Self-Experimentation for Behavior Change: Design and Formative Evaluation of Two Approaches

Jisoo Lee
School of Arts, Media and Engineering
Arizona State University
jisoo.lee@asu.edu

Erin Walker
School of Computing, Informatics, and Decision Systems Engineering
Arizona State University
erin.a.walker@asu.edu

Winslow Burleson
Rory Meyers College of Nursing
New York University
wb50@nyu.edu

Matthew Kay
Computer Science & Engineering | dub
University of Washington
mjskay@cs.washington.edu

Matthew Buman
School of Nutrition and Health Promotion
Arizona State University
mbuman@asu.edu

Eric B. Hekler
School of Nutrition and Health Promotion
Arizona State University
ehekler@asu.edu

ABSTRACT

Desirable outcomes such as health are tightly linked to behaviors, thus inspiring research on technologies that support people in changing those behaviors. Many behavior-change technologies are designed by HCI experts but this approach can make it difficult to personalize support to each user's unique goals and needs. This paper reports on the iterative design of two complementary support strategies for helping users create their own personalized behavior-change plans via self-experimentation: One emphasized the use of interactive instructional materials, and the other additionally introduced context-aware computing to enable user creation of "just in time" home-based interventions. In a formative trial with 27 users, we compared these two approaches to an unstructured sleep education control. Results suggest great promise in both strategies and provide insights on how to develop personalized behavior-change technologies.

Author Keywords

Behavior change; self-experimentation; just-in-time interventions; context-aware computing

ACM Classification Keywords

H.5.2. User Interfaces: User-Centered Design; Theory & Methods; J.4 Computer Applications; Social & Behavioral Sciences.

INTRODUCTION

Extensive evidence suggests the importance of people's sustained engagement in behaviors to improve health, productivity, and wellbeing [39,50]. For example, daily brushing is important for oral health [1] and regular physical activity can reduce risk of cardiovascular disease, obesity, and colon cancer [50]. Patients with Type 2 diabetes are recommended a number of behaviors such as monitoring glucose, taking medications, physical activity, and eating low sugar diets [17]. However, it is common for individuals to struggle with initiating and sustaining health behaviors [60]. This issue has inspired a large effort in the human-computer interaction (HCI) community to generate plausible solutions for supporting behavior change [e.g., 10].

Many of these behavior-change technologies (see related work) are designed, implemented, and evaluated by experts. An alternative and complementary approach for supporting more personalized and precise behavior change could be to help individuals create their own behavior change plans. Behavior change plans are the approaches a person takes to initiate and maintain a desired behavior, including the use of behavior-change techniques from the scientific literature but also, plausibly, other self-created approaches. This self-creation approach is linked to the Quantified Self (QS) movement, where individuals work to better understand themselves through self-tracking/self-study, including methods that they create [9,35]. Choe et al. [9] found that *"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."* A number of studies have been conducted in the HCI community focused on providing improved resources for self-experimentation, such as data collection and interpretation tools [36].

In this paper, we explore theoretically grounded mechanisms for supporting users' self-experimentation. Karkar et al. define self-experimentation as requiring three phases: formulating a hypothesis, testing the hypothesis with N-of-1 trial designs, and examining the results of the study [24]. Our work extends the concept of self-experimentation to the systematic study of the behavior-change plans one could use to initiate and maintain health behaviors, which we label **self-experimentation for behavior change**. We investigated two approaches for facilitating self-experimentation for behavior change. First, we designed **interactive instructional materials** to support users in the creation of and experimentation with behavior-change plans. This approach focuses on giving users tools to design and implement behavior-change plans compatible with their goals and lifestyle. Second, we used end-user programmable sensing and feedback to support the design of **"just-in-time" (JIT) interventions**, which provide triggers to engage in a desired behavior during states when a person has both the opportunity to engage in the behavior and the receptivity to interact with the system [48]. Just-in-time interventions are a logical target for self-experimentation for behavior change because JIT strategies are often context-sensitive and idiosyncratic. For example, if a person is trying to improve diet, a JIT intervention requires insights on when, where, with whom, and in what state (e.g., stress-eating) a person may be in when eating too much to define the JIT states when a prompt would actually be helpful.

In this paper, we first describe prior work in HCI focused on behavior-change technologies. Next, we describe our iterative design process in the creation of our interactive instructional materials and our context-aware JIT intervention system. We then report on a 7-week formative evaluation, which tests these two approaches for improving sleep relative to a sleep education control. We hypothesized that both self-experimentation for behavior change approaches would produce significantly improved sleep relative to an unstructured self-experimentation/education-only control condition. The key contributions of our study include:

1) Empirical results in favor of our structured self-experimentation for behavior change strategies for improving sleep relative to our unstructured control.
2) Concrete suggestions for personalization of behavior change interventions via self-experimentation. These suggestions generalize to other interventions attempting to scaffold self-experimentation for behavior change.
3) The use of a Bayesian statistical approach to conduct our formative evaluation, extending previous work [24]. This concrete use-case of Bayesian statistics for formative work provides details on what is gained from these analyses and can serve as a template for the HCI community.

RELATED WORK

Behavior-Change Technologies in HCI

HCI has become increasingly interested in studying the use of computing technology to promote behavior change [15,21]. A key approach has focused on improving a users' self-awareness, typically via sensing technologies for self-tracking and feedback from data. For example, *Affective Diary* [55] facilitated users' affective interpretation and reflection on daily experiences, by providing abstract body figures that represented movement and arousal levels throughout the day. MAHI [42] provided a website where diabetes patients and their educators communicated via diaries, with an explicit goal of fostering improved reflective skills among the patients. Li asserted the usefulness of users' exploration of multiple types of contextual and behavioral information in a single interface to support identification of factors that affect behavior [36]. Bentley and his colleagues [3] created a system that automatically finds correlations between a variety of contextual factors (weight, sleep, step count, etc.) and people's health and wellbeing.

Another popular strategy for supporting behavior change involves goal-setting and self-monitoring. In *UbiFit* [10] users were invited to establish a weekly goal for various activities (Cardio, Strength, Flexibility) and then provided feedback via the growth of a virtual garden. *Fish'n'Steps* [38] invites users to set their daily step goal, gathers player's step counts, and presents users' activity achievements via changes to a virtual character such as growth and facial expressions. In *Kunini* [6], players set goals that required them to run specific distances or paces before a specific date. More recently, Rabbi et al. [53] explored a system-driven personalization approach to goal-setting within *MyBehavior*. Specifically, *MyBehavior* gives suggestions on physical activity and dietary behavior based on continuously collected information on each user's behavior. In addition to goal-setting and self-monitoring, HCI has adopted a broad range of concepts from behavioral theory, including just-in-time information (e.g., Nawyn et al. [49]), priming (e.g., Consolvo et al. [10]), social validation (e.g., Toscos et al. [57]), and behavioral economics (e.g., Lee et al. [34]).

These examples involve HCI researchers encapsulating behavior-change techniques into a technical system for users. Behavior-change techniques are *"observable, replicable, and irreducible component[s] of a [behavioral] intervention designed to alter or regulate behavior; that is, a technique is proposed to be an 'active ingredient' (e.g., feedback, self-monitoring, and reinforcement)"* [45]. While users may benefit from this exposure to behavior-change techniques, they may feel a lack of agency in the implementation of these techniques, or may feel as though the techniques are not relevant to their individual needs. For example, King et al. [26] developed three smartphone apps focused on improving mid-life and older adults' physical activity that used different behavior-change techniques (e.g. social, analytic, or more game-like). Qualitative work suggested many individuals requested a "mix-and-match" approach at different times, thus reinforcing the need for personalization over time.

Self-Experimentation & Behavior Change

There is great opportunity for creating tools that enable the end-user in selection and personalization of behavior-change techniques [21,46]. This aligns with practices Quantified Selfers' engage in when attempting to find answers for themselves. A number of studies have investigated Quantified Selfers' self-tracking needs and methods for data collection and analysis to empower end-users in their own self-discovery process [9,35]. To enhance such self-discovery, Karkar et al. [24] established a framework for self-experimentation that includes formulating a hypothesis, conducting rigorous n-of-1 trials to support evidence-based decision-making, and reviewing results to facilitate further learning and study. There framework was developed for users interested in understanding how various actions (e.g., foods eaten) impact a symptom or other clinically-relevant outcome for that individual (e.g., irritable bowel syndrome symptoms). More recently, *SleepCoacher* [13] provides personalized recommendations related to sleep based on a person's self-tracking data (e.g., "…when your bedtime is consistent you tend to fall asleep faster. For the next {N} days, try going to bed at a consistent bedtime, around {N}am/pm") and to then test that recommendation with a n-of-1 trial. Results from this work highlight how experimenting on recommendations can improve sleep. A complementary set of tools could support individuals in the selection and personalization of behavior-change techniques and then, through self-experimentation, examine if those techniques help an individual initiate and maintain a behavior, which we call self-experimentation for behavior change.

Self-experimentation for behavior change is distinct and complementary to Karkar's self-experimentation for discovery. Karkar's framework provides a thoughtful strategy for helping an individual to select which behaviors to change to produce a desired outcome (e.g., better mood, reduced stress, reduced symptoms). After a person knows what they "should" do, a separate process is needed to study *how to change and maintain* the targeted behavior over time. A great deal of prior work highlights that there is a gap between what individuals intend to do vs. what they actually do [54], thus establishing the need for self-experimentation for behavior change. The creation and formative evaluation of approaches for self-experimentation for behavior change is the core focus of our work.

Self-Experimentation & Context-Aware Computing

One interesting strategy for supporting personalized behavior change could involve end-user programming of context-aware computing systems to enable individuals in the creation of their own personalized just-in-time interventions. Just-in-time adaptive interventions are an emerging class of behavioral interventions focused on providing support during the "just-in-time" moments when a person is vulnerable to engaging in negative behaviors and/or whenever opportunities for positive changes arise while also being receptive to interacting with a system [48]. For instance, *SitCoach* [59] is an intervention that provides office workers with a message to be physically active whenever the computer has detected 30-minutes of non-active work. For this message to be useful, an individual needs to both have the opportunity to be active soon after receiving the message and is receptive to receiving the message (e.g., not deep in thought on a task). Poppinga et al. [52] identified factors (time of day, phone position) that provide insights on receptivity and found that notifications are more likely to be answered before 8:21 a.m. and after 8:20 p.m., but not late at night, while the phone is in the person's hand. Theoretically, just-in-time notifications are meant to inspire action in that moment and, by extension, rely less on memory and self-control [47]. Further, based on the high likelihood that defining a JIT moment is both highly context-dependent and idiosyncratic [22], a plausible and, as of yet, under-studied area, could be in the development of technology that supports end-users in creating JIT interventions.

There is a long tradition in end-user development research for technologies like JIT interventions [8,37]. End-user development is characterized by the use of techniques that allow non-technical people to create applications [11]. This strategy is valuable when the problem space could benefit from intimate knowledge about activities and environments to design useful solutions such as context-aware applications [14]. Accordingly, there has been considerable research on end-user development tools for creation of context-aware applications in home environments [14,18]. However, most existing tools for users' creation of context-aware applications are to support control of appliances or environmental equipment [14,18]. Little attention has been given to the provision of toolkits focused on behavior change, thus establishing the key gap filled by our work.

ITERATIVE DESIGN OF THE INTERACTIVE INSTRUCTIONAL MATERIALS

In this section, we describe the creation and iterative testing of interactive instructional materials to support users in self-experimentation for behavior change. More specifically, we conducted two user studies to create the protocol used in our formative evaluation trial. For details of each iteration, see [31,32]. This formative work revealed several design strategies one could take to support goal-setting, the selection of behavior-change techniques, and self-monitoring.

As a starting point, we define **self-experimentation for behavior change** as a process executed by users to formulate, test, and iterate on hypotheses related to how well behavior plans can produce desired behavioral outcomes, including behavioral initiation and maintenance. Formulating, testing, and iterating on a behavioral plan corresponds to the three phases of self-experimentation delineated by Karkar et al. but, in this context, the focus is on testing a behavioral plan, not on testing if an action influences an outcome. A **behavioral plan** includes: 1) a goal; 2) a consciously chosen behavior-change technique(s) that is personalized by and for the individual; and 3) self-monitoring of the behavioral target to examine if the behavior-change technique fostered achievement of the goal.

Based on personal experience with training students in this approach within a class setting, our formative work [31,32], and design practices common in HCI, we explicitly did not include n-of-1 trials (e.g., [28]) within our 7-week trial. The key reason was because we have found that introduction of n-of-1 trials too early into a design process can have the unintended consequence of undermining the creative process, which is essential when formulating a hypothesis about a useful behavioral plan. We discuss this in detail in the discussion. To simplify the design process, we explicitly chose to develop highly structured protocols administered by a research assistant, which eventually could be used to design an interactive digital tutorial. This protocol was fully scripted and images were provided in succession via a presentation. The protocol had five steps: 1) Choosing a behavior to attempt to change (the target behavior), 2) Setting a goal, 3) Generating ideas for attainment of the goal by applying behavior-change techniques, 4) Formulating a final plan, consisting of one or more complementary behavior-change techniques, and 5) Devising self-tracking measures to determine if the goal was accomplished.

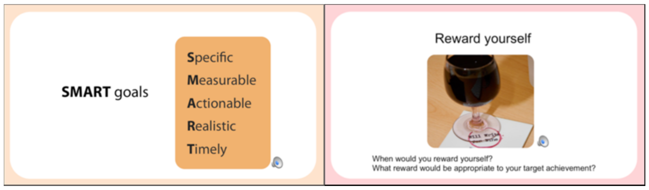
Our first test of this protocol included 2 sessions. In the first, users completed the above five steps. In the second, one week later, users reviewed results of their implementation of the behavioral plan including reviewing their self-tracking data. Results of our first user study revealed two problems. First, individuals generated under-specified goals that were not actionable. Second, individuals did not use the behavior-change techniques provided in the training (e.g., they often used the first behavior-change idea they came up with rather than the evidence-based suggestions).

Based on these results, we refined the protocol in two ways. First, during goal creation (Step 2), we included the concept of SMART goals [30]. The SMART (Specific, Measurable, Actionable, Realistic, and Timely) goal concept is a reinterpretation of Locke and Latham's goal setting theory [41]. According to this concept, goals that meet each of the acronym's words (e.g., specific, measurable) will be more useful for behavior change. During Step 2, users generated SMART goals by: (1) Reflecting on the issues they wanted to work on, (2) learning about the concept of 'behavioral goals,' in contrast to 'outcome goals', and (3) learning about the concept of a SMART goal with instructions on how to create one (Figure 1).

Second, during step 3 (behavior-change techniques), we provided an organizing structure to help individuals in the selection and personalization of behavior-change techniques. We leveraged two existing meta-models, Fogg's behavior model [16], and Michie's COM-B model [44], which were developed to help professionals create interventions. We simplified these models for the purposes of the instructional materials to a phrase, "A behavior occurs when, opportunity, ability, motivation, and a trigger all align." As part of this, we provided definitions for each concept. To support iteration and "self-diagnosis" of factors that impact behavioral initiation and maintenance, during the first session, users were provided only a single behavior-change technique from each domain (opportunity, motivation, ability, and trigger, see Figure 1 for one example slide). In the second session, users were presented with the meta-model phrase and each technique they tried the previous week was linked to the meta-model. They were then asked to self-diagnose if there might be a lack of opportunity, ability, motivation, or triggers when it comes to enacting their behavioral goal. Following this self-diagnosis, users were presented with additional behavior-change techniques for the diagnosed problem domain (e.g., given all of the opportunity behavior-change techniques). Note that if users self-diagnosed the problem as including multiple domains (e.g., both opportunity and motivation), then they were provided with both sets of added behavior-change techniques.

In terms of self-monitoring, our formative work indicated the value of providing individuals with two types of self-monitoring; open-ended journaling, which was particularly valuable for generating hypotheses about behavioral plans, and phone-based surveys with clear quantification of the target outcome(s) and any process variables (e.g., stress) that they thought might influence their outcome. We did not change this strategy in the second iteration.

We conducted a second test, this time using a 3-week protocol to better understand users' iterative process. As described elsewhere [31,32], these changes resulted in better specified goals, improved engagement with behavior-change techniques, and increased iteration in terms of further personalizing the behavior-change techniques chosen.



SMART goals

Specific
Measurable
Actionable
Realistic
Timely

Reward yourself

When would you reward yourself?
What reward would be appropriate to your target achievement?

Figure 1. Exemplar slides of the prototype guiding users' behavioral plan creation.

JUST-IN-TIME INTERVENTION TOOL

Our second self-experimentation for behavior change strategy focused on the self-creation and testing of JIT interventions using context-aware computing. We developed a context-sensitive application, which integrates off-the-shelf hardware and software to support the creation of context-aware JIT interventions. The system enables rapid prototyping of simple rule-based systems that include physical sensing, data storage, and media event components. Scripted sequences of media events are triggered based on time, sensed activity, and/or history of behavior.

We favored a rule-based approach due to its logical simplicity and flexibility across situations [14,18,58]. For example, a rule for detecting meal preparation might be: 'IF resident is in the kitchen AND (resident accessed cupboard AND resident accessed plates OR utensils cabinet) OR resident used an appliance THEN a meal was prepared' [12].

For sensing, we adopted X10 (www.x10.com) and Insteon (www.insteon.net) home automation sensors. Currently, the tool supports X10 wireless open/closed magnetic sensors, X10 wireless motion sensors, and Insteon on/off modules (see Figure 2). The JIT intervention toolkit also includes two types of prompting methods: audio content via wireless speakers and text messages via mobile devices. The audio prompts can include machine speech of user-inputted text or it can play user-added/selected sound files (e.g., music or other mp3 files). The use of sound as a prompt has several advantages over visual display. First, audio can catch a person's attention even if the user is not looking at the device. In addition, sound, especially music, is well known to influence emotions [27,40]. However, the audio prompts are limited by the need for individuals to be near the speakers. We additionally used text messages for prompting users for those times when a person is not near a speaker but has their phone [51]. The commercially available home automation software, Indigo (www.indigodomo.com), was employed for hardware communication (i.e., Apple computer, X10 and Insteon sensors, Apple wireless speaker system, mobile phones) and application programming.

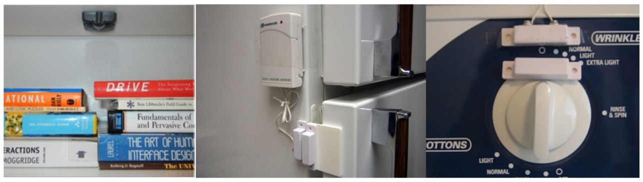


Figure 2. Sensor Use Examples; X10 motion sensor to detect users' taking a book (the left), X10 door sensor to detect users' opening the refrigerator (the middle), X10 door sensor to detect start/end of the laundry (the right).

FORMATIVE EVALUATION STUDY

Overview

We conducted a formative evaluation of the two self-experimentation for behavior change approaches on users' sleep quality. We chose sleep because sleep is an essential factor that affects an individual's physical vitality, emotional balance, and productivity, poor sleep is common, and various factors influence sleep such as other behaviors (e.g., bedtime, diet), psychological disturbances, pain, medical conditions, genetic factors, stress, age, physiological and cognitive arousal [29]. Further, sleep hygiene includes well-researched behavioral strategies to improve sleep [29], including: 1) Go to bed and get up at the same time each day; 2) Avoid napping; 3) Have a regular schedule for meals, medications, chores, and other activities; 4) Avoid stimulants such as caffeine (e.g., coffee, chocolate), nicotine, and alcohol near bedtime; and 5) Stay away from large meals near bedtime. Previous research suggests knowledge of sleep hygiene alone does not translate to improved sleep [56], thus making sleep hygiene a good control as it provides likely new information that will inspire the sort of unstructured self-experimentation common by QS'ers.

We randomly assigned users to one of three conditions: 1) Sleep Hygiene alone (SH), 2) Sleep Hygiene+Self-Experimentation for Behavior Change Tutorial (SH+SBT), and 3) Sleep Hygiene+JIT Self-Experimentation for Behavior Change Tutorial+JIT Intervention Tool (SH+SBT+JIT).

We hypothesized both self-experimentation for behavior change approaches would improve sleep quality compared to the SH control over 7 weeks. Our primary outcome was the Pittsburgh Sleep Quality Inventory (PSQI) [5], but we also collected daily self-reported sleep satisfaction and activity-monitor-measured sleep duration. A priori, we did not anticipate major shifts in sleep duration as the trial was only 7 weeks long, which is often too short for improvements.

Methods

Users

We recruited users with sleep complaints but no diagnosed sleep disorder. Users were informed that they would be given a sleep and activity monitor (i.e., the Jawbone UP Move) for participating in the study. Inclusion criteria included: 1) significant complaints with their sleep; 2) a smartphone (i.e,. Android or iPhone) to be used to gather self-tracking data via the app Paco; and 3) no plans to travel during the 7 weeks of the study. Exclusion criteria included: 1) diagnosed sleep disorder; 2) co-sleeping with someone else in the same bed/bedroom; and 3) disruptive and uncontrollable sleep schedules, such as night shift workers.

In total, 27 users (14 male, 13 female) were enrolled. Unfortunately, despite random assignment, distribution of ages was not balanced (see Table 1). While the majority of users were students (N=19), there was an imbalance in students vs. non-students (i.e., 8 non-students in SH, 6 in SH+SBT and 5 in SH+SBT+JIT conditions). Participants' survey responses to questions on perceived difficulty with sleep habits, importance of good sleep, and belief their sleep must be fixed (on a 7-point Likert scale) during session 1 indicated high motivation to improve sleep, as shown in Table 2.

Age range	SH		SH-SBT		SH-SBT-JIT	
	Male	Female	Male	Female	Male	Female
18-20	1	0	0	1	2	0
21-29	4	3	3	2	0	2
30-39	0	1	1	1	2	0
40-49	0	0	0	1	0	1
50-59	0	0	0	0	1	1

Table 1. Users' age distribution.

	SH	SH-SBT	SH-SBT-JIT
Difficulty with sleep habits	5.1 (.9)	5.6 (1.2)	5 (1.6)
Importance of good sleep	6.2 (.8)	6.1 (.8)	6.2 (1)
Belief sleep must be fixed	6.1 (.8)	5.8 (.8)	5.9 (1.2)

Table 2. Participants' motivation (1-strongly disagree to 7 strongly agree; Mean (SD).

Protocol & study design

All users met with research personnel for 5 sessions (see Table 3), received sleep hygiene information, and used the self-tracking strategies (see measures).

Session 1	Session 2	Session 3	Session 4	Session 5
Self-tracking tools setup	Initial creation	First revision	Second revision	Wrapping up

Table 3. Study procedure.

The SH group developed a behavioral plan for improving sleep with only information about sleep hygiene. During each subsequent section, they were asked to report what they observed based on self-tracking and then asked to change their plans as appropriate, again without any additional support. This was done to create the sort of unstructured self-experimentation common by QS'ers, while controlling for in-person interactions and education. The SH+SBT and SH+SBT+JIT conditions were trained in self-experimentation for behavior change based on the interactive instructional materials described earlier. Both of these self-experimentation groups were provided with worksheets to generate goals and ideas (see Figure 3).

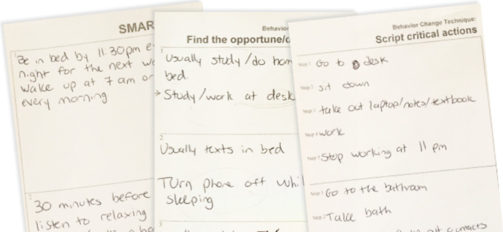


Figure 3. Users' worksheets.

For users in the SH+SBT+JIT condition, the interactive instruction was customized with content specifically aimed at JIT interventions. This group was taught the concept of JIT interventions and context-aware computing, and received JIT examples of behavior-change techniques. For each example, we used a slideshow to describe the targeted behavior, the behavior-change technique used, and a series of images showing how the technique could be implemented in a real-world context via our JIT tool. Users were asked to generate their ideas of the rules for implementation within the tool using sticky notes (Figure 4). We introduced this format to reinforce the rules-based logic and to enable easy rule changes via post-it note movement.

After each session, all participants were asked to write up their plan in an email and to send it to themselves as a reminder during the 2-week testing period. The SH group was asked to describe 'Things to do', and the other two conditions sent 'Goals' and 'Plans'. For the SH+SBT+JIT group, initial application descriptions were written by the researcher, and revised by users. The researcher developed JIT tool applications and installed them at users' homes. Participants were told that they could check past self-tracked data whenever they found it necessary.

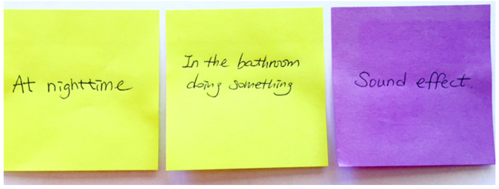


Figure 4. Example ideation of JIT applications.

Sessions 3, 4, & 5 started with a review of self-tracking data. They completed survey questions about normality of daily life, sleep, and plan implementation. Then they reported verbally with related cues (e.g., "How was your sleep?"' "Tell me how you carried out your plans?").

Measures

Our primary outcome measure was the PSQI [5]. The PSQI is a well-validated and extensively used measure of overall sleep quality that is used in a wide range of sleep research. The PSQI items (i.e., 7 components including subjective sleep quality, sleep latency, sleep duration, sleep efficiency, sleep disturbance, use of sleep medication, daytime dysfunction) correspond to the behavioral changes that our interventions and SH should influence. All users completed this questionnaire each session.

We also collected daily self-tracked sleep quality via a daily sleep diary and the wristband activity and sleep tracker, UP Move made by Jawbone (www.jawbone.com). We consider these secondary outcomes as the PSQI is a well regarded measure within the sleep research community. Further, as is common within behavioral evaluations, the self-tracked data was conceptualized as part of the intervention whereas the PSQI remained separate for a clean evaluation.

The sleep diary, which is used in sleep studies, included four questions asked every morning: 1) when did you go to bed? 2) how long did it take for you to fall asleep? 3) when did you wake up?, and 4) how satisfied were you with your sleep (rate from 1 to 10, higher scores are better)? Users installed PACO (www.pacoapp.com), which triggers a reminder inviting a user to answer the survey.

The Up Move collects a variety of data including total sleep time. As we did not intend to assess users' physical activity level, use of the activity tracking function during the day was optional but not required. All users were asked to download the Jawbone app to enable syncing.

All sessions were video recorded and user-generated materials were collected including worksheets, JIT app ideation using sticky notes, and notes of behavioral plans.

Analysis

We used Bayesian analysis instead of frequentist analysis, because it offers a more principled way to handle the uncertainty in small formative trials [19,25]. Our primary outcome was sleep quality as measured via the PSQI. Results from our secondary outcomes (daily sleep satisfaction and duration) can be found in our online supplement. We estimated the difference between the baseline and the phase 3 (final phase) outcomes for each measure. For the PSQI model, we used the estimated difference from baseline (measured at session 2) to phase 3 (measured at session 5) for each user as the response. All models included fixed effects for condition: the control (SH) and both intervention conditions (SH+SBT and SH+SBT+JIT).

We constructed Bayesian models using robust mixed effects linear regression.[1] Inspired by Howard et al. [23], for each outcome we ran three Bayesian analyses: an *uninformed* analysis, a *skeptical* analysis, and an *optimistic* analysis. These analyses differ only in the priors we set. The *uninformed* model uses *uninformed* priors, which produce estimates similar to frequentist models, which enables incorporation of beliefs postulated by a frequentist approach into our final averaged estimate. The *skeptical* priors are centered at zero (i.e. they assume it is most likely the treatment has no effect) and roughly cover the range of plausible effect sizes (e.g. reductions in PSQI of 4 or more are rarely seen for interventions like this; effects this size and larger have very low probability in our skeptical prior). Our *optimistic* priors assume the intervention will have a clinically meaningful effect around the size seen in the literature (i.e. a reduction of about 2 points on the PSQI) *but not larger*; thus, even this prior gives low probability to a reduction in PSQI of 4 points or more. The skeptical and optimistic priors were set by one of the authors who has expertise in sleep (see online supplement for full details).

As a final step, we average all three Bayesian models, weighted by the WAIC of each model (the Widely-Applicable Information Criterion, an estimate of out-of-sample prediction error). Models that are better at predicting the data out-of-sample are weighted higher. After McElreath [43], this approach acknowledges the uncertainty we have in our models (including our prior beliefs), and is applicable to small-sample studies.

We also report frequentist results to establish a comparator to statistics commonly reported. Our frequentist models used linear regression/ mixed effects linear regression[2] to examine the within-group changes over time and the between-group differences in our three target outcomes.

Outcome Results

Descriptive trends, prior to any statistical analyses of the PSQI, are visualized in Figure 5. These descriptives suggest that all groups reduced their PSQI scores (lower scores are better on the PSQI). These within-group changes for all conditions were confirmed in all of our models including our final averaged model.

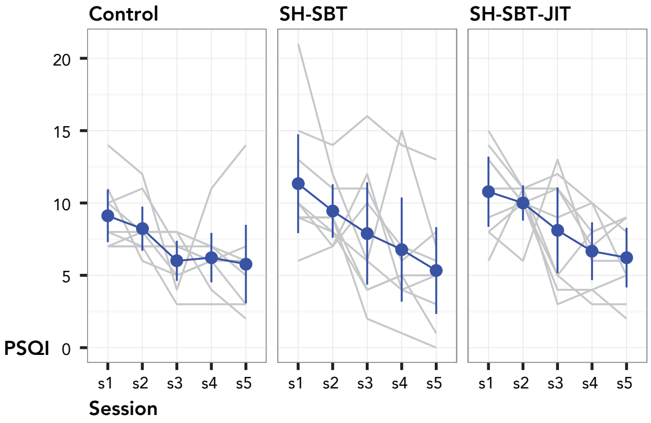


Figure 5. PSQI changes in each group. Each grey line shows one user. The blue line shows the mean for each session with 95% CI. S2 is our true baseline as measurement only occurred between s1 and s2 to help individuals generate more realistic PSQI estimates during s2.

The frequentist models suggest significant within-group changes for all conditions as indicated by the 95% confidence intervals not including 0 in the left column of Figure 6 (equivalent to $p < 0.05$). The central effect size estimates are approximately SH=-2, SH+SBT+JIT=-3, & SH+SBT= -4. These indicate clinically (and likely unrealistically) large effect sizes, which suggests that these models may be subject to a *magnitude error*, the tendency of small trials to overestimate effect sizes in frequentist analysis [20]. Results of the between-group comparisons based on the frequentist analyses revealed no significant differences between groups, but there are wide confidence intervals.

The plausibility of different effect sizes based on different prior beliefs are shown as probability distributions in light blue above the credibility intervals in Figure 6 (taller indicates increased probability that the effect is that size, such as -2). The final averaged Bayesian model suggests a very high probability (over 99% chance) that all conditions produced a change from baseline to phase 3 that was less than 0 (lower is better for PSQI). Our aggregate model estimates a 35% chance of reduction of 2 or more and 80% chance of reduction of 1 or more in the control, suggesting a moderate effect. Further, there was a high probability of a clinically significant effect for the SH+SBT and SH+SBT+JIT conditions (a 95% and 92% respective chance of reduction in PSQI of 2 or more; left column of Figure 6). There is also a good chance that there was some greater reduction in PSQI for both interventions than in the control (~90% chance that the between-group effects are < 0; right column) and an OK chance that those additional reductions are meaningful (~75% chance the differences are < -1) but a low chance that the effects are large (~35% chance that the differences are <-2). Finally, the estimated difference in PSQI of SH+SBT+JIT – SH+SBT is 0.05 with a wide 95% credibility interval (-1.73 to 2.45). There is a 73% chance that the difference in reduction between the two is 1 or less (i.e., that they have outcomes of similar clinical significance).

[1] *brms* package in R [4], which uses the Stan modeling language [7], and a Student *t* error distribution in place of a Normal error distribution for robustness [33].

[2] *lme4* package in R [2]

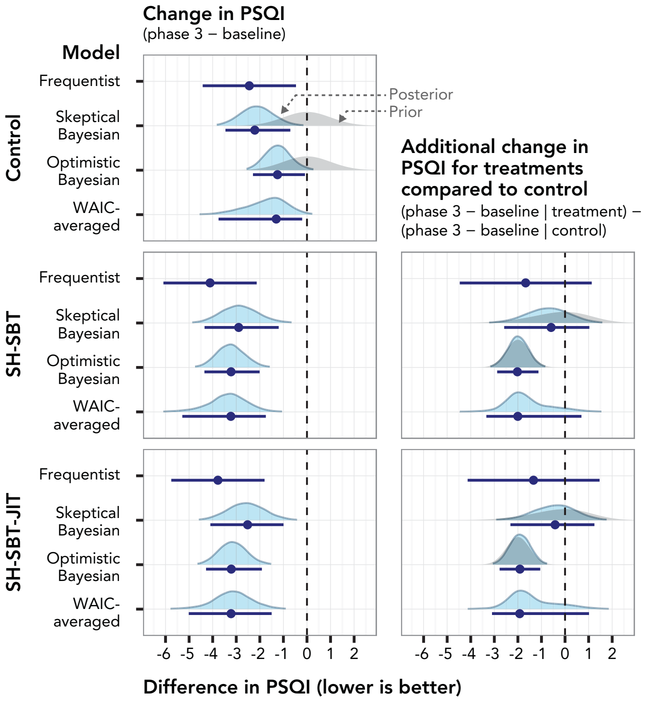


Figure 6. Results from our PSQI analyses with mode and 95% confidence intervals (95% quantile credibility intervals for Bayesian models). Prior and posterior probability distributions are shown where applicable. The uninformed Bayesian model (omitted) has estimates similar to the frequentist. The WAIC-averaged model is weighted 26% on the uninformed, 21% on the skeptical, & 53% on the optimistic model.

Process Results

Behavioral plan quality

Across conditions, users selected similar goals based on sleep hygiene suggestions. Three goals were particularly popular, 'Adjusting/sticking to a sleep schedule' (16 users), 'Doing relaxing routines near bedtime' (15 users) and 'Physical activity' (12 users). We asked all participants to rate how well they achieved their goals on a 0 to 10 scale, and the result shown in Table 4 indicates improvement.

	SH	SH-SBT	SH-SBT-JIT
Session 5	6.4 (2.9)	6.8 (1.8)	8.4 (1.2)
Session 5–Session 3	.8 (1.2)	2.9 (2.4)	1.9 (2.0)

Table 4. Goal achievement (1 to 10 with 10 indicating perfectly meeting the goal, Mean (SD)).

Participants in the interventions were also asked to rate the perceived quality of plans on 7-point scales: 1) Overall, I am satisfied with my goal; 2) The plan fits my lifestyle well; and 3) The plan will be essential for me to achieve my goal. Results summarized in Table 5 indicate both improvement and near maxed out scale responses.

		SH-SBT	SH-SBT-JIT
Satisfying	Session 4	5.9 (1.3)	6 (1.5)
	Session 4-Session 2	.8 (1.0)	0.0 (.5)
Fitting	Session 4	6.0 (1.0)	6.4 (1.1)
	Session 4-Session 2	.3 (1.1)	.5 (1.1)
Essential	Session 4	6.3 (.9)	6.3 (1.2)
	Session 4-Session 2	.9 (1.3)	.6 (1.1)

Table 5. Evaluation on created plans (1-strongly disagree to 7 strongly agree; Mean (SD)).

One key difference between the interventions and control, which is in line with our goals, was that the behavioral plans within the SH-SBT and SH-SBT-JIT conditions were more specific. For example:

P28 (SH), Session 2
Go to bed between 10 and 11 pm, and wake up between 5:30 and 7 am. In the morning, don't go back to bed. Don't take a nap; Avoid stimulants near bedtime.
P08 (SH-SBT), Session 2
Goals: Go to bed at 11-11:30 pm
Wake up at 8:30 am
Plans: Set up calendar reminder at 10:30 pm to get ready for bed; Turn off electronics at that time; Start getting ready for bed; Relax before sleep; Don't use electronics before bed; Keep computer on desk; Reward: if I go to bed on time for a week, go shopping
P09 (SH-SBT-JIT), Session 2,
Goals: No phone use near bedtime. 10 PM.
Open curtains in morning. Keep room light.
Plans: Remove chargers earlier than 10 PM, and transport all devices and chargers into another room, and finish up any tasks related to the computer within the guest room, and do not take back any of the devices to your room; When waking up at 8:00 AM, as a first activity, walk to curtains and open them to ensure a steady flow of natural light.

This pattern continued as users made revisions in Sessions 3 and 4. In the SH control, users did not typically change plans but, instead, tried different sleep hygiene strategies. Specifically, in session 3, 6 users added 1 or 2 new goals and 4 removed 1 or 2 goals. In session 4, 3 users added 1 new goal and 4 users removed 1 goal. In contrast, in the SH+SBT & SH+SBT+JIT conditions, users **did not change** their targeted sleep hygiene behaviors/goals as often. In the SH-SBT condition, during session 3, 3 users added 1 new goal and 2 users removed 1. In session 4, 2 added 1 new goal and 1 user removed 1 goal. Instead, users in the intervention conditions (SH-SBT and SH-SBT-JIT) made **changes to their behavioral plans** including adding or removing behavior-change techniques or modifying and further personalizing behavior-change techniques, such as adding elements that had been poorly defined. For instance, P08 (SH-SBT) specified items to avoid near bedtime (TV, movies, phone) in Session 3, which was only labeled '*elec-*

tronics' in session 2. In Session 3, 4 of the SH condition users did not make any changes in their plans, while 2 of the SH-SBT and 1 of the SH-SBT-JIT did not. In Session 4, 5 users in the SH condition did not change their plans, while 2 users in the SH-SBT and all of the SH-SBT-JIT changed their plans.

Overall, individuals across conditions that created more realistic, specified, and personalized plans had greater sleep improvements, though our sample is too small for firm conclusions. For example, P37 (SH) had more specific solutions compared to others in the SH group. In the revisions, she gradually modified her solutions to be more realistic and personalized (initial 6:15 AM wake up time changed to 6:45AM and 7AM; initial warm bath with soothing music changed to reading or writing journal). Her PSQI score improved by 5 (baseline=8, phase 3=3). In contrast, P10 realized the need to define activities during the nighttime in Session 3 and added '*Make a relaxing bedtime routine'*, but with no further details. By session 4, she reported that her bedtime routine was not more relaxing and her PSQI worsened (baseline=12, phase 3=14).

JIT-specific results

Most participants found JIT support beneficial, with two general themes. First, the system reminded participants of goals. It helped to stop preoccupying activity, which has been continuing longer than necessary: P19, 'I usually spent long time using computer [sic]', P31, 'Yeah, like if I was distracted, playing video game or working on the homework, it was nice to get that text message…and then I realize it's late [sic]…' Second, it inspired positive emotions. For example, P13 (who had the music play when she came home after work, which was designed to remind her about prep for the next day) stated, 'not necessarily about snack/lunch prep. Now you're are at home… now [I am] relaxed[sic]', P03, 'I really liked the music when I open the closet, and on Friday mornings. Though I failed in reaching the exercise goal, it was just fun, good to hear. [sic]' Despite this positive perception of the JIT support, our quantitative PSQI results indicate a low likelihood that our JIT component improved sleep beyond our tutorial.

One possible explanation may have been a misalignment between triggers and plans. Sometimes, no triggers were created for a goal. While most users created triggers for the majority of their targets, P33 included a trigger for only 1 among his 4 targets, and P24, for 2 targets among her 5/7 (depending on session). Users with the greatest sleep quality improvements in the SH+SBT+JIT group appeared to have better alignment between their plans and triggers. For example, P24, who made only minimal sleep quality improvement, only incorporated application responses for one target behavior among six. For waking up, she designed her application to play peaceful music at 5:45 AM and switch to loud rock music at 6AM if she did not awake. For her other behavioral targets including drinking water, no working in bed, increased exercise, relaxation near bedtime, and no liquid after 9 PM, she did not create any triggers. In contrast, P27, who did have improved sleep quality, created triggers for most of his goals. For exercise in the morning, he designed his application to play music when he entered the kitchen and to help him eat smaller meals, he made the application play music when he entered the kitchen after work. For no coffee after 4pm, he created a SMS reminder at 4PM. The only behavior he did not create a trigger for was going to bed between 10:30 and 11:30PM.

A second explanation could be under-utilization of context-aware computing. Users mostly developed time-based triggers, such as P15 sending himself an SMS at 11AM on Sunday, saying "Meal plan." Action-based conditions that involve sensors (e.g., when opening the refrigerator between 7 and 8 PM, play sound to invite preparation of snack/lunch for tomorrow) were limited including 3 users (P15, P31, and P33) who used only time-based triggers. The participants that used action-based triggers, in general, had improved sleep quality. For example, P09 had a targeted goal of establishing a bedtime routine and, as part of that, developed a series of triggers focused on supporting the routine. The first trigger at 9 PM was a text message saying "charge the devices." At 10 PM if the smartphone was not being changed, music would play in the bedroom. If the phone was charged on time (meaning plugged in prior to 10 PM for 3 nights in a row) AND when a person opened a box of candy THEN happy music would play. If the phone had not been charged and the box opened then sad music played. The user also created a trigger at 8 AM to play happy music to invite them to open the blinds.

A third plausible explanation for limited response was some participants engaging in only limited iteration on their JIT interventions. For example, P27 added a trigger to not drink coffee after 4 PM in session 3 after realizing that was important but not specified in session 2. Existing triggers were also modified. For example, P13 added a trigger to an existing one to support going to bed. Initially, only music played in the living room at 9:45 PM, but in session 3, another piece of music played in the bedroom 10 minutes later. Those individuals that engaged in these small tweaks appeared to benefit most from the JIT intervention, thus suggesting the need for more explicit support in iteration.

DISCUSSION

Results indicate that all three self-experimentation strategies, including our unstructured control, appeared to improve sleep quality over 7 weeks, with the high likelihood that our two interventions resulted in a small to moderate improvement in sleep quality relative to the control. Further, results indicated success with goals in all conditions with greater improvements in goal achievement in the interventions relative to control. Results also indicated near maxed out scales of positive perceptions of plans created in the two intervention conditions by the final session. Taken together, these results suggest the value of our structured self-experimentation approaches for creating goals and plans relative to an unstructured self-experimentation control, which was designed to mimic self-experimentation commonly done by QS'ers. Our results suggest the value of our tutorials for supporting self-creation of personalized behavior-change plans. Others should consider building on our tutorials if there is a need to support users in the self-creation of personalized behavioral goals and plans (see online supplement).

Our current work placed greatest emphasis on the first part of the self-experimentation for behavior change framework; namely formulating a hypothesis about how to change a behavior. While our approach did include self-tracking and a systematic way to review those data to iterate on behavioral plans, we did not include a formal n-of-1 trial evaluation of the plans. This was done intentionally as our personal experience teaching behavior change in our courses is that individuals struggle with what to change, which others have studied [13,24] and devising a viable personalized plan on how to change and sustain the behavior. While technically an n-of-1 trial could have been incorporated during the 2 weeks in between sessions, we explicitly excluded n-of-1 trials because we wanted participants to formulate a robust hypothesis and not get overly focused on "the answer" from the trial over other information. Indeed, the subtle changes made during each iteration were exactly what we sought. If we had run a longer trial, we would likely have incorporated n-of-1 trials (e.g., [28]) after the week 5 or 7 mark, depending on how confident each participant was in their goals and plans. We contend that only at that point would a person have enough experience to translate an under-specified hunch about a behavioral plan (i.e., what they often start with) into a well-reasoned hypothesis (i.e., something appropriate for an n-of-1 trial). Future work should flesh out when to use self-experimentation for discovery [13,24], self-experimentation for behavior change without n-of-1 trials (what we did) and self-experimentation for behavior change that includes n-of-1 trials.

While our JIT intervention did produce improved outcomes relative to control, results indicated no significant advantage beyond our self-experimentation tutorial. With that said, careful examination of our results suggests several targets for future work. First, if notifications can be sent during JIT states, then individuals appear to appreciate them, thus suggesting the potential for this type of intervention. It appeared that many of the issues arose from the wide variety exhibited in how individuals programmed their own JIT interventions. Specifically, individuals that clearly specified JIT plans for each goal, took full advantage of context-aware computing (i.e., did not merely use time but also used the sensors for defining rules), and iterated to improve their systems, exhibited the greatest improvements. Future work could build on the basic self-experimentation protocol we created (see online supplement) but add greater support for and emphasis of these three points.

Beyond self-experimentation, our work also provides clear guidance for other HCI researchers interested in using Bayesian analyses when evaluating their systems. As illustrated in our discussion, the Bayesian estimates give us a richer—and appropriately more conservative—view of the systems by allowing us to quantify the probability of clinically significant effect sizes. This more nuanced information can better support decision-making of others. In particular, we argue that the probability of small to medium effects justifies continued design work, including development of these protocols within automated online systems and better support for individuals to create more complex JIT state trigger rules. These conclusions would have been hard to draw if using purely frequentist approaches, which, from a between-group perspective, would have indicated no effect because of the wide confidence intervals. It is our hope that this case study could be used as a starting template for others interested in using Bayesian analyses in formative evaluations; to help, as Kay et al. put it [25], "free design and engineering researchers from the shackles of meaningless *p*-values in small-*n* studies."

CONCLUSION

Our results indicate the value of our tutorials for helping individuals generate personalized behavioral goals and plans for achieving said goals over a more unstructured form of self-experimentation. Results from our JIT intervention suggests that these systems work particularly well if JIT plans are created for each behavioral goal, if context-aware computing is fully used for inferring JIT states, and if individuals are empowered to iterate. Beyond this, we have also proposed a unique framework for self-experimentation for behavior change that is complementary to previous work [24]. Our Bayesian analyses could also be used as a starting template for others interested in using Bayesian analyses in their work, with full details available within our online supplement. Future work should explore how to help individuals devise realistic, specific, and personalized behavioral plans; help individuals use more rigorous n-of-1 trials in self-experimentation (including the possibility of a combined self-experimentation paradigm that includes both Karkar's self-experimentation for discovery and our self-experimentation for behavior change paradigms along with careful timing on when NOT to use n-of-1 trials); and provide individuals with better training on the concepts of opportunity and receptivity for helping individuals in designing their own JIT triggers via context-aware computing.

ACKNOWLEDGEMENTS

This work was supported by a Google Research Faculty Award (PI: Hekler). Thanks to Bob Evans from Google for being our sponsor, for support on the early conceptualization of this work, and for providing the tool, Paco, for our use.

REFERENCES

1. Thomas Attin and E. Hornecker. 2005. Tooth brushing and oral health: how frequently and when should tooth brushing be performed?. Oral health & preventive dentistry, 3, 3: 135-40.
2. Douglas Bates, Martin Maechler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67, 1: 1-48.
3. Frank Bentley, Konrad Tollmar, Peter Stephenson, Laura Levy, Brian Jones, Scott Robertson, Ed Price, Richard Catrambone, and Jeff Wilson. 2013. Health Mashups: Presenting Statistical Patterns between Wellbeing Data and Context in Natural Language to Promote Behavior Change. ACM Trans. Comput.-Hum. Interact. 20, 5, Article 30, 27 pages.
4. Paul-Christian Buerkner. brms: An R Package for Bayesian Multilevel Models using Stan. Journal of Statistical Software (in press).
5. Daniel J. Buysse, Charles F. Reynolds, Timothy H. Monk, Susan R. Berman, and David J. Kupfer. 1989. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry research 28, 2: 193-213.
6. Taj Campbell, Brian Ngo, and James Fogarty. 2008. Game design principles in everyday fitness applications. In Proceedings of the 2008 ACM conference on Computer supported cooperative work (CSCW '08), 249-252.
7. Bob Carpenter, Andrew Gelman, Matt Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Michael A. Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2016. Stan: A probabilistic programming language. Journal of Statistical Software (in press).
8. Jeannette S. Chin, Victor Callaghan, and Graham Clarke. 2006. An end-user programming paradigm for pervasive computing applications. In 2006 ACS/IEEE International Conference on Pervasive Services, 325-328.
9. Eun Kyoung Choe, Nicole B. Lee, Bongshin Lee, Wanda Pratt, and Julie A. Kientz. 2014. Understanding quantified-selfers' practices in collecting and exploring personal data. In Proceedings of the 32nd annual ACM conference on Human factors in computing systems (CHI '14), 1143-1152.
10. Sunny Consolvo, David W. McDonald, Tammy Toscos, Mike Y. Chen, Jon Froehlich, Beverly Harrison, Predrag Klasnja, Anthony LaMarca, Louis LeGrand, Ryan Libby, Ian Smith, and James A. Landay. 2008. Activity sensing in the wild: a field trial of ubifit garden. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08), 1797-1806.
11. Allen Cypher and Daniel Conrad Halbert. Watch what I do: programming by demonstration. MIT press, 1993.
12. Siddharth Dalal, Majd Alwan, Reza Seifrafi, Steve Kell, and Donald Brown. 2005. A rule-based approach to the analysis of elders activity data: Detection of health and possible emergency conditions. In AAAI Fall 2005 Symposium, pp. 2545-2552.
13. Nediyana Daskalova, Danaë Metaxa-Kakavouli, Adrienne Tran, Nicole Nugent, Julie Boergers, John McGeary, and Jeff Huang. 2016. SleepCoacher: A Personalized Automated Self-Experimentation System for Sleep Recommendations. In Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16). ACM, New York, NY, USA, 347-358.
14. Anind K. Dey, Timothy Sohn, Sara Streng, and Justin Kodama. 2006. iCAP: Interactive prototyping of context-aware applications. In International Conference on Pervasive Computing, 254-271.
15. Brian J. Fogg. 2002. Persuasive Technology: Using Computers to Change What We Think and Do. Morgan Kaufmann.
16. Brian J. Fogg. 2009. A behavior model for persuasive design. In Proceedings of the 4th international Conference on Persuasive Technology, 40.
17. Martha M. Funnell, Tammy L. Brown, Belinda P. Childs, Linda B. Haas, Gwen M. Hosey, Brian Jensen, Melinda Maryniuk et al. 2009. National standards for diabetes self-management education. Diabetes care 32, no. Supplement 1: S87-S94.
18. Manuel García-Herranz, Pablo A. Haya, and Xavier Alamán. 2010. Towards a Ubiquitous End-User Programming System for Smart Spaces. J. UCS 16, 12: 1633-1649.
19. Andrew Gelman and David Weakliem. 2009. Of beauty, sex and power: Too little attention has been paid to the statistical challenges in estimating small effects. American Scientist 97, 4: 310-316.
20. Andrew Gelman and John Carlin. 2014. Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors. Perspectives on Psychological Science 9, 6: 641–651.
21. Eric B. Hekler, Predrag Klasnja, Jon E. Froehlich, and Matthew P. Buman. 2013. Mind the theoretical gap: interpreting, using, and developing behavioral theory in HCI research. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13), 3307-3316.
22. Eric B. Hekler, Susan Michie, Misha Pavel, Daniel E. Rivera, Linda M. Collins, Holly B. Jimison, Claire Garnett, Skye Parral, Donna Spruijt-Metz. 2016. Ad-

vancing Models and Theories for Digital Behavior Change Interventions. American Journal of Preventive Medicine, 51,5: 825-832.

23. George S. Howard, Scott E. Maxwell, and Kevin J. Fleming. 2000. The proof of the pudding: An illustration of the relative strengths of null hypothesis, meta-analysis, and Bayesian analysis. Psychological methods 5, 3: 315–332.

24. Ravi Karkar, Jasmine Zia, Roger Vilardaga, Sonali R. Mishra, James Fogarty, Sean A. Munson, and Julie A. Kientz. 2015. A framework for self-experimentation in personalized health. Journal of the American Medical Informatics Association: ocv150.

25. Matthew Kay, Gregory L. Nelson, and Eric B. Hekler. 2016. Researcher-Centered Design of Statistics: Why Bayesian Statistics Better Fit the Culture and Incentives of HCI. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16), 4521-4532.

26. Abby C. King, Eric B. Hekler, Lauren A. Grieco, Sandra J. Winter, Jylana L. Sheats, Matthew P. Buman, Banny Banerjee, Thomas N. Robinson, and Jesse Cirimele. 2013. Harnessing different motivational frames via mobile phones to promote daily physical activity and reduce sedentary behavior in aging adults. PloS one 8, 4: e62613.

27. Vladimir J. Konečni. 2008. Does music induce emotion? A theoretical and methodological analysis. Psychology of Aesthetics, Creativity, and the Arts 2, 2: 115.

28. Richard L. Kravitz, Naihua Duan, eds, and the DEcIDE Methods Center N-of-1 Guidance Panel. Design and Implementation of N-of-1 Trials: A User's Guide. AHRQ Publication No. 13(14)-EHC122-EF. Rockville, MD: Agency for Healthcare Research and Quality; February 2014. www.effectivehealthcare.ahrq.gov/N-1-Trials.cfm.

29. Patricia Lacks and Monique Rotert. Knowledge and practice of sleep hygiene techniques in insomniacs and good sleepers. Behaviour research and therapy 24, 3: 365-368.

30. Gary P. Latham. 2003. Goal Setting: A Five-Step Approach to Behavior Change. Organizational Dynamics 32, 3: 309-318.

31. Jisoo Lee, Erin Walker, Winslow Burleson, and Eric B. Hekler. 2014. Exploring users' creation of personalized behavioral plans. In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication (UbiComp '14 Adjunct), 703-706.

32. Jisoo Lee, Erin Walker, Winslow Burleson, and Eric B. Hekler. 2015. Understanding Users' Creation of Behavior Change Plans with Theory-Based Support. In Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '15), 2301-2306.

33. John K Kruschke. 2013. Bayesian estimation supersedes the t test. Journal of Experimental Psychology: General 142, 2: 573–603. http://doi.org/10.1037/a0029146

34. Min Kyung Lee, Sara Kiesler, and Jodi Forlizzi. 2011. Mining behavioral economics to design persuasive technology for healthy choices. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11), 325-334.

35. Ian Li, Anind K. Dey, and Jodi Forlizzi. 2011. Understanding my data, myself: supporting self-reflection with ubicomp technologies. In Proceedings of the 13th international conference on Ubiquitous computing (UbiComp '11), 405-414.

36. Ian Li. 2011. Personal Informatics and Context: Using Context to Reveal Factors that Affect Behavior. Ph.D Dissertation. Carnegie Mellon University, Pittsburgh, Pennsylvania..

37. Henry Lieberman, Fabio Paternò, Markus Klann, and Volker Wulf. 2006. End-user development: An emerging paradigm. In End user development, 1-8.

38. James J. Lin, Lena Mamykina, Silvia Lindtner, Gregory Delajoux, and Henry B. Strub. 2006. Fish'n'Steps: Encouraging physical activity with an interactive computer game. In International Conference on Ubiquitous Computing, 261-278.

39. Helen Lindner, David Menzies, Jill Kelly, Sonya Taylor, and Marianne Shearer. 2003. Coaching for behaviour change in chronic disease: a review of the literature and the implications for coaching as a self-management intervention. Australian Journal of Primary Health 9, 3: 177-185.

40. Steven R. Livingstone, Ralf Mühlberger, Andrew R. Brown, and Andrew Loch. 2007. Controlling musical emotionality: An affective computational architecture for influencing musical emotions. Digital Creativity 18, 1: 43-53.

41. Edwin A. Locke and Gary P. Latham. 2002. Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. American psychologist 57, 9: 705.

42. Lena Mamykina, Elizabeth Mynatt, Patricia Davidson, and Daniel Greenblatt. 2008. MAHI: investigation of social scaffolding for reflective thinking in diabetes management. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 477-486.

43. Richard McElreath. 2016. Statistical rethinking: A Bayesian course with examples in R and Stan. Vol. 122. CRC Press.

44. Susan Michie, Maartje M. van Stralen, and Robert West. 2011. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implementation Science 6, 1: 1.

45. Susan Michie, Michelle Richardson, Marie Johnston, Charles Abraham, Jill Francis, Wendy Hardeman, Martin P. Eccles, James Cane, and Caroline E. Wood. 2013. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Annals of behavioral medicine 46, 1: 81-95.

46. Susan Michie, Lou Atkins, & Robert West. 2014. The Behaviour Change Wheel: A Guide To Designing Interventions. Silverback Publishing.

47. Mark Muraven and Roy F. Baumeister. 2000. Self-regulation and depletion of limited resources: Does self-control resemble a muscle?. Psychological bulletin 126, 2: 247.

48. Inbal Nahum-Shani, Eric B. Hekler, and Donna Spruijt-Metz. 2015. Building health behavior models to guide the development of just-in-time adaptive interventions: A pragmatic framework. Health Psychology 34, S: 1209.

49. Jason Nawyn, Stephen S. Intille, and Kent Larson. 2006. Embedding behavior modification strategies into a consumer electronic device: a case study. In Proceedings of the 8th international conference on Ubiquitous Computing (UbiComp'06), 297-314.

50. Miriam E. Nelson, W. Jack Rejeski, Steven N. Blair, Pamela W. Duncan, James O. Judge, Abby C. King, Carol A. Macera, and Carmen Castaneda-Sceppa. 2007. Physical activity and public health in older adults: recommendation from the American College of Sports Medicine and the American Heart Association. Circulation 116, 9: 1094.

51. Virpi Oksman and Pirjo Rautiainen. 2003. "Perhaps it is a Body Part": How the Mobile Phone Became an Organic Part of the Everyday Lives of Finnish Children and Teenagers. Machines that become us: The social context of communication technology: 293-308.

52. Benjamin Poppinga, Wilko Heuten, and Susanne Boll. 2014. Sensor-based identification of opportune moments for triggering notifications. IEEE Pervasive Computing 13, 1: 22-29.

53. Mashfiqui Rabbi, Min Hane Aung, Mi Zhang, and Tanzeem Choudhury. 2015. MyBehavior: automatic personalized health feedback from user behaviors and preferences using smartphones. In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15), 707-718.

54. Paschal Sheeran. 2002. Intention—behavior relations: A conceptual and empirical review. European review of social psychology 12, 1: 1-36.

55. Anna Ståhl, Kristina Höök, Martin Svensson, Alex S. Taylor, and Marco Combetto. 2009. Experiencing the affective diary. Personal and Ubiquitous Computing 13, 5: 365-378.

56. Edward J. Stepanski and James K. Wyatt. 2003. Use of sleep hygiene in the treatment of insomnia. Sleep medicine reviews 7, 3: 215-225.

57. Tammy Toscos, Anne Faber, Shunying An, and Mona Praful Gandhi. 2006. Chick clique: persuasive technology to motivate teenage girls to exercise. In CHI '06 Extended Abstracts on Human Factors in Computing Systems (CHI EA '06), 1873-1878.

58. Khai N. Truong, Elaine M. Huang, and Gregory D. Abowd. 2004. CAMP: A magnetic poetry interface for end-user programming of capture applications for the home. In International Conference on Ubiquitous Computing, 143-160.

59. Van Dantzig, Saskia, Gijs Geleijnse, and Aart Tijmen van Halteren. 2013. Toward a persuasive mobile application to reduce sedentary behavior.Personal and ubiquitous computing 17, 6: 1237-1246.

60. Fang Xu, Machell Town, Lina S. Balluz, William P. Bartoli, Wilmon Murphy, Pranesh P. Chowdhury, William S. Garvin et al. 2013. Surveillance for certain health behaviors among States and selected local areas—United States, 2010. MMWR Surveill Summ 62, 1: 1-247.