and offered a shortcut to the sleep summary page. Thus, we can conclude that widgets afford frequent self-reflection.

Self-reflection with SleepTight

To examine what people learned during self-reflection with SleepTight, we analyzed qualitative data gathered from weekly surveys and exit interviews, in which we asked, "What did you learn while using SleepTight?" Due to the qualitative nature of the data, we did not seek measurable differences between the groups. Here, we present the types of self-reflection with example quotes from both groups.

Being able to capture both target behaviors and contributing factors allowed participants **in both conditions** to self-reflect on their sleep behaviors in various meaningful ways. However, participants in the FS condition brought up concerns with respect to projecting personal data on their widgets, which we will address later in the discussion.

Analyzing participants' self-reflection descriptions, we identified two dimensions. The first dimension was ***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.* The other dimension was ***topic***—for example, whether a description was about *sleep patterns*; *other activity patterns*; *relationships between sleep and other factors*; or *tracking habits*. Table 2 shows the category summary and example quotes for each category.

The majority of self-reflection descriptions were about *findings* on participants' *sleep patterns*. We suspect that this result was due to the consistent capturing of sleep behaviors using the daily sleep diary. From the aggregated sleep data, participants were able to figure out their sleep patterns such as to-bed time and wake-up time and the consistency of their sleep pattern. 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.

Both versions of SleepTight also increased participants' awareness of *other activities besides sleep*, such as their drinking habits (e.g., "*I don't drink as much alcohol as I thought I did*"), nighttime activities, or non-routine events.

Another type of self-reflection was about *findings* on the

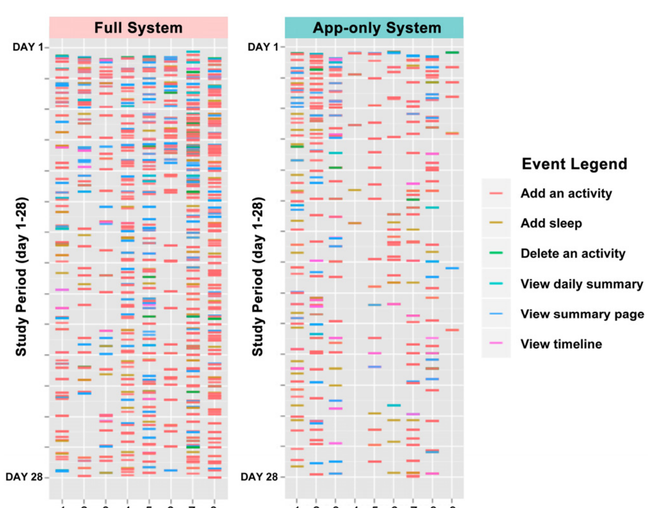

Figure 7. Chromograms of usage over the entire study period by condition. Each column represents each participant's data, *N* = 17.

Table 2. Categories of self-reflection description and example quotes.

Level of Certainty	Topics	Example Quotes
Conclusive Finding	Sleep 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 at any sort of daily tracking activity [FS-7]." [Confirmation]
Hypothesis	Sleep 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]."

relationship between *sleep and other factors*. Because SleepTight allows people to track multiple factors at a time, participants identified associations 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*." They also became aware of how their nighttime activities affect 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.*"

Some participants described with care what they had learned, acknowledging that there could be flaws in their reflection due to few data points or other confounding factors. We marked those cases with '*hypothesis.*'

In summary, participants in both conditions were able to learn their sleep patterns and other activity patterns with SleepTight. In doing so, it was helpful to capture target behaviors and contributing factors as well as seeing feedback on the Add Activity tab and Sleep Summary tab. Participants found the Comparison tab "very data centric" and not helpful in identifying relationships among the captured factors. They identified associations among the factors from their careful observation and self-awareness of behavior rather than feedback from the Comparison tab. Some participants inferred a causal relationship between different factors and sleep quality, which may well be incorrect.

DISCUSSION

In this section, we discuss lessons learned, limitations of SleepTight, and implications for self-monitoring technology design focusing on what makes manual tracking successful. We begin by revisiting the three initial design goals and then extend our discussion to other implications.

Capturing both Target Behaviors and Contributing Factors

SleepTight was designed as a self-management tool that people can use without the help of a medical professional. Identifying target behaviors and contributing factors is not always straightforward because what seems to be a target behavior could actually be a contributing factor and vice versa (e.g., lack of sleep results in an increased caffeine intake). Working closely with a domain expert (a sleep clinician in our case) is crucial in configuring the initial tracking setting and determining the default activities to be tracked. Furthermore, systems like SleepTight can be best used for the purpose of hypothesis generation rather than hypothesis testing. After generating a plausible hypothesis, rigorous self-experimentation could be conducted to test the hypothesis. It would then require a more advanced statistical approach to model the relationships between multiple explanatory and dependent variables.

Reducing the Capture Burden through Widgets

SleepTight's widgets contributed to higher sleep diary adherence rate. We suspect that the widgets mediated this effect by serving as a visual reminder and reducing the capture and access burden. Our study has implications for what it means to reduce the capture burden. First, it should be easy to remember to capture data. Although researchers often use time-based notification (e.g., [3]) to facilitate reminders, our study showed the power of *visual reminder*. Second, it should be effortless to access the capturing tool. Widgets not only served as a capturing tool but also provided a direct link to the Sleep Diary page and Add Activity tab where people can capture data. While data entry itself should be easy, a reminder and shortcut can aid with timely data capture. As we relaxed the data precision to reduce the capture burden in the SleepTight design, it will be interesting to explore ways to capture fine-grained data by finding the right balance between automated sensing and manual tracking in future work.

Leveraging Manual Tracking in Self-reflection

An important finding from the SleepTight study was ***when*** people self-reflect. Aside from times when they were entering data, participants rarely took a time to look at the feedback and ponder upon it. The very difference compared to a fully automated tool is this extra opportunity for data awareness. 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 daytime activities. Participants had to think about how many drinks or caffeinated beverages they had when they enter data. Moreover, visual feedback was an important linkage between people's awareness and motivation to track. Looking at the empty bar (missing data) on the Sleep Summary tab encouraged people to track data in a prompt manner next time. We will further explore ways to communicate complicated information—such as correlations among multiple factors—by leveraging the visual feedback as our Comparison tab design was too text and data heavy.

Projecting Personal Data onto Widgets in a Positive Light

As SleepTight supports capturing personal data and projects the captured data onto lock screen, it runs a risk of making people become overly anxious. A recent study on people's phone unlocking behaviors showed that on average, people unlock their phones between 4.8 and 105.3 times per day [38]. Because SleepTight projects individuals' sleep, alcohol, caffeine, tobacco, and other behaviors on the widgets, it could cause added stress, especially when the data shows negative information about oneself, which was also reported by Consolvo and colleagues [11]. Not wanting to see negative information (e.g., a bright red frowny face for negative sleep quality) every time a person unlocks the phone, one participant (FS-2) entered skewed data overestimating his behavior. Thus, feedback from the widget should be encouraging and judgment-free and yet correctly conveying the current state, which is particularly challenging when the data contains negative information about oneself. Moreover, two participants in the Full-system condition brought up privacy concerns regarding projecting their sleep behaviors and other behavioral factors onto the lock screen because it can be easily seen by other people around them. People were genuinely interested in understanding how various factors (e.g., sexual activity, stress) affect their sleep, and thus, they frequently added these factors as custom items. At the same time, they did not want others to know that they were tracking these factors, not to mention projecting them onto the lock screen. Although having only five items on the widget was a design decision made from space limitations, providing ways to reorder the items turned out to be helpful for some people who wanted to have a control over what to show and hide. Thus, the tool did afford an ability to hide or mask items deemed too private for the lock screen.

Identifying and Capturing Anomalies

Existing self-monitoring tools, including SleepTight, are designed to capture everyday behaviors, but do not distinguish anomalies from routines. We asked participants to collect nighttime activities with an assumption that what they do right before bed would influence people's sleep quality. However, it turned out that most people have a set of activities that they do every day—called "nighttime routines"—which did not have much relationship with their sleep quality. These activities included brushing their teeth, watching TV, talking to their spouse, or reading a book. What affected people's sleep more were non-routine activities and events—for example, having friends come over, travelling, or preparing for an exam. Therefore, once a self-monitoring tool identifies people's routines, it should distinguish routines from anomalies and encourage people to collect anomalies. Rare events are valuable data points.

Limitations

We note that good data capture by itself does not necessarily lead to behavior change and that studying the effects of SleepTight on long-term behavior change warrants future research. However, people can leverage and make sense of the data once they collect it, so easy data capture is an important requisite for behavior change to occur.

We could not show how much the widget contributed to the increased awareness in addition to the App-only condition because people's level of self-awareness was reported qualitatively. A quantitative approach to measure self-awareness and self-reflection could help examine these issues. Also, we assumed that people have one tracking widget installed on their mobile phone; however, having multiple tracking widgets on a smartphone might pose extra privacy concerns and access burdens for the person.

CONCLUSION

We presented the design and evaluation of the SleepTight system, a lightweight manual tracking system for helping people capture sleep and other contributing factors. To evaluate the effects of widgets on tracking adherence, information access, and self-reflection, we conducted a four-week, between-subjects deployment study comparing the Full-system (lock screen, home screen, and app) to the App-only system. We found that SleepTight's widgets helped people capture more data by serving as visual reminders and providing a shortcut to the full app. Widgets also helped people capture events close to their actual time, although social activities and workflow prevented them from capturing events real-time. Participants in both conditions were able to reflect on their sleep behaviors and sleep-related activities and identified findings and hypotheses about their sleep patterns, other activity patterns, and relationships among multiple factors. The results from this study demonstrated the value of manual tracking with the widgets including quick way to capture contextual data, self-reflection, and engagement, which can augment the benefits of self-monitoring.

ACKNOWLEDGMENT

We would like to thank our study participants, and Chang Whan Jung and Sooyong Wang for their assistance with the initial system development. This work was supported by Intel Science & Technology Center for Pervasive Computing and NSF Grant #IIS-1344613.

REFERENCES

1. Aura. http://www.withings.com/us/withings-aura.html.
2. Bauer, J.S., Consolvo, S., Greenstein, B., Schooler, J., Wu, E., Watson, N.F., & Kientz, J.A. (2012). ShutEye: encouraging awareness of healthy sleep recommendations with a mobile, peripheral display. *CHI '12*.
3. Bentley, F. & Tollmar, K. (2013). The power of mobile notifications to increase wellbeing logging behavior. *CHI '13*, 1095–1098.
4. 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 (TOCHI), 20*(5), 1–27.
5. Brown, G. D., & Chater, N. (2001). The chronological organization of memory: Common psychological foundations for remembering and timing. *Time and memory: Issues in philosophy and psychology*, 77-110.
6. 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.
7. Chen, Z., Lin, M., Chen, F., Lane, N. D., Cardone, G., Wang, R., Li, T., Chen, Y., Choudhury, T., & Campbell, A.T. (2013). Unobtrusive sleep monitoring using smartphones. *PervasiveHealth '13*. 145–152.
8. Choe, E.K., Consolvo, S., Watson, N.F., & Kientz, J.A. (2011). Opportunities for computing technologies to support healthy sleep behaviors. *CHI '11*, 3053–3062.
9. Choe, E.K., Lee, N.B., Lee, B., Pratt, W., & Kientz, J.A. (2014). Understanding Quantified-Selfers' practices in collecting and exploring personal data. *CHI '14*, 1143–1152.
10. Choe, E.K., Lee, B., & schraefel, m.c. (2015). Characterizing visualization insights from Quantified-Selfers' personal data presentations. *IEEE Computer Graphics and Applications, 35*(4), 28–37.
11. 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. *Ubicomp '08*.
12. Fitbit. http://www.fitbit.com.
13. Flurry. http://www.flurry.com/blog/flurry-insights/health-and-fitness-apps-finally-take-fueled-fitness-fanatics#.VADBjmSwKVt.
14. Hernandez, J., Paredes, P., Roseway, A., & Czerwinski, M. (2014). Under pressure: sensing stress of computer users. CHI *'14,* 51–60.
15. Inside Wearables. http://endeavourpartners.net/assets/Wearables-and-the-Science-of-Human-Behavior-Change-EP4.pdf.
16. Kanfer, F.H. & Gaelick-Buys, L. (1991). *Self-management methods.* Pergamon Press.
17. 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. *UbiComp '12*, 226–234.
18. Kazdin, A.E. (1974). Reactive self-monitoring: the effects of response desirability, goal setting, and feedback. *J Consult Clin Psych, 42*(5), 704–716.
19. Kim, S., Kientz, J.A., Patel, S.N., & Abowd, G.D. (2008). Are you sleeping?: sharing portrayed sleeping status within a social network. *CSCW '08*, 619–628.
20. Kong, A., Beresford, S.A., Alfano, C.M., Foster-Schubert, K.E., Neuhouser, M.L., Johnson, D.B., Duggan, C., Wang, C.Y., Xiao, L., Jeffery, R.W., Bain, C.E., & McTiernan, A. (2012). Self-monitoring and eating-related behaviors are associated with 12-month weight loss in postmenopausal overweight-to-obese women. *Journal of the Academy of Nutrition and Dietetics, 112*(9), 1428–1435.
21. Kong, F. & Tan, J. (2012). DietCam: Automatic dietary assessment with mobile camera phones. *Pervasive and Mobile Computing, 8*(1), 147–163.
22. Korotitsch, W.J. & Nelson-Gray, R.O. (1999). An overview of self-monitoring research in assessment and treatment. *Psychological Assessment, 11*(4), 415.
23. Lawson, S., Jamison-Powell, S., Garbett, A., Linehan, C., Kucharczyk, E., Verbaan, S., Rowland, D.A., & Morgan, K. (2013). Validating a mobile phone application for the everyday, unobtrusive, objective measurement of sleep. *CHI '13*, 2497–2506.
24. Li, I., Dey, A. K., & Forlizzi, J. (2012). Using context to reveal factors that affect physical activity. *ACM Transactions on Computer-Human Interaction (TOCHI), 19*(1), 7.
25. Mattila, E., Parkka, J., Hermersdorf, M., Kaasinen, J., Vainio, J., Samposalo, K., Merilahti, J., Kolari, J., Kulju, M., Lappalainen, R., & Korhonen, I. (2008). Mobile diary for wellness management—results on usage and usability in two user studies. *Information Technology in Biomedicine, 12*(4), 501–512.
26. Min, J.K., Doryab, A., Wiese, J., Amini, S., & Zimmerman, J. (2014). Toss 'N' Turn: smartphone as sleep and sleep quality detector. *CHI '14*, 477–486.
27. Morris, D., Saponas, T.S., Guillory, A., & Kelner, I. (2014). RecoFit: using a wearable sensor to find, recognize, and count repetitive exercises. *CHI '14*, 3225–3234.
28. Moves. https://www.moves-app.com.
29. Nelson, R.O. & Hayes, S.C. (1981). Theoretical explanations for reactivity in self-monitoring. *Behavior Modification, 5*(1), 3–14.
30. Oliver, N. & Flores-Mangas, F. (2007). HealthGear: automatic sleep apnea detection and monitoring with a mobile phone. *Journal of Communications*, *2*(2), 1–9.
31. Perlis, M. L., Jungquist, C., Smith, M. T., & Posner, D. (2006). *Cognitive behavioral treatment of insomnia: A session-by-session guide (Vol. 1).* Springer Science & Business Media.
32. Quantified Self. http://quantifiedself.com.
33. 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. *IJHCS, 71*(9), 878–888.

34. Sleep Cycle. http://www.sleepcycle.com.
35. Stepanski, E.J. & Wyatt, J.K. (2003). Use of sleep hygiene in the treatment of insomnia. *Sleep medicine reviews, 7*(3), 215–225.
36. Stone, A.A., Shiffman, S., Schwartz, J.E., Broderick, J.E., & Hufford, M.R. (2003). Patient compliance with paper and electronic diaries. *Control Clin Trials, 24*(2), 182–199.
37. Thomas, D.R. (2006). A general inductive approach for analyzing qualitative evaluation data. *Am J Eval, 27*(2), 237–246.
38. Truong, K.N., Shihipar, T., & Wigdor, D.J. (2014). Slide to X: unlocking the potential of smartphone unlocking. *CHI '14,* 3635–3644.