Investigating Contextual Notifications to Drive Self-Monitoring in mHealth Apps for Weight Maintenance

Yu-Peng Chen
Department of CISE, University of Florida
Gainesville, Florida, United States
yupengchen@ufl.edu

Julia Woodward
Department of Computer Science and Engineering, University of South Florida
Tampa, Florida, United States
juliaevewoodward@usf.edu

Dinank Bista
Department of CISE, University of Florida
Gainesville, Florida, United States
dbista@ufl.edu

Xuanpu Zhang
Department of CISE, University of Florida
Gainesville, Florida, United States
xzhang13@ufl.edu

Ishvina Singh
Department of CISE, University of Florida
Gainesville, Florida, United States
ishvina.singh@ufl.edu

Oluwatomisin Obajemu
Department of CISE, University of Florida
Gainesville, Florida, United States
o.obajemu@ufl.edu

Meena N. Shankar
Department of Clinical and Health Psychology, University of Florida
Gainesville, Florida, United States
mns@phhp.ufl.edu

Kathryn M. Ross
Department of Clinical and Health Psychology, University of Florida
Gainesville, Florida, United States
kmross@phhp.ufl.edu

Jaime Ruiz
Department of CISE, University of Florida
Gainesville, Florida, United States
jaime.ruiz@ufl.edu

Lisa Anthony
Department of CISE, University of Florida
Gainesville, Florida, United States
lanthony@cise.ufl.edu

ABSTRACT

Mobile health applications for weight maintenance offer self-monitoring as a tool to empower users to achieve health goals (e.g., losing weight); yet maintaining consistent self-monitoring over time proves challenging for users. These apps use push notifications to help increase users' app engagement and reduce long-term attrition, but they are often ignored by users due to appearing at inopportune moments. Therefore, we analyzed whether delivering push notifications based on time alone or also considering user context (e.g., current activity) affected users' engagement in a weight maintenance app, in a 4-week in-the-wild study with 30 participants. We found no difference in participants' overall (across the day) self-monitoring frequency between the two conditions, but in the context-based condition, participants responded faster and more frequently to notifications, and logged their data more timely (as eating/exercising occurs). Our work informs the design of notifications in weight maintenance apps to improve their efficacy in promoting self-monitoring.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**; *Smartphones*; *Empirical studies in ubiquitous and mobile computing*; • **Applied computing → Consumer health**.

KEYWORDS

Mobile Health, Notifications, Health Behavior Change

ACM Reference Format:
Yu-Peng Chen, Julia Woodward, Dinank Bista, Xuanpu Zhang, Ishvina Singh, Oluwatomisin Obajemu, Meena N. Shankar, Kathryn M. Ross, Jaime Ruiz, and Lisa Anthony. 2024. Investigating Contextual Notifications to Drive Self-Monitoring in mHealth Apps for Weight Maintenance. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA*. ACM, New York, NY, USA, 21 pages. https://doi.org/10.1145/3613904.3641993

1 INTRODUCTION

Mobile health (mHealth) applications (apps) can promote behavioral changes toward healthier habits by assisting in users' self-monitoring process. Prior work has demonstrated that a system that facilitates users' self-monitoring and provides feedback can

CHI '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0330-0/24/05
https://doi.org/10.1145/3613904.3641993

effectively increase physical activity [64] and motivate healthy eating [21]. Therefore, self-logging is a critical component of many well-being systems. Encouraging users to log personalized health data (e.g., food and exercise) not only aids in their process of achieving health behavior change, as described above, but also allows the system to provide tailored feedback and health services. However, self-logging is often difficult to sustain at regular intervals over many weeks [10, 11].

To address this challenge, we revisited behavior change theories and turned to the Fogg Behavior Model [28], which highlights the role of triggers in boosting engagement and adherence. Prior work has shown that mHealth app notifications increase users' app engagement [67], adherence to health goals [35], and long-term retention [79]. For instance, Bentley and Tollmar [6] found that including mHealth app notifications increased users' logging of dietary intake data from 12% to 63%. However, prior work has also shown that mHealth notifications are often ignored by users, either due to appearing at inopportune moments [16, 27, 57] or leading to habituation (i.e., decreased response to repeated stimulation) over time [5, 38, 83]. To design effective notifications, extensive research has studied when to ideally interrupt users, but mainly in areas other than mHealth. These previous studies have shown that sending notifications based on contextual factors, such as location, time, and task engagement level, reduces perceived disruption, decreases response time, and increases click rate [27, 59, 74, 76, 84].

To increase mHealth app engagement for promoting health behavior change, prior work explored using context-based notification systems to reach users when they are at a high level of receptivity [22, 50, 67, 68]. Nonetheless, the results of these efforts have produced a mixture of conclusions, contributing to a landscape of both successes and ambiguities. Although some studies have demonstrated the efficacy of contextual factors in enhancing notification effectiveness and user engagement [36, 50], other investigations have not yielded statistically significant results in support of the same concept [40]. For example, Horsch et al. [40] found no effect of notification type (context-based vs. time-based reminders) on user engagement of a sleep diary app over a one-week study. Overall, it is unclear if sending mHealth notifications based on contextual factors adds additional benefit, especially long-term. Therefore, we focus on investigating how the timing of notifications (context-based vs. time-based) influences logging frequency and time required for users to respond to notifications.

Contextual factors, such as time of day, have been found to be effective in estimating user interruptibility. Prior work also found that ideal moments for interruption occur at transitions between different physical activities, mainly because that moment represents when a self-initiated task interruption occurs [44, 70, 71]. Therefore, we selected time of day and physical activity transitions as our contextual factors for sending context-based notifications. We designed a within-subjects study employing notification timing as the independent variable with two conditions: time-based notifications and context-based notifications. In the time-based condition, our app sends notifications based only on the user's fixed time preferences, whereas in the context-based condition, the app sends notifications based on both the user's time preferences and their physical activity transitions. To simulate a real-world use case, we selected weight maintenance as a health goal for participants to focus on. Informed by our team of health experts, we selected weights, dietary intake (lunch and dinner), physical activities, and self-rated measures of weight-related variables (e.g., hunger) as self-monitoring data points.

To investigate our research question, we developed an instrumented companion app with custom functionality to pair with FatSecret [23], a commercially popular weight management app that offers an API [24], logging interfaces, and comprehensive databases for tracking weight-related data. We conducted an in-the-wild study with 30 participants recruited from our local community or cities within the same time zone over 4 weeks. We counterbalanced the order effect by dividing 15 participants into a context-first group and the remaining 15 into a time-first group. Our results showed that while the average overall log rate was greater in the context-based condition (58.87%) than in the time-based condition (55.54%), there was no significant effect of notification timing on the overall log rate. This suggests that contextual notifications did not necessarily trigger more daily self-monitoring behavior. However, we found a significant effect of notification timing on notification click response time, click rate, and completion rate. Participants (1) clicked on notifications faster (12.33 minutes vs. 18.42 minutes), (2) clicked on notifications more frequently (19.05% vs. 13.96%), and (3) logged more health data (21.77% vs. 17.32%) within 60 minutes after the notification was sent in the context-based condition than in the time-based condition. Our results suggest that context-based notifications were effective in triggering more timely responses to notifications and more timely self-monitoring behaviors. Additionally, insights gathered from interviews highlighted individual differences in behavioral preferences that introduced variability into our overall self-monitoring behavior measure.

The contributions of our paper are:

- integration of insights from interruptibility literature into an mHealth context, using an app that was designed with health experts' guidance,
- a month-long in-the-wild study in which participants used our app in their daily lives deployed on their own devices,
- a deeper understanding of how using contextual cues for sending notifications can trigger more **timely** self-monitoring behaviors,
- a set of implications for future research on finding the right balance between real-world individual differences and optimal health-related guidelines, and
- a discussion of how our results are informative for future design of mHealth apps.

2 RELATED WORK

We focus our review of relevant prior work on four major categories: (1) reasons for using mHealth apps to support health behavior change, (2) health intervention strategies leveraging mobile devices, (3) context-based notifications in mHealth apps, and (4) opportune moments to interrupt users and send notifications.

2.1 Mobile Devices for Health Behavior Change

The World Health Organization defines Mobile Health (mHealth) as the use of mobile and wireless technologies to support the achievement of health objectives [101]. Mobile devices such as smartphones and commercial wearables are rapidly becoming an important platform for the delivery of behavior change interventions to promote healthy lifestyles [91]. The widespread adoption of mobile devices provides a scalable platform that can be used by a large share of the population [78]. By overcoming geographical and temporal barriers, these mobile devices enabled a more convenient delivery of healthcare services at a lower cost compared to traditional healthcare services [91]. Additionally, the ubiquitous and ever-present nature of mobile devices allows for continuous collection of an individual's contextual data and provides opportunities to deliver health interventions at the right place and time [49]. Unlike desktop computers or laptops, mobile devices, especially smartphones, are nearly always close to the person who owns them [19]. Previous work on smartphone proximity found that individuals were within arm's reach of their smartphones 53% of the time and within the same room as their phones 88% of the time [19]. Furthermore, mobile devices including smartphones and smartwatches have become personal and intimate objects [46, 98]. These devices are usually used for various activities in the user's daily routines, including reading emails, social networking, financial tasks, and entertainment. As a result, they often contain highly personal information such as pictures, text messages, and financial information [46, 98]. This personal relationship can increase acceptance and usage of mHealth apps, which in turn facilitate the delivery of health interventions in the user's daily life [49].

2.2 Health Intervention Strategies Leveraging Mobile Devices

With their ubiquitous nature, mobile platforms offer an opportunity for providing behavior change guidance in the user's daily life. According to behavior change theories, setting a challenging yet attainable goal motivates an individual to achieve behavior change [56]. Additionally, Bandura's social cognitive theory (SCT) [2] asserts that self-efficacy is a primary determinant of behavior change, representing an individual's confidence in their capability to perform a specific behavior within a given context. SCT underscores the pivotal role of repeated successes in bolstering self-efficacy, as previous accomplishments significantly shape one's self-perception [2]. Leveraging the personal information collected by a mobile device, mHealth apps have been used to suggest adaptive health goals to promote physical activity. Zhou et al. [107] built *CalFit*, an app that uses a reinforcement learning-based (RL-based) algorithm to suggest adaptive daily step goals based on the user's past goal achievements and step counts. A 10-week study with 64 participants demonstrated the effectiveness of the adaptive goals in promoting physical activity. Similarly, Miyake et al. [64] developed the *StepUp Forecast* app, which provides a prediction of the user's daily and hourly steps, and showed the effectiveness of presenting the predictions in increasing the user's self-efficacy and step count in a 5-week study with 36 participants. Moreover, Rabbi et al. [81] designed *MyBehavior*, an app that recognizes physical activity based on sensor data and generates contextualized, actionable, and low-effort recommendations. The authors conducted a 3-week pilot study with 17 participants and found that participants who received *MyBehavior*'s personalized suggestions walked significantly more. Lastly, Bracken and Waite [9] completed an online survey with 112 *MyFitnessPal* users to investigate the relationship between their self-efficacy for healthy eating (SE-HE) and their achievement of nutrition goals. The authors found that higher SE-HE and app usage were significantly related to greater reported goal achievement. This finding indicates that an individual's belief in their ability to eat healthy is a major predictor of nutrition goal achievement, which aligns with SCT [2].

In addition to goal-setting, self-monitoring is another essential behavior change technique that has been widely used in mHealth apps. Self-monitoring increases the user's awareness of their progress and provides an opportunity for them to adjust their strategy [4]. For example, self-monitoring of weight, dietary intake, and physical activity is a key strategy for weight management in individuals with obesity [12]. Daily tracking of weight and weight-related factors empowers individuals to evaluate their progress toward both immediate and long-term goals, while also increasing their awareness of connections between specific behaviors and weight, assisting in future goal setting [47]. Research has shown that individuals who engage in more frequent self-monitoring lose more weight during weight loss programs and regain less weight after program completion [10, 12]. Compared to more traditional paper-and-pencil techniques, mHealth apps offer advantages for self-monitoring, such as real-time feedback on goal progress and visualizations. Using mHealth apps to self-monitor may also be more socially acceptable. For example, using a smartphone to access nutrition information or log food/drinks consumed may be less stigmatizing compared to the use of a traditional paper log or a calorie reference book [12]. Popular ways of tracking progress include automated logging of sensor data and a journaling mechanism for users to report data on their own [34, 55]. For instance, the *Fish'n'Steps* app [55] utilized a pedometer for automated step count tracking, encouraged users to set daily step goals, and provided progress feedback.

mHealth apps have been utilizing both goal setting and self-monitoring as essential techniques to empower users to achieve health goals in various health fields, including physical activity promotion [53, 81], weight loss [29, 48], nutrition recommendation [31, 103], smoking cessation [89], alcohol use [1], stress reduction [26, 43], sleep management [17, 62] and chronic disease self-management [33, 104]. However, without personalized and actionable guidance, the app's effect on health behavior change has been shown to be limited [56, 65]. Thus, an important direction for mHealth apps to promote health behavior change involves collaboration with health experts/interventionists.

2.3 Context-Based Notifications for mHealth

Through sending push notifications (e.g., as reminders for health-related tasks), mHealth apps encourage users to self-monitor to achieve their goals of behavioral change. App notifications increase users' engagement with the apps and their adherence to health objectives [6, 8, 35, 40, 75]. Bidargaddi et al. [8] analyzed mHealth app engagement over 89 days with 1,255 participants and found that sending a notification with a tailored health message resulted in more app interaction within the next 24 hours than not sending notifications. Patrick et al. [75] investigated the effect of text-based interventions on helping users lose weight over 4 months. The participants who received text-based tailored notifications 2 to 5 times a day instead of printed material lost more weight. Although notifications can aid in adherence and app engagement, users' responsiveness and interest decreased as they received more notifications, especially if the content was similar [6, 44, 87, 88]. Additionally, abandonment of health technologies prior to achieving one's goals is still common [14, 30, 37, 51].

To increase notification responsiveness and mHealth app engagement, prior work has explored using context-based notification systems to reach users when they are at a high level of receptivity [22, 50, 67, 68]. Previous studies have used a range of contextual factors such as time, physical activity, calendar, and weather [36, 92, 99, 100]. For example, A-CHESS is an existing just-in-time adaptive intervention (JITAI) that provides location-based notifications to support recovery from alcohol abuse [36]. When a user approaches a high-risk location (e.g., bar), A-CHESS sends a notification alert to the user. Based on a study with 349 participants over 8 months, the participants who had access to traditional counseling plus A-CHESS (170 participants), compared to just traditional counseling, reported significantly fewer risky drinking days. Künzler et al. [50] examined factors affecting users' receptivity towards JITAIs. The authors conducted a study with 189 participants over a period of 6 weeks, in which participants received notifications from a chatbot-based digital coach, *Ally*, to improve their physical activity levels. They found that intrinsic factors, such as personality and age, had an effect on responsiveness, as well as contextual factors (e.g., time, activity). For example, participants were more likely to answer prompts delivered between 10 am to 6 pm, and while walking instead of being still.

However, some studies have not found evidence that context-based notifications provide additional benefits for user engagement in mHealth apps. Morrison et al. [66] investigated different timings of push notifications for a mobile stress-management app. The notifications were sent either occasionally (not every day), daily, or based on context (i.e., location, movement, and time of day). The study had a total of 77 participants who used the app for an average of 3 days. The authors found that participants in the "occasional" group responded significantly less often and there was no difference between the daily and context-based groups. Horsch et al. [40] examined the effect of context-based and time-based reminders on adherence to a sleep diary app with relaxation exercises. Participants received context-based reminders for one week, time-based for another week, and then no reminders for one week. In the context-based group, reminders were sent when participants were in one location for an hour, just ended a phone call, or were using another app. The authors found no difference in completed exercises and diary entries between the time-based and context-based groups, only a significant difference between reminders and no reminders.

Overall, it is unclear if context-based notifications are more beneficial for mHealth app engagement and adherence than time-based notifications, especially long-term. In these previous studies, the participants only interacted with the apps for approximately 3 to 7 days, which is a relatively short time to gauge app engagement and to allow for user behaviors to emerge. Therefore, in our study, we compare context-based and time-based notifications over 4 weeks.

2.4 Interruptibility

In addition to analyzing when to send notifications based on users' context (e.g., time, location), prior work has investigated the *interruptibility* of users in different domains. A key problem with the design of push notifications in software is that they often arrive at inopportune moments for the user [16, 44, 57]. Poorly timed notifications negatively impact users' ongoing tasks, attention, focus, and cognitive load, and can even result in users deleting the app [25]. To reduce the cost of ill-timed interruptions and increase user engagement with the suggested content, previous studies have attempted many different methods of determining the ideal interruption time (e.g., detecting users' physical activity transition to estimate less disruptive times [71]) [13, 60, 80, 96, 105]. These studies have found that contextual factors such as users' engagement with ongoing tasks and types of current activity can be used to infer their interruptibility [27, 41, 60, 77]. Mehrotra et al. [60] created an app called *My Phone and Me* and conducted a study in which 20 participants were prompted to report on their interactions with notifications for two months. The authors found that users' perceived disruption increased as the complexity of an ongoing task rose. Other contextual factors that have been found to be effective in estimating user interruptibility include time of the day and location [80, 90, 105].

Previous studies have also found that ideal moments for interruption occur at transitions between different physical activities (i.e., when a self-initiated task interruption occurs) [44, 70, 72]. Okoshi et al. [70, 71] created the *Attelia* notification management system that detects breakpoints in users' daily lives in real-time and defers notifications until such moments. The results of their in-the-wild study showed that Attelia was able to reduce users' cognitive load by 33% and their response time by 13% compared to sending notifications at random times. In collaboration with Yahoo! JAPAN, Okoshi et al. [74] conducted a large-scale in-the-wild user study with more than 680,000 participants by deploying their context-aware system to the Yahoo! JAPAN Android app. The authors showed that deferring notification delivery until a breakpoint inferred from users' physical activity transitions resulted in faster response time, increased click rate, and improved user engagement.

These previous studies have shown promising results in estimating the user's interruptible moments based on contextual factors. However, these prior studies were not conducted in the mHealth domain. Therefore, we wanted to understand how applying the results from previous interruptibility studies in daily life to the context of mHealth can contribute to users' responsiveness and engagement with mHealth app notifications.

3 METHOD

3.1 Intervention Design

To investigate how the timing of notifications affects users' engagement in an mHealth app, we designed a 4-week within-subjects study that utilizes notification timing as the independent variable with two conditions: time-based notifications and context-based notifications. Informed by previous research, we selected time of the day and physical activity transitions as the two contextual factors for sending context-based notifications. Time-based notifications were sent based only on the user's time preferences at fixed times, while context-based notifications were sent based on both the user's

Table 1: Five types of self-monitoring data points.

Data Type	Self-Monitoring Frequency
Weight	Daily
Lunch	Daily
Dinner	Daily
Activity	Daily
Questionnaire (Self-rated measures)	Twice every week

time preferences and physical activity transitions we detected with our app.

To simulate a real-world use case, we selected weight maintenance as a potential health goal for the user. Previous research has demonstrated the effectiveness of involving experienced domain experts in the design process of mHealth studies [26, 69, 106]. Thus, our team included a faculty member in clinical psychology with 17 years of experience in obesity treatment and a registered dietitian with 21 years of research experience. We consulted our health experts for selecting (1) key self-monitoring data points required for weight maintenance, and (2) the timing of our notifications. Informed by our health experts and prior work [85], we selected *weight*, *dietary intake* (lunch and dinner), *physical activities*, and *self-rated measures* of weight-related variables (e.g., hunger) as self-monitoring data points (Table 1). We did not include breakfast due to its proximity to the recommended weighing time, and to prevent an excessive number of notifications.

The self-rated measures were delivered through one weekly and one end-of-week check-in questionnaires; all other data points were daily. Notifications for the weekly questionnaire were sent on a random day between Monday and Saturday, and those for the end-of-week questionnaire were always sent on Sunday. For the other four types of health data (i.e., weight, lunch, dinner, and activity), a daily notification was sent if the data was still missing by the scheduled notification time. Thus, participants received at most five notifications in a day, which aligns with suggestions on notification frequency from previous research [67]. Over a 4-week study period, for each participant, we designed a total of 2 (questionnaires) $\times$ 4 (weeks) + 4 (data types) $\times$ 28 (days) = 120 opportunities for sending notifications (60 for each condition), each associated with an opportunity for logging one of the five types of health data.

3.2 App Design

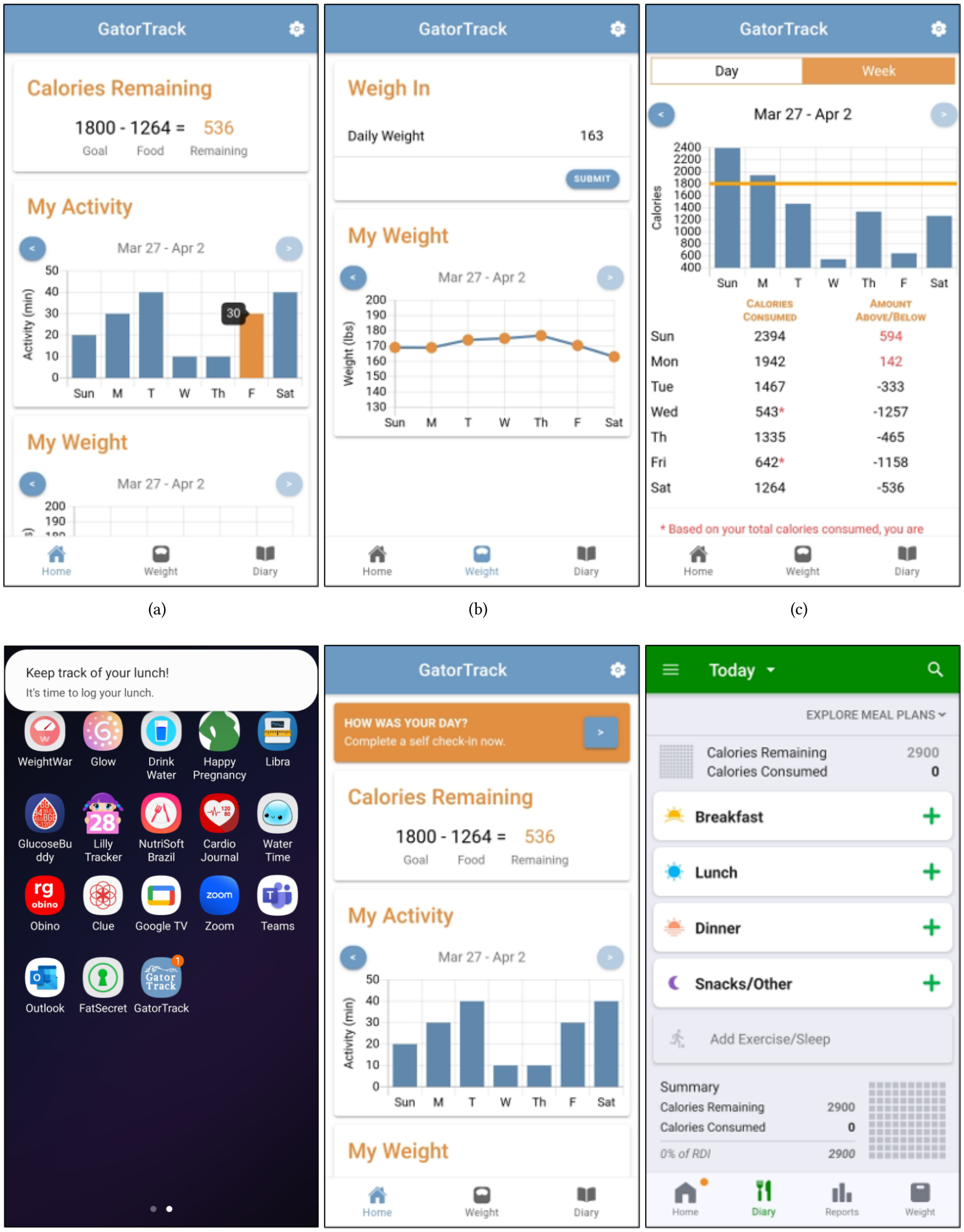
Investigating our research question requires an instrumented app with custom functionality because no commercial app supported all of our requirements (e.g., self-rated measures delivered through the weekly questionnaires). However, building an app from scratch can be resource-intensive, especially when our main focus is on assessing the impact of our notification design, as opposed to undertaking a comprehensive app development project. Hence, we designed and developed a companion app, GatorTrack, through a human-centered iterative design process that included guidance from our health experts. The app integrated Google's Activity Recognition Transition API [18] to detect physical activity transitions for sending context-based notifications. This companion app was created to pair with FatSecret [23], a commercially popular weight management app that offers an API [24], logging interfaces, and comprehensive databases for tracking food and exercise. Figure 1 shows the resulting app we developed for this study. Participants logged daily weights (Figure 1-b) and answered questionnaires through GatorTrack. At the bottom of the app, we implemented three tabs for easy navigation. The "Diary" tab directs the user to the FatSecret app for recording dietary intake and physical activities.

3.3 Notification Design

As mentioned in Section 3.1, for each participant, there were 120 opportunities for sending a notification, each associated with a notification record (i.e., a log of the notification and its result) for one of the five types of health data (i.e., weight, lunch, dinner, activity, and questionnaire). In this subsection, we describe how we (1) formulated a notification record and (2) designed the notification timings.

Existing research has shown that user responses to notifications often involve multiple stages [13, 96]. Thus, our notification record design aims to capture (1) the user's immediate reaction to the notification and (2) whether the user acted on the target task. Figure 2 shows the logical flow of our notification records based on user behavior in the app. Each notification record contains three binary flags (True/False) indicating the status of the record: *sent*, *clicked*, and *completed*. The *sent* flag indicates if the notification was sent. A notification would not be sent if its corresponding health data had already been logged before the scheduled notification time or transition (Figure 2-a). The *clicked* flag indicates if the notification was **clicked** within 60 minutes after being sent. The *completed* flag indicates if the notification's corresponding health data was **logged** within 60 minutes after the notification was sent. When the *clicked* flag was set to *True*, the *sent* flag must also be *True*, but the *completed* flag could be either *True* or *False*, meaning that a user might have clicked on a notification without logging the corresponding health data. When the *completed* flag was set to *True*, the *sent* flag must also be *True*, but the *clicked* flag could be either *True* or *False*, meaning that a user might have logged the data without clicking on the notification. For each flag, if its value is *True*, the notification record will also contain its corresponding timestamp.

Next, we collaborated with our health experts to establish notification timing based on users' daily routines in both conditions. User profiling has been shown to enhance the capacity of mHealth apps to offer personalized services, aiding users in attaining their health objectives [34, 103]. Additionally, previous studies have found that customization in mHealth apps can increase users' engagement



(a) (b) (c)

(d) (e) (f)

Figure 1: Screenshots of our weight maintenance system: (a) Home screen of the GatorTrack app, including a balance equation for displaying the daily calorie goal, a bar chart for displaying weekly activity in minutes, and a line graph for displaying weekly weight summary in pounds. (b) The weigh-in screen for recording the user's daily weight. (c) The screen for displaying daily and weekly dietary intake details. (d) An example of the push notification. (e) An example of the in-app notification. We chose an orange background to complement the main gray-blue color theme to create an effective visual effect. (f) A screenshot of the FatSecret app. Participants logged their food and exercise in this commercial app.

through empowering user autonomy [93]. Thus, instead of allowing users to set timers for each of the five health data types, we implemented an interface for users to set their personal schedule for

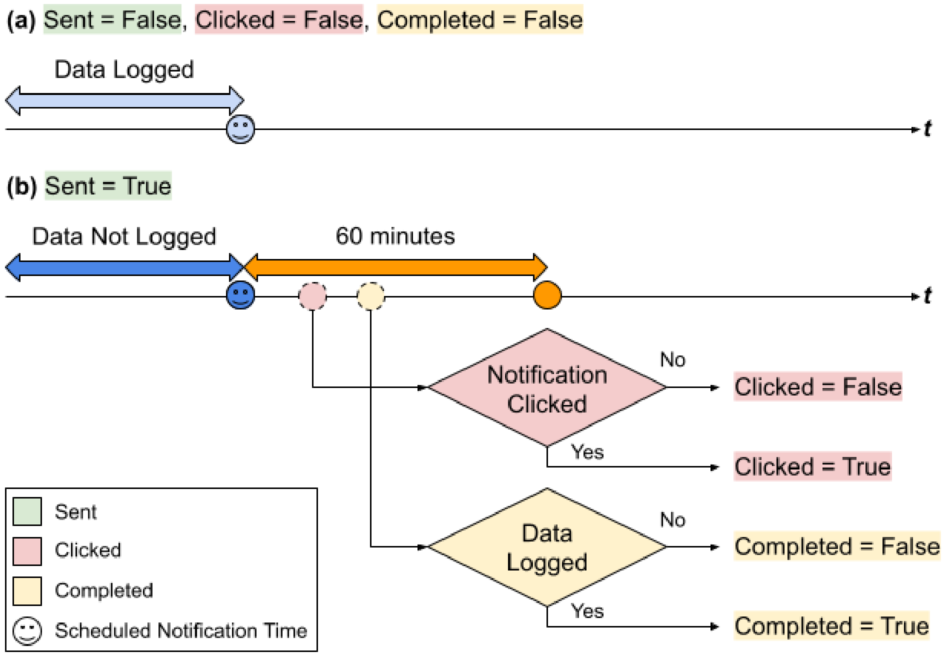


Figure 2: Flags of a notification record. (a) The notification was not sent if its corresponding health data had been logged before the scheduled time. In this case, both *clicked* and *completed* could only be *False*. (b) In the case where the notification was sent, within 60 minutes, *clicked* was set to *True* if the notification was clicked and *completed* was set to *True* if the corresponding data was logged.

Table 2: Notification time settings in the time-based condition.

Data Type	Notification Time
Weight	Wake-up time + 10 mins
Lunch	Lunch time + 1 hr
Dinner	Dinner time + 1 hr
Activity	Dinner time + 1.5 hrs
Questionnaire	Bedtime - 1 hr

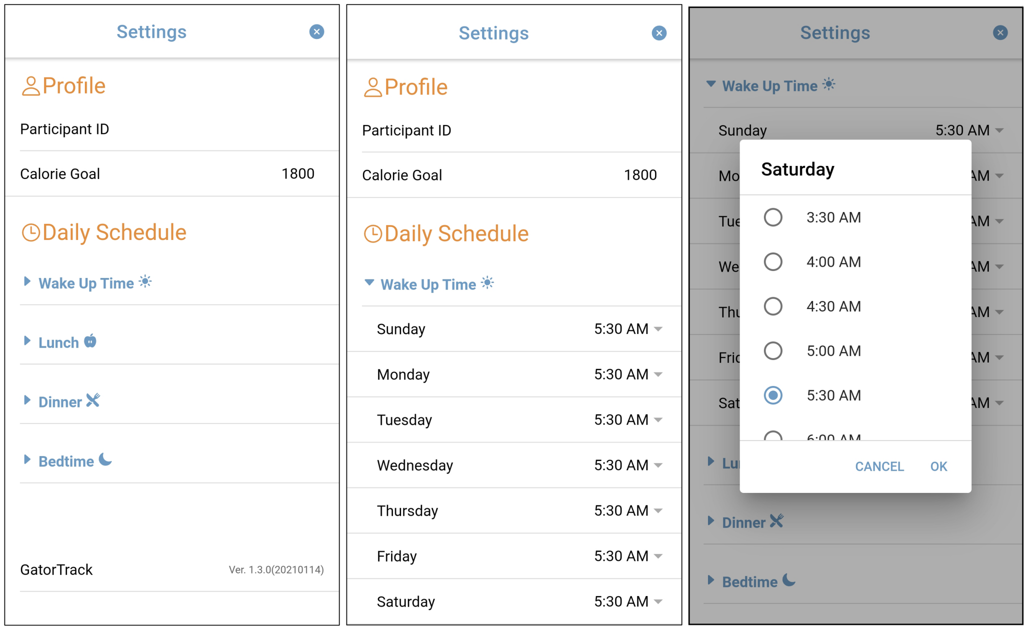
wake-up time, lunch time, dinner time, and bedtime (Figure 3). As shown in Table 2, timing of weight notifications is decided based on wake-up time since our health experts suggested that fasting morning weights are more accurate. Timing of activity notifications is decided based on dinner time. More specifically, our health experts suggested that activity notifications should be sent after dinner time because of local weather conditions in the southeastern region of the United States and work hours. Timing of the questionnaire notifications corresponds to bedtime since the questionnaire was intended to be a reflection at the end of the day.

In the time-based condition (Table 2), we shifted the corresponding user time preference for each type of data by a specific duration, based on input from our health experts, and configured GatorTrack to send notifications at those fixed times. For example, if the user set their wake-up time at 5:30 AM, the app will send a weight notification at 5:40 AM.

In the context-based condition (Table 3), we selected a specific time range for each type of data based on users' daily routine. As shown in Figure 4, upon entering the corresponding time range, the GatorTrack app started listening for transitions and sent a notification when the corresponding type of transition (Table 3) was detected. Thus, our context-based notifications took into account both contextual factors: time of day and physical activity transitions. If no transitions were detected by the end of the time range, a notification was sent, which we denoted as a "timeout" notification. Previous studies in the field of weight maintenance highlighted the importance of recording one's health data shortly after performing the target behavior (e.g., eating and/or exercising). Thus, notifications in both conditions were designed to stay in the notification panel for at most 60 minutes before disappearing, in order to measure the user's timely response to the notifications. The specific duration was decided based on our survey of prior work in the interruptibility field [60, 71, 80].

3.4 Testing Scenarios and Pilot Test

GatorTrack utilizes Google's Transition API [18] to detect transitions between five types of physical activity, including Still, Walking, Running, On Bicycle, and In Vehicle. To ensure that the API would be able to capture transitions fast and accurately enough for our research goal, prior to the study, three researchers conducted tests using a prototype app in 214 simulated scenarios that included six types of transitions: Still to Walking, Walking to Still, Running to Walking, Running to Still, Walking to In Vehicle, and In Vehicle



(a) (b) (c)

Figure 3: Screenshots of the "Settings" screen in GatorTrack: (a) Clicking on the "Gear" icon on the app home screen (Figure 1-a) directs the user to this "Settings" screen, which displays four categories for the user's daily schedule. (b) Clicking on one of the categories (e.g., "Wake Up Time") displays a drop-down menu that allows the user to set their time preferences for each day of the week. (c) Clicking on one of the days (e.g., "Saturday") displays a list of 48 time options, each separated by 30 minutes.

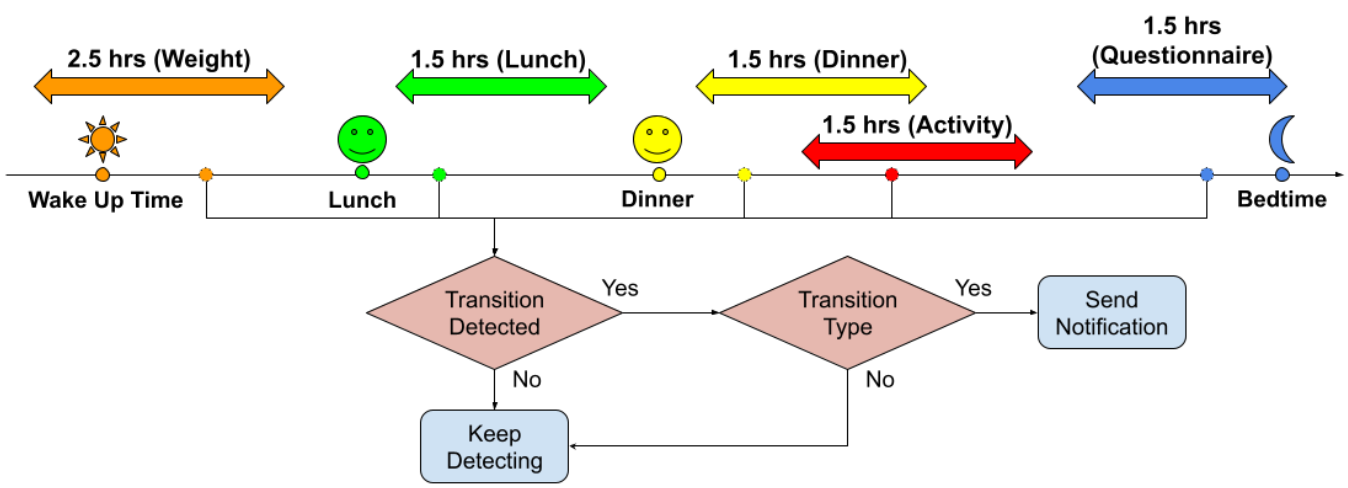


Figure 4: An illustration of the context-based notification scheme. Using the example provided in Figure 3, in which the "Wake Up" time is set at 5:30 AM on Saturday, and assuming the user has not logged their daily weight, in the time-based condition, GatorTrack will send a weight notification at 5:40 AM on Saturday, according to Table 2. In the context-based condition, GatorTrack will listen for transitions from 5:00 AM to 7:30 AM on Saturday, and send a weight notification if any type of transition is detected within this time range, according to Table 3.

to Walking. These scenarios were crafted through collaborative brainstorming sessions involving our HCI researchers and health experts, aiming to capture how individuals might interact with our app in their daily lives. As shown in Table 4, in these testing scenarios, we varied (1) the physical activity, (2) the duration, and (3) the position of the phone. For example, we used the following testing

Table 3: Notification time settings in the context-based condition.

Data Type	Detection Start Time	Detection End Time	Transition Type
Weight	Wake-up time - 30 mins	Wake-up time + 2 hrs	Any transition
Lunch	Lunch time + 30 mins	Lunch time + 2 hrs	Any movement to Still
Dinner	Dinner time + 30 mins	Dinner time + 2 hrs	Any movement to Still
Activity	Dinner time + 1 hr	Dinner time + 2.5 hrs	Any movement to Still
Questionnaire	Bedtime - 1.5 hrs	Bedtime	Any movement to Still

Table 4: Variables in our API testing scenarios.

Variables	Options
Physical Activity	Sit, Walk, Run, Drive
Duration	10s, 30s, 1m, 2m, 5m, 30m, 1hr
Phone Position	Being used, Held in the hand but not used, On the desk, In the pocket

scenario for the "Still to Walking" transition: "[user] **sits** at his desk for **1h** with his phone lying **on the desk** not used, and then he stands up and **walks** around while **using the phone** for **30s**." Over the course of one month, three researchers collectively tested the 214 scenarios one at a time, following the procedure of creating a log file locally on the testing device through the prototype app, performing the activity while timing themselves, and clicking on the notification if it was triggered by the target transition, which would update the log file with the detected timestamp. Our results from this preliminary testing showed that the API requires 56 seconds on average to register a transition with 92% accuracy (i.e., it failed to detect a transition in 17 total scenarios).

Following the full development and release of the GatorTrack app on Google Play, we conducted a two-week pilot study with 3 participants recruited via word-of-mouth to identify potential usability and technical issues. Based on the feedback obtained during the pilot study, we updated our onboarding materials to include a more comprehensive description of functionality offered by FatSecret. Regarding GatorTrack, the initial design allowed users to select daily schedule time options within the time frame of 6 AM to 11 PM, with intervals of 30 minutes. However, in the pilot study, we observed that one participant reported waking up as early as 4:30 AM to 5 AM, while another participant reported a bedtime of approximately 1 AM to 2 AM. Thus, we expanded the available time options to cover a full 24-hour cycle, thereby accommodating more diverse user schedules, as shown in Figure 3.

3.5 Recruitment and Participants

We recruited 30 participants who met our inclusion criteria from the local community and cities within the same time zone through word of mouth, university email lists, posted flyers, and in-person recruiting at local places such as libraries and museums. Our inclusion criteria were as follows: (1) being in the age range of 18 to 70 and (2) owning a Samsung Android phone. We exclusively recruited Samsung phone users because during our beta testing, the Activity Recognition Transition API [18] performed the best on Samsung phones compared with other Android phone brands. Additionally, a weighing scale was necessary for the study. Participants without a weighing scale received one at no cost and were allowed to keep it. There were 20 self-identified males and 10 self-identified females who ranged in age from 18 to 54 (M = 27.93; SD = 8.47). 27 participants lived in our local community, while 3 lived in other cities within the same time zone. 5 participants identified as White, 2 as Black or African American, 19 as Asian, 1 as Hispanic or Latino, 2 as Middle Eastern, and 1 preferred not to answer. 18 participants stated they had experience with mHealth apps, while 12 did not. We opted not to restrict our participant selection to individuals actively pursuing weight-related goals, as our primary focus was on investigating the comparative effectiveness of contextual notifications in triggering self-monitoring behaviors, regardless of whether participants had a predefined goal. We also chose not to impose restrictions on the use of other apps or travel, primarily to maintain the natural flow of users' daily lives. It's crucial to highlight that our implementation automatically adjusted timestamps based on the time zone, ensuring the accuracy of notification times even in situations involving travel across time zones.

3.6 Study Procedure

Our study included three Zoom meetings: the onboarding meeting, followed by two weeks of the study, the check-in meeting, another two weeks, and lastly, the final meeting (Figure 5). During the onboarding meeting, the participant provided electronic informed consent. We then created accounts in both GatorTrack and FatSecret for the participant and demonstrated a comprehensive walk-through of the two apps. We counterbalanced condition order by alternately assigning participants to either a context-first group or a time-first group. Participants were asked to record their weight, dietary intake, and physical activity every day, and self-rated measures of weight-related variables twice per week. Depending on the condition, participants received context-based or time-based notifications as reminders only through GatorTrack. We asked participants to disable all notifications from FatSecret. After two weeks, we conducted a mid-study check-in interview, in which we asked participants a set of semi-structured questions focusing on their preferences and usage of the apps and notifications, and addressed any potential technical issues they might have encountered. After

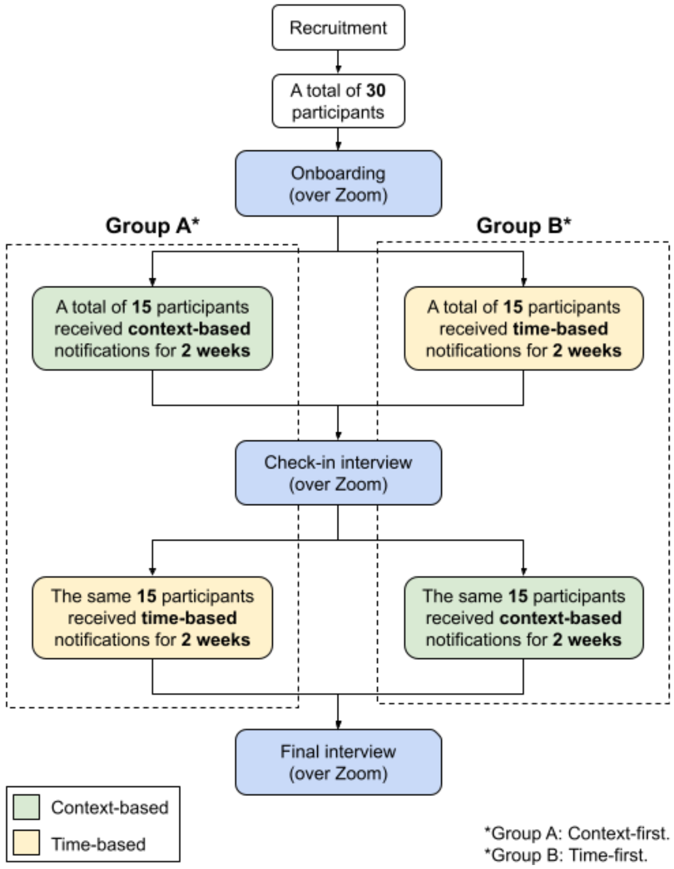


Figure 5: Study procedure. The order of conditions was counterbalanced across participants.

another two weeks, we concluded the study with a final interview, which included another set of semi-structured questions focusing on participants' overall experience and a debriefing in which we explained the difference between the two types of notifications they had received. Interested readers can find in the supplementary material our interview questions. For confidentiality, each participant was assigned a unique 4-digit ID number, ensuring that logged information remained disconnected from individual identities. While the Transition API required location services to be enabled on participants' smartphones over the course of the study, we neither accessed nor stored geographical coordinates. Participants received $20 per week as compensation if they continued to log at least one type of health data per week, totaling $80 upon study completion. Our study protocol was approved by our university's Institutional Review Board.

3.7 Collected Data

Before we conducted our data analysis, we ran several sanity checks on the collected data. We found that, for two participants (one in the context-based condition and the other in the time-based condition), we encountered a technical issue that resulted in one notification (out of 120) not being sent over the course of the study. Our app also sent additional weekly check-in questionnaires to 5 participants (2 in the context-based condition and 3 in the time-based condition), totaling 5 extra notifications, due to another technical issue. To mitigate these limitations in our data analysis, we normalized the log counts by the actual number of notifications sent.

Our study was conducted over 179 days between November 2022 and May 2023, with each of the 30 participants using our tracking system for 28 days. As described in Section 3.3, our system was designed to create 120 notification records per participant, with each corresponding to an opportunity for logging one of the five health data types (i.e., weight, lunch, dinner, activity, and questionnaire), totaling 3600 notification records. However, due to the technical issue described above, our system created a total of 3605 notification records, associated with 3605 daily logging opportunities. Among these 3605 logging opportunities, participants recorded a total of 2060 (57.14%) daily logs, with an average of 68.67 per participant (min = 22; max = 112; SD = 20.68). Regarding notification records, a total of 2922 (81.05%) notifications were *sent*, with an average of 97.4 sent notifications per participant (min = 44; max = 118; SD = 14.39). Out of these sent notifications, a total of 441 (15.09%) notifications were *clicked*, with an average of 14.7 clicked notifications per participant (min = 2; max = 55; SD = 10.03). A total of 514 (17.59%) notifications were *completed*, with an average of 17.13 completed notifications per participant (min = 3; max = 46;

SD = 10.78) (recall that a notification could be *completed* without being *clicked*).

3.8 Data Analysis

Our study was a within-subjects study employing one independent variable with two levels: *notification timing* (i.e., condition, time-based or context-based). We conducted quantitative analysis on four dependent variables, detailed in the next section, capturing our participants' engagement with the notifications from GatorTrack and self-monitoring features in both apps. We first ran a repeated measures one-way ANOVA on our data. We initially selected a parametric test because our dependent variables were continuous numeric values. However, since empirically sampled data often does not conform to the assumptions of normality, leading to the violation of prerequisites for parametric tests, we also ran the non-parametric Wilcoxon Signed-Rank test. We present the results for the repeated measures one-way ANOVA only if both the parametric and non-parametric tests showed the same patterns of significant difference between two conditions ($p < 0.05$). Additionally, to identify themes that could help illuminate the patterns we observed in the quantitative data, we transcribed the 60 interview recordings and analyzed the responses through affinity diagramming–a bottom-up, inductive method for organizing large-scale qualitative data [7]. We opted not to calculate inter-rater reliability, as it is not recommended when the research focus is on identifying concepts and themes [58]. Similar to prior work [102], to create the affinity diagram, five researchers broke down the transcripts into individual utterances and converted each of them into sticky notes using Miro [63], an online whiteboard tool for remote collaboration. Subsequently, we iteratively grouped a total of 818 sticky notes into themes. Initially, we identified 14 central categories, which we further combined into 7 main themes.

4 RESULTS

To investigate the influence of notification timing on user engagement with self-monitoring, we selected four dependent variables to measure user engagement: *overall log rate*, *notification click response time*, *notification click rate*, and *notification completion rate*. Table 5 shows the definitions and results of our repeated measures one-way ANOVA on the four dependent variables with a within-subjects factor of *notification timing* (i.e., condition). Figure 6 provides an overview of our results in box and whisker plots while Figure 7 demonstrates the raw data and effective size.

4.1 Overall Log Rate

To measure the user's daily self-monitoring behavior, we investigated the *overall log rate*. We define the *overall log rate* as the overall count of daily logs for each data type (i.e., weight, lunch, dinner, physical activity, and questionnaire) divided by the anticipated total number of daily logging events. We only counted one data log for each data type per day; for example, a participant who logs two items for lunch, would have the same performance measure as a participant who logs only one item for lunch, emphasizing the daily monitoring practice.

In general, in each condition, the target was to have 60 daily logging events per participant. However, as mentioned in Section 3.7, due to technical issues, we adjusted this count for certain participants. Specifically, for 5 participants who received an additional weekly-questionnaire notification, we increased the target count by one event in the corresponding condition, making it 61. For 2 other participants who did not receive one notification as a result of technical issues, we decreased the target count by one, bringing it down to 59.

In the time-based condition, the average overall log rate was 55.54% (min = 16.67%; max = 91.67%; SD = 17.85%). In the context-based condition, the average overall log rate was 58.87% (min = 16.39%; max = 95%; SD = 20.51%). The context-based condition had a greater average overall log rate than the time-based condition by 3.33% (95% CI: [-3.07, 9.73]). A repeated measures one-way ANOVA on *overall log rate* with a within-subjects factor of *notification timing* found no significant difference between the two conditions ($F_{1,29}$ = 1.1302, $p > 0.05$). This result suggests that being in the context-based condition had no significant effect on the overall log rate.

4.2 Notification Click Response Time

To investigate if sending context-based notifications can motivate the user to respond more promptly to these notifications, we ran analysis on the *notification click response time*. We define the *notification click response time* as the time elapsed between the time when the notification was clicked and the time when it was sent. This decision is motivated by prior research in the field of weight management, which emphasized the importance of promptly documenting one's health data following the execution the target behavior, such as eating or exercising. This measure can only be computed if the notification was clicked. We removed 4 participants (2 for each condition) from this analysis because they did not click on any notifications in the second half of their study.

In the time-based condition, the average notification click response time was 18.42 minutes (min = 3.29m; max = 46m; SD = 10.99m). In the context-based condition, the average notification click response time was 12.33 minutes (min = 2.47m; max = 36.25m; SD = 8.5m). The time-based condition had a greater average notification click response time than the context-based condition by 6.09 minutes (95% CI: [1.72, 10.47]). A repeated measures one-way ANOVA on *notification click response time* with a within-subjects factor of *notification timing* found a significant difference between the two conditions ($F_{1,25}$ = 8.2484, $p < 0.01$). This result suggests that context-based notifications were clicked within significantly less elapsed time than time-based notifications.

4.3 Notification Click Rate

To investigate the effect of context-based notifications on the frequency of user responses, we also ran analysis on the *notification click rate*. We define the *notification click rate* as the number of *clicked* notifications divided by the number of *sent* notifications, where a *clicked* notification indicates that the notification was clicked within 60 minutes after being sent.

In the time-based condition, the average notification click rate was 13.96% (min = 0%; max = 73.08%; SD = 14.01%). In the context-based condition, the average notification click rate was 19.05% (min = 0%; max = 61%; SD = 14.99%). The context-based condition had a greater average notification click rate than the time-based

Table 5: Definitions and results of our repeated measures one-way ANOVA on the four dependent variables with a within-subjects factor of notification timing. Starred (*) rows are significant at the p < .05 level; those with a + are significant at the p < .01 level.

Dependent Variable	Definition	p-value
Overall log rate	Overall count of daily logs / Total logging opportunities	0.2965
Notification click response time (m)	Notification *clicked* time - Notification *sent* time	0.0082+
Notification click rate	Number of *clicked* notifications / Number of *sent* notifications	0.0113*
Notification completion rate	Number of *completed* notifications / Number of *sent* notifications	0.0110*

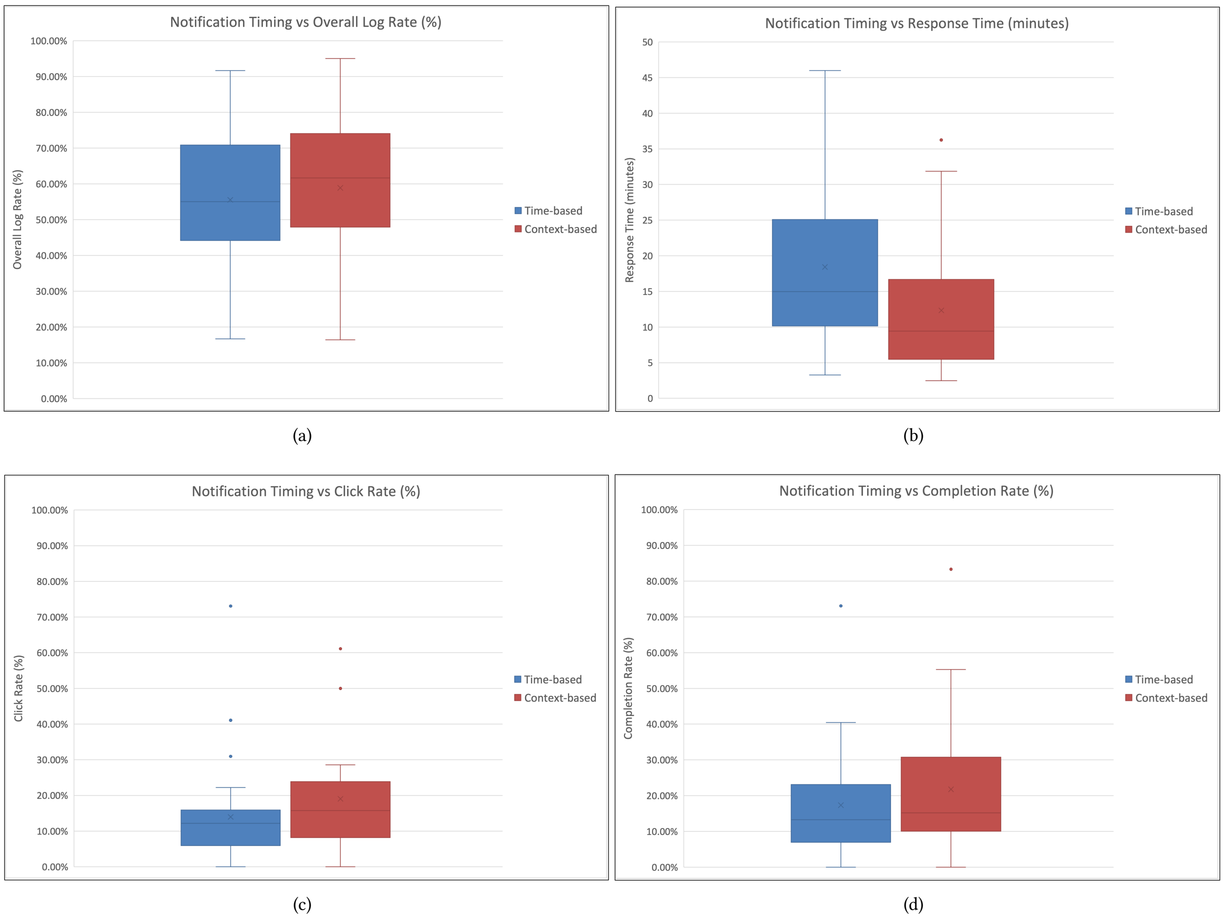


Figure 6: Box and whisker plots illustrating the notification timing conditions against the four dependent variables: (a) overall log rate, (b) notification click response time, (c) notification click rate, and (d) notification completion rate. Refer to Table 5 for the definition of these dependent variables. The notification timing conditions include time-based and context-based conditions.

condition by 5.09% (95% CI: [1.24, 8.94]). A repeated measures one-way ANOVA on *notification click rate* with a within-subjects factor of *notification timing* found a significant difference between the two conditions ($F_{1,29} = 7.3207$, $p < 0.05$). This result suggests that context-based notifications were *clicked* significantly more often when compared to time-based notifications.

4.4 Notification Completion Rate

To investigate the effect of notification timing on the user's timely self-monitoring behavior throughout the day, we ran analysis on the *notification completion rate*. We define *notification completion rate* as the number of *completed* notifications divided by the number of *sent* notifications, where a *completed* notification indicates that the notification's corresponding health data was logged within 60

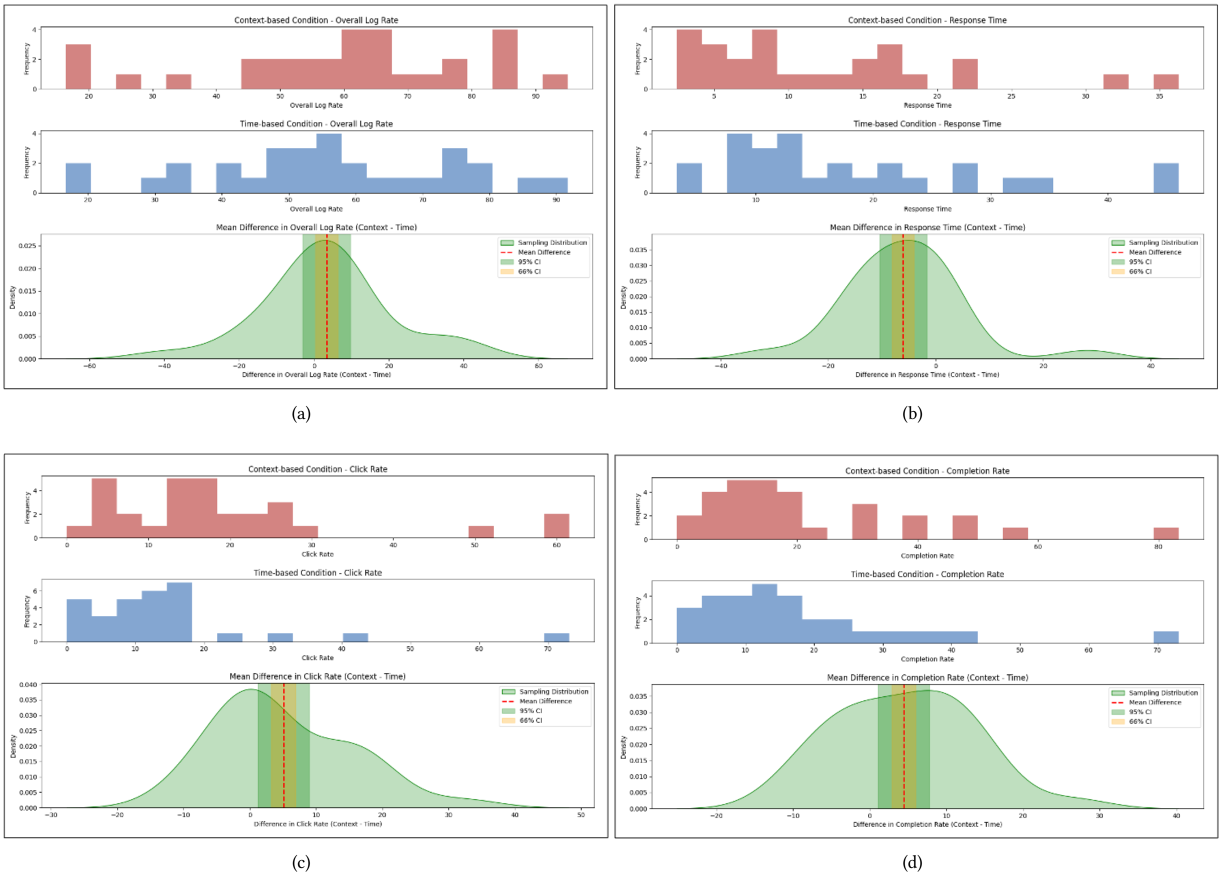


Figure 7: Estimated mean differences in (a) overall log rate, (b) notification click response time, (c) notification click rate, and (d) notification completion rate between context-based and time-based conditions, with 66% CI, 95% CI, and kernel density estimate of the sampling distribution.

minutes after the notification was sent. Because self-monitoring can occur at anytime during the day depending on individual differences, in order to measure the direct effect of notifications on self-monitoring behaviors, we aimed to measure the logging behaviors that were directly associated with the notifications we sent.

In the time-based condition, the average notification completion rate was 17.32% (min = 0%; max = 73.08%; SD = 14.93%). In the context-based condition, the average notification click response time was 21.77% (min = 0%; max = 83.33%; SD = 18.10%). The context-based condition had a greater average notification completion rate than the time-based condition by 4.45% (95% CI: [1.10, 7.78]). A repeated measures one-way ANOVA on *notification completion rate* with a within-subjects factor of *notification timing* found a significant difference between the two conditions ($F_{1,29} = 7.3831$, $p < 0.05$). This result suggests that context-based notifications were also *completed* significantly more often when compared to time-based notifications.

4.5 Main Categories and Themes

We present seven general themes identified through our qualitative analysis of the interview recordings, with specific examples from our interview transcripts.

4.5.1 Notification Interaction. During our interviews, participants described their daily interaction with notifications. Some participants logged their data immediately upon receiving notifications, with some explicitly mentioning that they did so because of concerns about the unreliability of their memory. For example, P1356 said, "*When I weigh myself in the morning, if I don't enter it right away, it is easy for me to forget the number.*" On the contrary, instead of logging data immediately upon receiving the notification, there were participants who chose to log all their data at once, often towards the end of the day, saying, for example, "*I like to input my information, after I had the final meals, like the final meals of the day.*" (P2610). Among those participants, some specifically described their strategies of handling notifications. For example, some participants mentioned the strategy of remembering the task and completing it later: "*Whenever I see [the notification], I'm normally in the middle of something. So I say, okay, I'll do it later.*" (P3792). Others indicated leaving notifications unread in the panel and attending to them when available or later in the day: "*If it's important, I leave it there. If not, I just delete it and then I'll check it later.*" (P9772). Additionally, some participants tended to treat notifications in the panel as summaries, choosing to clear all of them at once without clicking, and subsequently engaging with the associated tasks. For example, P7173 said, "*I guess I kind of like to dismiss a lot of the notifications because it's pretty much the same thing over and over.*"

4.5.2 Notification Effectiveness. In general, most participants found notifications effective in reminding them to log their health data. Additionally, in certain cases, notifications had the additional capability to trigger the associated behavior itself (e.g., eating), not just prompting the logging behavior. For example, P1356 said, "*It reminds me that I have to eat. It's less, for me, it's less about tracking because tracking has just become hand in hand with the eating for this period of time. So it's just, especially for lunchtime for me, it's like, hey, I need to take a break to go grab some lunch.*" Furthermore, some participants mentioned that they had developed a habit of tracking their health data during the study, saying, for example, "*It does [drive me more consistent] because it becomes a habit.*" (P8487). P9696 also said, "*But it's three or four weeks into the study. I'm used to it now. In the beginning, they were useful, but right now it's just, I don't really think about the notifications. I just do it.*"

4.5.3 Notification Preferences. Participants exhibited personalized preferences in various aspects. Regarding notification content, while some preferred the simplicity and consistency of our design of the notification, others suggested that the notification content could be more intriguing. For example, P7173 said, "*I don't know why I'm thinking about the wording so much, because in my head I just think of, like, jokes, some variation or playfulness. Maybe not sarcasm or insulting. That's the kind of humor I like personally though. But like, I guess have some playfulness to it, you know, like, have you been active?*" In terms of notification frequency, most participants mentioned that the number of notifications was appropriate. For example, P4343 said, "*I remember it's like one or two times a day. So, it's a very good number.*"

4.5.4 User Agency. Participants expressed their need for more customizable settings in FatSecret, particularly regarding flexible food choices for home-cooked meals. For example, P1356 said, "*If I make something from scratch, it's hard for me to account for, you know, how much oil I put in it when I'm cooking it and how much of this? and how much that? ... A number of times I was going to have something I don't even know what to enter.*" Some participants expressed their desire for control over how data is displayed and their need for the flexibility to log for previous days. For example, P4343 said, "*I can only record my weight within 24 hours. Let's say I forgot to log it in. And when I log it in, like, let's say, 12:30 AM, then it will become another day, and, the previous day will be just skipped. So if there's any function that I can trace back to the previous day, that would be very helpful.*"

4.5.5 Context of Use. In general, most participants mentioned adhering to their usual routines without making adjustments. For example, P2610 mentioned, "*Usually my daily schedules are pretty consistent. Nothing really changed.*" Nonetheless, there were participants who engaged with their smartphone in varied ways depending on the context of their daily lives. Location affected how participants interacted with their smartphone. Some participants did not always have their phones with them at home, stating, for example, "*Like I said, I work at home, so I just leave it [the phone] close by, but I don't carry it anywhere, so if I walk through the house, I just leave it on the office [table].*" (P9689). Additionally, some participants had different phone usage patterns depending on weekdays or weekends. There were also specific events in participants' lives that influenced how they engaged with our weight maintenance system. For example, one participant (P9772) mentioned that final exams caused them to be less active, saying, "*It's just that sometimes I forget. Because lately it's been just finals and stuff, so, lots of stuff.*"

4.5.6 System Usefulness. Most participants perceived GatorTrack as simple and user-friendly, considering it easily accessible and navigable. For example, P1753 stated, "*I could log when I ate food, what I ate, log my weight and see all of the trends ... everything is very nicely laid out, easy to use, like, simple and intuitive user interface.*" Participants also found the real-time feedback, delivered through a calorie balance equation and summary graphs (Figure 1), to be helpful in increasing their self-awareness and motivating them to track. For example, P4664 said, "*I actually really plan to lose weight and when I see [the] graph, I see how much has changed. Kind of makes me think maybe I should watch my calories more.*"

4.5.7 System Integration. Most participants expressed their desire for an integrated app comprising of both GatorTrack's and FatSecret's functionalities. For example, P6213 stated, "*Whatever the other app [FatSecret] provides me, this [GatorTrack] is just a, a small subset of it. So instead I just use the FatSecret app because it provides me more flexibility and options.*" Some participants expressed frustration with delays in synchronization and inaccuracies in data from different sources. For example, P6827 stated, "*It would be cool if it [GatorTrack] could run on its own and not have to be with another app [FatSecret] sometimes ... it does take a little while to sync data.*" A common practice was to immediately confirm the accuracy of recorded data and verify correct synchronization between GatorTrack and FatSecret after self-weighing or data logging.

5 DISCUSSION

We discuss the overall conclusions we can draw from our results, as well as how variations in our participants' individual self-monitoring behaviors affected them, whether or not the use of physical activity transition detection made a difference in self-monitoring, and implications for design of future weight management apps.

5.1 Effects of Notifications on Self-Monitoring Behaviors in Our Study

Prior research in the domain of weight management has emphasized the significance of self-monitoring one's health data each day [10–12]. Therefore, in our study, we measured the participant's engagement with daily self-monitoring by calculating the *overall log rate*, which counted one data log for each data type per day. Based on our results (Table 5), while the average overall log rate was slightly higher in the context-based condition (58.87%) compared to the time-based condition (55.54%), this difference was not significant. This implies that participants did not necessarily engage with self-monitoring more in the context-based condition. According to the Fogg Behavior Model [28], a certain behavior will be performed when an individual (1) is sufficiently motivated, (2) believes they have the ability to perform the behavior, and (3) is triggered to perform the behavior. Thus, while our notifications acted as triggers (#3), these are just one component of the broader behavior change process. Some previous research on mHealth apps has explored the impact of goal-setting features (#1) and users' self-efficacy (i.e., their belief in their ability to perform a behavior) (#2) on health behavior change [9, 64, 107]. More studies are needed to understand how these factors, along with self-monitoring, interact to shape the user's behavior change process.

When considering just those cases when participants actually engaged in self-monitoring behaviors, that is, when they engaged with the notifications and/or actually logged data, context-based notifications provided a benefit. We investigated the effects of context-based notifications on timely self-monitoring behavior (within 60 minutes after the notification was sent) by looking at *notification click response time*, *click rate*, and *completion rate*; all three of these dependent variables showed significant effects of notification timing condition. On average, participants clicked notifications faster (12.33m vs. 18.42m) and more frequently (19.05% vs. 13.96%) when they were in the context-based condition compared to when they were in the time-based condition. Prior work in the interruptibility literature has shown that deferring notification delivery until a natural breakpoint occurs, inferred from users' physical activity transitions, resulted in faster response times and increased click rates [73, 74]. Our study expands this literature by showing similar results in an mHealth weight maintenance self-monitoring app. It is important to point out that responding more to notifications does not necessarily imply that the user also logs data more frequently. But, participants also on average *completed* notifications more frequently (21.77% vs. 17.32%) in the context-based condition compared to the time-based condition. Since the notification was marked as *complete* only if the corresponding data was logged within 60 minutes after the notification was sent, the *notification completion rate* captured the user's *timely* engagement with the self-monitoring behavior. In our study, we sent notifications based on the user's specified daily schedule in both conditions, as detailed in Table 2 and Table 3. The underlying assumption was that users would adhere to the schedules they have specified. Our interview data largely supported this assumption, with most participants indicating they maintained consistent routines without making changes. However, it is crucial to highlight a critical aspect of the relevance of notifications in triggering *timely* self-monitoring behaviors—there could be a potential scenario where the user changed their personal schedule without updating the settings in GatorTrack accordingly, which would lead to a situation where completing a notification within 60 minutes might not have been *timely* with regard to the self-monitoring of the target behavior (eating or exercising). Notably, utilizing a mobile device to capture the full picture of the user's life comes with the inherent limitation of the need to rely on the user's active and precise personal settings and self-report.

This result shows that sending context-based notifications had a significant positive effect on motivating the timely (within 60 minutes) daily self-monitoring behaviors recommended by best practices in the clinical health psychology literature [10–12] and our own health experts. The benefits of timely self-monitoring include enhanced self-awareness for adjustment within the day, mitigation of memory biases, and the potential for effective real-time health interventions and the development of predictive machine learning models for adaptive and personalized health services. We discuss the details of our implications for the design of future weight maintenance app in Section 5.4.

5.2 Individual Differences in Self-Monitoring Behaviors in Our Study

Our analysis of the significant *notification completion rate*, the count of data logged within 60 minutes after the associated notification was sent, also pointed to individual differences between participants' patterns of self-monitoring behavior that were interesting to examine in more detail, especially in contrast to the lack of significant differences in the *overall log rate* results. As illustrated in Figure 2, the overall log rate measurement takes into account daily logging events in three cases relative to the scheduled notification time: (C1) data logged before the notification time (preventing a notification from being sent at all) [the blue duration in Figure 2-a], (C2) data logged within 60 minutes after the notification was sent (notification completion rate) [the orange duration in Figure 2-b], and (C3) data logged more than 60 minutes after the notification was sent. The significant effect of notification timing condition on *notification completion rate* (C2) suggests that context-based notifications can trigger more timely self-monitoring behavior. However, the lack of a significant effect of notification timing condition on the overall log rate (C1+C2+C3) suggests that individual differences may have influenced our overall self-monitoring behavior measure.

As we saw from our qualitative analysis, some participants mentioned that they developed a habit of tracking their health data during the study after a period of time, which corresponds to C1. Additionally, there were participants who tended to log all their data at once, usually at the end of the day, which corresponds to C3, instead of immediately upon receiving a notification. Some participants stated that they would remember the task and do it afterwards, while others mentioned that they would leave the notifications unread in the panel and address them later in the day. We can infer from these participants' comments that, even though the notifications were not responded to within 60 minutes after being sent, they still had the effect of triggering participants' logging behavior by virtue of the reminder.

While prior work (e.g., A-CHESS [36]) focused on designing contextual notifications to help users develop a target health behavior (e.g., recovery from alcohol abuse), our work focused on an essential component for promoting health behavior change—self-monitoring, a practice proven to contribute to achieving the target health goal, such as weight maintenance [11, 12, 47, 64]. Apart from showing that contextual notifications significantly increased *timely* self-monitoring, our interview data revealed an interesting observation: when delivered before a meal, for some participants, notifications had the additional capability to trigger the relevant behavior itself (e.g., eating), beyond prompting the logging behavior. Although such situations (e.g., receiving lunch notifications before the participant had eaten lunch) suggest that the notifications were not timed as intended, in certain cases, the notifications appeared to contribute positively to supporting health behavior change. This observation hints at the potential use of contextual notifications for triggering relevant behaviors (e.g., eating and exercising) to support weight maintenance in future studies. However, given that the relevant behavior itself demands more effort than simply monitoring, additional behavior change techniques may be necessary.

5.3 Detection of Physical Activity Transitions Helped Increase Timely Self-Monitoring

Recall that we considered two contextual factors for sending our context-based notifications: time of day and physical activity transitions, detected through the use of Google's Transition API. In our design, we factored in the time of day by creating a time range (2.5 hours for weight data, 1.5 hours for other types) aligned to the participant's daily schedule in which to listen for their activity transitions (Table 3). If no transitions were detected by the end of the time range, a "timeout" notification was still sent. In our analysis of the notification completion rate, context-based notifications triggered significantly more data logging events compared to time-based notifications (21.77% vs. 17.32%). To understand the effect of each contextual factor, we further examined the percentages of notifications sent based on the detection of physical activity transitions (i.e., transition-triggered notifications) and those sent by the end of the specified time range when no transitions had been detected (i.e., timeout notifications). We found that among the 278 *completed* context-based notifications, 108 (38.85%) were transition-triggered notifications while 170 (61.15%) were timeout notifications. This result suggests that, while transitions were prevalent, they were not the only driving factor behind logging behaviors. In fact, some participants mentioned they did not always keep their phones on them, making activity transition detection infeasible. However, when we compared only the context-based "timeout" notifications with the purely time-based notifications, a repeated measures ANOVA on *notification completion rate* with a within-subjects factor of *notification timing* (timeout, M = 20.05%; SD = 16.02%; vs. time-based, M = 17.32%; SD = 14.93%) found no significant difference ($F_{1,29} = 2.0684$, $p > 0.05$). When we compared only the context-based transition-triggered notifications with the time-based notifications, a repeated measures ANOVA on the *notification completion rate* with a within-subjects factor of *notification timing* (transition-triggered vs. time-based) did find a significant difference ($F_{1,29} = 5.9181$, $p < 0.05$). The average completion rate for context-based transition-triggered notifications alone (M = 26.64%; SD = 25.83%) was higher than purely time-based notifications (M = 17.32%; SD = 14.93%). This result suggests that, while transition-triggered notifications were slightly less frequent overall than timeout notifications in the context-based condition, these transition-triggered notifications played a significant role in triggering more timely self-monitoring behaviors.

5.4 Implications for the Design of Future Weight Maintenance Apps

The weight management literature (and broader literature on self-monitoring [3]) recommends that, to achieve the greatest benefit from self-monitoring, people should log their food and activity throughout the day (i.e., as eating/exercising occurs) versus waiting until the end of the day to log [10–12]. The benefits of timely self-monitoring are threefold: (1) the practice of tracking and receiving real-time feedback fosters timely self-reflection, enhances self-awareness, and provides opportunities for adapting dietary or exercise plans [3, 47]; (2) logging health data promptly when the behavior occurs addresses the unreliability and recall bias of human memory [42]; and (3) the accurate logging of real-time data holds potential for effective real-time health interventions and the development of predictive machine learning models, enabling timely, adaptive, and personalized health services [68]. Logging as soon as possible after eating/exercising occurs helps individuals to be more mindful of their within-day progress, so that they have opportunities to make changes according to their daily calorie goal, for example, adjusting the portion size of their later meals based on earlier ones if needed, or adapting their exercise plans. It can also help individuals plan ahead for their daily health goals if there happen to be special events to consider or interruptions in their routine schedules. Timely self-monitoring further enables real-time feedback on users' recorded data, offering a chance for them to adapt and refine their goals and strategy—another key recommendation from the weight maintenance literature [20]. GatorTrack provided real-time feedback through a calorie balance equation and summary graphs (Figure 1), based on design recommendations from our health experts, that were viewed by participants as easily accessible and navigable.

However, we heard from some participants in our study that their logging pattern was to remember what they ate and log all data at the end of the day. In addition to these participants missing out on the chance for just-in-time adjustment, there is also the challenge of human memory being unreliable [42] when logging after the fact. Retrieving an event from memory and reconstructing it can unintentionally introduce biases and inaccuracies [45, 95], even within a single day's events. Moreover, the recall bias can be more pronounced for dietary intake due to the frequent nature of eating, making it easy to forget details [86].

Lastly, self-reporting, a primary data source in weight maintenance apps, provides health professionals with insights into users' daily routines for personalized feedback and treatment. Logging data as close to the occurrence of the behavior as possible offers real-time information crucial for effective real-time health intervention. For example, an adaptive weight maintenance intervention relies on data points with precise timestamps of when the behavior actually occurs to deliver the most effective intervention. Furthermore, accurate real-time data can potentially be used for developing predictive machine learning models (e.g., reinforcement learning-based models [61]) that are capable of empowering health professionals to provide real-time, adaptive, and personalized health services.

As discussed, the results from our study showed that contextual notifications were effective in triggering more timely self-monitoring. However, there is still room to improve participants' overall self-monitoring behavior to be in line with recommendations from health behavior change literature. More research is needed to understand how to find the right balance between individual differences (e.g., preferring to log at the end of the day) and health guidelines (e.g., recommending to log throughout the day). Perhaps an mHealth app could pre-populate a day's log based on asking the user a few simple questions at the beginning of the day or end of the prior day, or other onboard sensors could be used to detect mealtimes and meal contents (e.g., photos of meals a user is about to eat [48, 82]). Another option is to use newly emerging AI tools like large-language models (LLMs) [94] to support participants throughout the day in a kind of coaching relationship that strikes a better balance for the user experience. However, the use of AI to support mHealth applications is an emerging field and comes with many questions about participants' data privacy and issues of fairness, accountability, transparency, and ethics (FATE) central to human-AI interaction [54], not to mention basic questions of accessibility and cost of requiring mobile smartphones with data access to use these apps. Furthermore, while weight management literature recommends self-monitoring throughout the day, HCI research has documented the challenges users encounter in food journaling, highlighting the importance of tailoring the journaling process to users' specific health goals [15]. More research is needed at the intersection of mHealth and HCI to address these questions to motivate behavior change and eventually improve health outcomes.

6 LIMITATIONS AND FUTURE WORK

Our work has several limitations. First, to simulate a practical application, we selected weight maintenance as the targeted health goal for participants. Although this is a commonly employed scenario, it still might not encompass the full spectrum of use cases for mHealth apps. Future research should consider extending our findings to other mHealth scenarios to assess the general applicability of our results.

Second, as mentioned in Section 3.7, during the study, two bugs in our app implementation caused (a) one notification out of 120 to not be sent for two participants, and (b) one additional weekly check-in questionnaire notification to be sent to five participants. To address these issues, we normalized the log counts by the actual number of notifications sent to each participant in our data analysis. Additionally, it is important to note that we excluded four participants from our analysis of click response time because they did not click on any notifications during the second half of the study (though they continued logging and otherwise participating in the study).

Third, our population sample exhibited imbalance in terms of ethnicity and gender: 63.33% of our participants self-identified as Asian, and 66.67% self-identified as male. Additionally, our participants were recruited from the local community or cities within the same time zone. This demographic might not include comprehensive cultural or regional differences that could have an impact on app usage and responses to notifications. Future work should consider expanding the population considered in order to understand how these details may have biased our results. In particular, weight loss and weight maintenance studies in the clinical health literature typically attract more white women than other demographics [32, 39, 52], so generally speaking replicating this study with different population groups would be informative. Furthermore, due to our app implementation, we recruited only Samsung smartphone users. However, prior work in the health literature has shown that there is no difference in mHealth self-monitoring impacts on weight loss between Android and iOS phone users [97], so our results should still generalize.

Fourth, we must consider the possible effect of study participation itself on the frequency of overall self-monitoring behavior. Our study was conducted by deploying our app on participants' own phones over a period of one month with only three interactions with the research team, to help increase naturalness and validity. However, some participants did report that they were logging more often initially because they were participating in the study. As time progressed, though, many of them told us that this behavior evolved into a habit. Interestingly, several participants expressed their intention to continue logging and self-monitoring after the study ended. Others found daily logging to be less beneficial to them and had no intention of continuing it after the study concluded.

Fifth, as we discussed in Section 5.3, relying solely on a smartphone for detecting transitions between physical activities may not be optimal. During our study, participants told us that they occasionally left their phone behind when they moved about or ate meals, often for reasons such as charging their phone at their desk. Future work could consider integrating wearables such as smartwatches, potentially improving the ability to detect more accurate physical activity transitions for behavioral health intervention.

Lastly, although our study involves the use of a calorie tracking app, we did not include a detailed safety plan to mitigate potential issues such as promoting disordered eating. It is possible that calorie counting can potentially create challenges for certain individuals, such as feelings of shame or judgment [15]. In our study, participants were instructed to focus solely on tracking their health data without any emphasis on changing dietary intake, physical activity, or weight. Our health experts assessed this approach as posing minimal risk of potential harm. Nevertheless, future work involving self-monitoring of dietary intake (especially studies aiming to change dietary intake or those that set specific caloric intake goals) should consider including protocols to minimize and manage potential risks related to these concerns.

7 CONCLUSION

Many mHealth apps for weight maintenance rely on self-monitoring as a critical component to empower users to achieve their health goals. However, consistent adherence to self-monitoring tends to decrease over time. While push notifications in these apps can increase users' app engagement, adherence, and long-term retention, they are often ignored by users due to appearing at inopportune moments. Contextual factors including time of day and physical activity transitions have been found to be effective in estimating user interruptibility in other domains. Therefore, we conducted a within-subjects study in the context of an mHealth app for weight loss and weight maintenance, employing notification timing as the independent variable with two conditions: time-based notifications and context-based notifications. We integrated insights from the interruptibility literature into a weight maintenance app designed and developed in collaboration with health experts on our team, enabling us to achieve a closer alignment with real-world mHealth use cases. The outcomes of our four-week in-the-wild study with 30 participants suggest that, while notification timing had no significant effect on overall daily self-monitoring behavior, context-based notifications were effective in triggering faster and more frequent responses to the notifications, as well as in fostering more timely self-monitoring behaviors (i.e., as eating/exercise occurs). Our study offers a deeper understanding of how using contextual factors for sending notifications can trigger more **timely** self-monitoring behaviors. Finally, we discuss how our results can inform the future design of mHealth apps, and suggest implications for future research on finding the right balance between real-world individual differences in behavior and optimal health-related guidelines.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health Grant Awards #R01DK119244. Opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect these agencies' views.

REFERENCES

[1] Sangwon Bae, Tammy Chung, Denzil Ferreira, Anind K Dey, and Brian Suffoletto. 2018. Mobile phone sensors and supervised machine learning to identify alcohol use events in young adults: Implications for just-in-time adaptive interventions. *Addictive behaviors* 83 (Aug. 2018), 42–47. https://doi.org/10.1016/j.addbeh.2017.11.039

[2] Albert Bandura. 1977. Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review* 84, 2 (March 1977), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

[3] Albert Bandura. 1998. Health promotion from the perspective of social cognitive theory. *Psychology & Health* 13, 4 (July 1998), 623–649. https://doi.org/10.1080/08870449808407422

[4] Albert Bandura. 2004. Health Promotion by Social Cognitive Means. *Health Education & Behavior* 31, 2 (April 2004), 143–164. https://doi.org/10.1177/1090198104263660

[5] Dror Ben-Zeev, Christopher J Brenner, Mark Begale, Jennifer Duffecy, David C Mohr, and Kim T Mueser. 2014. Feasibility, Acceptability, and Preliminary Efficacy of a Smartphone Intervention for Schizophrenia. *Schizophrenia Bulletin* 40, 6 (Nov. 2014), 1244–1253. https://doi.org/10.1093/schbul/sbu033

[6] Frank Bentley and Konrad Tollmar. 2013. The Power of Mobile Notifications to Increase Wellbeing Logging Behavior. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. ACM, New York, NY, USA, 1095–1098. https://doi.org/10.1145/2470654.2466140

[7] Hugh Beyer and Karen Holtzblatt. 1999. Contextual Design. *Interactions* 6, 1 (Jan. 1999), 32–42. https://doi.org/10.1145/291224.291229

[8] Niranjan Bidargaddi, Daniel Almirall, Susan Murphy, Inbal Nahum-Shani, Michael Kovalcik, Timothy Pituch, Haitham Maaieh, and Victor Strecher. 2018. To Prompt or Not to Prompt? A Microrandomized Trial of Time-Varying Push Notifications to Increase Proximal Engagement With a Mobile Health App. *JMIR mHealth and uHealth* 6, 11 (Nov. 2018), e10123. https://doi.org/10.2196/10123

[9] Madison L Bracken and Bradly M Waite. 2020. Self-Efficacy and Nutrition-Related Goal Achievement of MyFitnessPal Users. *Health Education & Behavior* 47, 5 (June 2020), 677–681. https://doi.org/10.1177/1090198120936261

[10] Lora E Burke, Susan M Sereika, Edvin Music, Melanie Warziski, Mindi A Styn, and Arthur Stone. 2008. Using instrumented paper diaries to document self-monitoring patterns in weight loss. *Contemporary clinical trials* 29, 2 (March 2008), 182–193. https://doi.org/10.1016/j.cct.2007.07.004

[11] Lora E Burke, Jing Wang, and Mary Ann Sevick. 2011. Self-Monitoring in Weight Loss: A Systematic Review of the Literature. *Journal of the American Dietetic Association* 111, 1 (Jan. 2011), 92–102. https://doi.org/10.1016/j.jada.2010.10.008

[12] Chelsea A Carpenter, Umelo A Ugwoaba, Michelle I Cardel, and Kathryn M Ross. 2022. Using self-monitoring technology for nutritional counseling and weight management. *DIGITAL HEALTH* 8 (May 2022), 20552076221102774. https://doi.org/10.1177/20552076221102774

[13] Woohyeok Choi, Sangkeun Park, Duyeon Kim, Youn-kyung Lim, and Uichin Lee. 2019. Multi-Stage Receptivity Model for Mobile Just-In-Time Health Intervention. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 2, Article 39 (June 2019), 26 pages. https://doi.org/10.1145/3328910

[14] James Clawson, Jessica A. Pater, Andrew D. Miller, Elizabeth D. Mynatt, and Lena Mamykina. 2015. No longer wearing: investigating the abandonment of personal health-tracking technologies on craigslist. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan) *(UbiComp '15)*. ACM, New York, NY, USA, 647–658. https://doi.org/10.1145/2750858.2807554

[15] Felicia Cordeiro, Daniel A. Epstein, Edison Thomaz, Elizabeth Bales, Arvind K. Jagannathan, Gregory D. Abowd, and James Fogarty. 2015. Barriers and Negative Nudges: Exploring Challenges in Food Journaling. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 1159–1162. https://doi.org/10.1145/2702123.2702155

[16] Mary Czerwinski, Eric Horvitz, and Susan Wilhite. 2004. A diary study of task switching and interruptions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vienna, Austria) *(CHI '04)*. ACM, New York, NY, USA, 175–182. https://doi.org/10.1145/985692.985715

[17] 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* (Tokyo, Japan) *(UIST '16)*. ACM, New York, NY, USA, 347–358. https://doi.org/10.1145/2984511.2984534

[18] Android Developers. 2024. Detect when users start or end an activity. Retrieved January 3, 2024 from https://developer.android.com/develop/sensors-and-location/location/transitions

[19] Anind K. Dey, Katarzyna Wac, Denzil Ferreira, Kevin Tassini, Jin-Hyuk Hong, and Julian Ramos. 2011. Getting closer: an empirical investigation of the proximity of user to their smart phones. In *Proceedings of the 13th International Conference on Ubiquitous Computing* (Beijing, China) *(UbiComp '11)*. ACM, New York, NY, USA, 163–172. https://doi.org/10.1145/2030112.2030135

[20] Katerina Dounavi and Olga Tsoumani. 2019. Mobile Health Applications in Weight Management: A Systematic Literature Review. *American Journal of Preventive Medicine* 56, 6 (June 2019), 894–903. https://doi.org/10.1016/j.amepre.2018.12.005

[21] Mauro Dragoni, Ivan Donadello, and Claudio Eccher. 2020. Explainable AI meets persuasiveness: Translating reasoning results into behavioral change advice. *Artificial Intelligence in Medicine* 105 (2020), 101840.

[22] Jonas Elslander and Katsumi Tanaka. 2013. A Notification-Centric Mobile Interaction Survey and Framework. In *Social Informatics*, Adam Jatowt, Ee-Peng Lim, Ying Ding, Asako Miura, Taro Tezuka, Gaël Dias, Katsumi Tanaka, Andrew Flanagin, and Bing Tian Dai (Eds.), Vol. 8238. Springer International Publishing, Cham, 443–456. https://doi.org/10.1007/978-3-319-03260-3_38

[23] FatSecret. 2023. FatSecret - Calorie Counter and Diet Tracker for Weight Loss. Retrieved September 6, 2023 from https://www.fatsecret.com/

[24] FatSecret Platform. 2023. The Largest Global Nutrition Database, Recipe and Food API. Retrieved September 6, 2023 from https://platform.fatsecret.com/api/

[25] Adrienne Porter Felt, Serge Egelman, and David Wagner. 2012. I've got 99 problems, but vibration ain't one: a survey of smartphone users' concerns. In *Proceedings of the Second ACM Workshop on Security and Privacy in Smartphones and Mobile Devices* (Raleigh, North Carolina, USA) *(SPSM '12)*. ACM, New York, NY, USA, 33–44. https://doi.org/10.1145/2381934.2381943

[26] Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. 2017. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Mental Health* 4, 2 (June 2017), e19. https://doi.org/10.2196/mental.7785

[27] James Fogarty, Amy J. Ko, Htet Htet Aung, Elspeth Golden, Karen P. Tang, and Scott E. Hudson. 2005. Examining task engagement in sensor-based statistical models of human interruptibility. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Portland, Oregon, USA) *(CHI '05)*. ACM, New York, NY, USA, 331–340. https://doi.org/10.1145/1054972.1055018

[28] BJ Fogg. 2009. A behavior model for persuasive design. In *Proceedings of the 4th International Conference on Persuasive Technology* (Claremont, California, USA) *(Persuasive '09)*. ACM, New York, NY, USA, Article 40, 7 pages. https://doi.org/10.1145/1541948.1541999

[29] Evan M Forman, Stephanie G Kerrigan, Meghan L Butryn, Adrienne S Juarascio, Stephanie M Manasse, Santiago Ontañón, Diane H Dallal, Rebecca J Crochiere, and Danielle Moskow. 2019. Can the artificial intelligence technique of reinforcement learning use continuously-monitored digital data to optimize treatment for weight loss? *Journal of Behavioral Medicine* 42, 2 (April 2019), 276–290. https://doi.org/10.1007/s10865-018-9964-1

[30] Yoshimi Fukuoka, Caryl Gay, William Haskell, Shoshana Arai, and Eric Vittinghoff. 2015. Identifying Factors Associated With Dropout During Prerandomization Run-in Period From an mHealth Physical Activity Education Study: The mPED Trial. *JMIR mHealth uHealth* 3, 2 (April 2015), e34. https://doi.org/10.2196/mhealth.3928

[31] Elliot G. Mitchell, Elizabeth M. Heitkemper, Marissa Burgermaster, Matthew E. Levine, Yishen Miao, Maria L. Hwang, Pooja M. Desai, Andrea Cassells, Jonathan N. Tobin, Esteban G. Tabak, David J. Albers, Arlene M. Smaldone, and Lena Mamykina. 2021. From Reflection to Action: Combining Machine Learning with Expert Knowledge for Nutrition Goal Recommendations. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. ACM, New York, NY, USA, Article 206, 17 pages. https://doi.org/10.1145/3411764.3445555

[32] Daniel Gero, Bálint File, Noreen Hinrichs, Matteo Mueller, István Ulbert, Zoltán Somogyvári, and Marco Bueter. 2020. Mental and emotional representations of "weight loss": free-word association networks in members of bariatric surgery-related social media communities. *Surgery for Obesity and Related Diseases* 16, 9 (Sept. 2020), 1312–1320. https://doi.org/10.1016/j.soard.2020.05.032

[33] Enying Gong, Shaira Baptista, Anthony Russell, Paul Scuffham, Michaela Riddell, Jane Speight, Dominique Bird, Emily Williams, Mojtaba Lotfaliany, and Brian Oldenburg. 2020. My Diabetes Coach, a Mobile App–Based Interactive Conversational Agent to Support Type 2 Diabetes Self-Management: Randomized Effectiveness-Implementation Trial. *Journal of Medical Internet Research* 22, 11 (Nov. 2020), e20322. https://doi.org/10.2196/20322

[34] Mitchell Gordon, Tim Althoff, and Jure Leskovec. 2019. Goal-setting And Achievement In Activity Tracking Apps: A Case Study Of MyFitnessPal. In *The World Wide Web Conference* (San Francisco, CA, USA) *(WWW '19)*. ACM, New York, NY, USA, 571–582. https://doi.org/10.1145/3308558.3313432

[35] Franz Gravenhorst, Amir Muaremi, Jakob Bardram, Agnes Grünerbl, Oscar Mayora, Gabriel Wurzer, Mads Frost, Venet Osmani, Bert Arnrich, Paul Lukowicz, and Gerhard Tröster. 2015. Mobile phones as medical devices in mental disorder treatment: an overview. *Personal and Ubiquitous Computing* 19, 2 (Feb. 2015), 335–353. https://doi.org/10.1007/s00779-014-0829-5

[36] David H. Gustafson, Fiona M. McTavish, Ming-Yuan Chih, Amy K. Atwood, Roberta A. Johnson, Michael G. Boyle, Michael S. Levy, Hilary Driscoll, Steven M. Chisholm, Lisa Dillenburg, Andrew Isham, and Dhavan Shah. 2014. A Smartphone Application to Support Recovery From Alcoholism: A Randomized Clinical Trial. *JAMA Psychiatry* 71, 5 (May 2014), 566–572. https://doi.org/10.1001/jamapsychiatry.2013.4642

[37] Daniel Harrison, Paul Marshall, Nadia Bianchi-Berthouze, and Jon Bird. 2015. Activity tracking: barriers, workarounds and customisation. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan) *(UbiComp '15)*. ACM, New York, NY, USA, 617–621. https://doi.org/10.1145/2750858.2805832

[38] William L Haskell, I-Min Lee, Russell R Pate, Kenneth E Powell, Steven N Blair, Barry A Franklin, Caroline A Macera, Gregory W Heath, Paul D Thompson, and Adrian Bauman. 2007. Physical Activity and Public Health: Updated Recommendation for Adults From the American College of Sports Medicine and the American Heart Association. *Circulation* 116, 9 (Aug. 2007), 1081–1093. https://doi.org/10.1161/CIRCULATIONAHA.107.185649

[39] Christina F Haughton, Valerie J Silfee, Monica L Wang, Andrea C Lopez-Cepero, David P Estabrook, Christine Frisard, Milagros C Rosal, Sherry L Pagoto, and Stephenie C Lemon. 2018. Racial/ethnic representation in lifestyle weight loss intervention studies in the United States: A systematic review. *Preventive medicine reports* 9 (March 2018), 131–137. https://doi.org/10.1016/j.pmedr.2018.01.012

[40] Corine Horsch, Sandor Spruit, Jaap Lancee, Rogier van Eijk, Robbert Jan Beun, Mark Neerincx, and Willem-Paul Brinkman. 2017. Reminders make people adhere better to a self-help sleep intervention. *Health and technology* 7, 2 (Nov. 2017), 173–188. https://doi.org/10.1007/s12553-016-0167-x

[41] Eric Horvitz, Paul Koch, Raman Sarin, Johnson Apacible, and Muru Subramani. 2005. Bayesphone: precomputation of context-sensitive policies for inquiry and action in mobile devices. In *Proceedings of the 10th International Conference on User Modeling* (Edinburgh, UK) *(UM'05)*. Springer-Verlag, Berlin, Heidelberg, 251–260. https://doi.org/10.1007/11527886_33

[42] Michael R. Hufford and Saul Shiffman. 2003. Assessment Methods for Patient-Reported Outcomes. *Disease Management and Health Outcomes* 11, 2 (Feb. 2003), 77–86. https://doi.org/10.2165/00115677-200311020-00002

[43] Becky Inkster, Shubhankar Sarda, and Vinod Subramanian. 2018. An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. *JMIR mHealth and uHealth* 6, 11 (Nov. 2018), e12106. https://doi.org/10.2196/12106

[44] Shamsi T. Iqbal and Brian P. Bailey. 2008. Effects of intelligent notification management on users and their tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. ACM, New York, NY, USA, 93–102. https://doi.org/10.1145/1357054.1357070

[45] Jared B Jobe. 2000. Cognitive processes in self-report. In *The science of self-report: Implications for research and practice*, Arthur A Stone, Jaylan S Turkkan, Christine A Bachrach, Jared B Jobe, Howard S Kurtzman, and Virginia S Cain (Eds.). Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US, 25–29.

[46] Norazlyn Kamal Basha, Eugene Cheng-Xi Aw, and Stephanie Hui-Wen Chuah. 2022. Are we so over smartwatches? Or can technology, fashion, and psychographic attributes sustain smartwatch usage? *Technology in Society* 69 (May 2022), 101952. https://doi.org/10.1016/j.techsoc.2022.101952

[47] Frederick H Kanfer. 1970. Self-monitoring: Methodological limitations and clinical applications. *Journal of Consulting and Clinical Psychology* 35, 2 (1970), 148–152. https://doi.org/10.1037/h0029874

[48] Ho Heon Kim, Youngin Kim, and Yu Rang Park. 2021. Interpretable Conditional Recurrent Neural Network for Weight Change Prediction: Algorithm Development and Validation Study. *JMIR mHealth and uHealth* 9, 3 (March 2021), e22183. https://doi.org/10.2196/22183

[49] Predrag Klasnja and Wanda Pratt. 2012. Healthcare in the pocket: Mapping the space of mobile-phone health interventions. *Journal of Biomedical Informatics* 45, 1 (Feb. 2012), 184–198. https://doi.org/10.1016/j.jbi.2011.08.017

[50] Florian Künzler, Varun Mishra, Jan-Niklas Kramer, David Kotz, Elgar Fleisch, and Tobias Kowatsch. 2020. Exploring the State-of-Receptivity for mHealth Interventions. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 4, Article 140 (Sept. 2020), 27 pages. https://doi.org/10.1145/3369805

[51] Brian Yoshio Laing, Carol M. Mangione, Chi-Hong Tseng, Mei Leng, Ekaterina Vaisberg, Megha Mahida, Michelle Bholat, Eve Glazier, Donald E. Morisky, and Douglas S. Bell. 2014. Effectiveness of a Smartphone Application for Weight Loss Compared With Usual Care in Overweight Primary Care Patients: A Randomized, Controlled Trial. *Annals of Internal Medicine* 161, 10_Supplement (Nov. 2014), S5–S12. https://doi.org/10.7326/M13-3005

[52] Stephenie C Lemon, Milagros C Rosal, Jane Zapka, Amy Borg, and Victoria Andersen. 2009. Contributions of weight perceptions to weight loss attempts: differences by body mass index and gender. *Body image* 6, 2 (March 2009), 90–96. https://doi.org/10.1016/j.bodyim.2008.11.004

[53] Peng Liao, Kristjan Greenewald, Predrag Klasnja, and Susan Murphy. 2020. Personalized HeartSteps: A Reinforcement Learning Algorithm for Optimizing Physical Activity. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 1, Article 18 (March 2020), 22 pages. https://doi.org/10.1145/3381007

[54] Q. Vera Liao, Hariharan Subramonyam, Jennifer Wang, and Jennifer Wortman Vaughan. 2023. Designerly Understanding: Information Needs for Model Transparency to Support Design Ideation for AI-Powered User Experience. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. ACM, New York, NY, USA, Article 9, 21 pages. https://doi.org/10.1145/3544548.3580652

[55] 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 *Proceedings of the 8th International Conference on Ubiquitous Computing* (Orange County, CA) *(UbiComp'06)*. Springer-Verlag, Berlin, Heidelberg, 261–278. https://doi.org/10.1007/11853565_16

[56] 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 (Sept. 2002), 705–717. https://doi.org/10.1037/0003-066X.57.9.705

[57] Gloria Mark, Victor M. Gonzalez, and Justin Harris. 2005. No task left behind? examining the nature of fragmented work. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Portland, Oregon, USA) *(CHI '05)*. ACM, New York, NY, USA, 321–330. https://doi.org/10.1145/1054972.1055017

[58] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and Inter-Rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 72 (Nov. 2019), 23 pages. https://doi.org/10.1145/3359174

[59] Abhinav Mehrotra, Robert Hendley, and Mirco Musolesi. 2016. PrefMiner: mining user's preferences for intelligent mobile notification management. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Heidelberg, Germany) *(UbiComp '16)*. ACM, New York, NY, USA, 1223–1234. https://doi.org/10.1145/2971648.2971747

[60] Abhinav Mehrotra, Veljko Pejovic, Jo Vermeulen, Robert Hendley, and Mirco Musolesi. 2016. My Phone and Me: Understanding People's Receptivity to Mobile Notifications. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. ACM, New York, NY, USA, 1021–1032. https://doi.org/10.1145/2858036.2858566

[61] Marianne Menictas, Mashfiqui Rabbi, Predrag Klasnja, and Susan Murphy. 2019. Artificial intelligence decision-making in mobile health. *The Biochemist* 41, 5 (Oct. 2019), 20–24. https://doi.org/10.1042/BIO04105020

[62] Jun-Ki Min, Afsaneh Doryab, Jason Wiese, Shahriyar Amini, John Zimmerman, and Jason I. Hong. 2014. Toss 'n' turn: smartphone as sleep and sleep quality detector. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. ACM, New York, NY, USA, 477–486. https://doi.org/10.1145/2556288.2557220

[63] Miro. 2024. Miro: The Visual Collaboration Platform for Every Team. Retrieved January 1, 2024 from https://miro.com

[64] Asuka Miyake, Masami Takahashi, Ryo Hashimoto, and Momoko Nakatani. 2021. StepUp Forecast: Predicting Future to Promote Walking. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction* (Toulouse & Virtual, France) *(MobileHCI '21)*. ACM, New York, NY, USA, Article 16, 12 pages. https://doi.org/10.1145/3447526.3472020

[65] David C Mohr, Pim Cuijpers, and Kenneth Lehman. 2011. Supportive Accountability: A Model for Providing Human Support to Enhance Adherence to eHealth Interventions. *Journal of Medical Internet Research* 13, 1 (March 2011), e30. https://doi.org/10.2196/jmir.1602

[66] Leanne G Morrison, Charlie Hargood, Veljko Pejovic, Adam W A Geraghty, Scott Lloyd, Natalie Goodman, Danius T Michaelides, Anna Weston, Mirco Musolesi, Mark J Weal, and Lucy Yardley. 2017. The Effect of Timing and Frequency of Push Notifications on Usage of a Smartphone-Based Stress Management Intervention: An Exploratory Trial. *PLOS ONE* 12, 1 (Jan. 2017), e0169162. https://doi.org/10.1371/journal.pone.0169162

[67] Frederick Muench and Amit Baumel. 2017. More Than a Text Message: Dismantling Digital Triggers to Curate Behavior Change in Patient-Centered Health Interventions. *Journal of Medical Internet Research* 19, 5 (May 2017), e147. https://doi.org/10.2196/jmir.7463

[68] Inbal Nahum-Shani, Shawna N Smith, Bonnie J Spring, Linda M Collins, Katie Witkiewitz, Ambuj Tewari, and Susan A Murphy. 2018. Just-in-Time Adaptive Interventions (JITAIs) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support. *Annals of Behavioral Medicine* 52, 6 (May 2018), 446–462. https://doi.org/10.1007/s12160-016-9830-8

[69] Linda Neuhauser, Gary L Kreps, Kathleen Morrison, Marcos Athanasoulis, Nikolai Kirienko, and Deryk Van Brunt. 2013. Using design science and artificial intelligence to improve health communication: ChronologyMD case example. *Patient Education and Counseling* 92, 2 (Aug. 2013), 211–217. https://doi.org/10.1016/j.pec.2013.04.006

[70] Tadashi Okoshi, Jin Nakazawa, and Hideyuki Tokuda. 2014. Attelia: sensing user's attention status on smart phones. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (Seattle, Washington) *(UbiComp '14 Adjunct)*. ACM, New York, NY, USA, 139–142. https://doi.org/10.1145/2638728.2638802

[71] Tadashi Okoshi, Julian Ramos, Hiroki Nozaki, Jin Nakazawa, Anind K. Dey, and Hideyuki Tokuda. 2015. Attelia: Reducing user's cognitive load due to interruptive notifications on smart phones. In *2015 IEEE International Conference on Pervasive Computing and Communications (PerCom)* (St. Louis, MO, USA). IEEE, Piscataway, NJ, USA, 96–104. https://doi.org/10.1109/PERCOM.2015.7146515

[72] Tadashi Okoshi, Julian Ramos, Hiroki Nozaki, Jin Nakazawa, Anind K. Dey, and Hideyuki Tokuda. 2015. Reducing users' perceived mental effort due to interruptive notifications in multi-device mobile environments. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan) *(UbiComp '15)*. ACM, New York, NY, USA, 475–486. https://doi.org/10.1145/2750858.2807517

[73] Tadashi Okoshi, Kota Tsubouchi, Masaya Taji, Takanori Ichikawa, and Hideyuki Tokuda. 2017. Attention and engagement-awareness in the wild: A large-scale study with adaptive notifications. In *2017 IEEE International Conference on Pervasive Computing and Communications, PerCom 2017* (Kona, HI, USA). IEEE, Piscataway, NJ, USA, 100–110. https://doi.org/10.1109/PERCOM.2017.7917856

[74] Tadashi Okoshi, Kota Tsubouchi, and Hideyuki Tokuda. 2019. Real-World Product Deployment of Adaptive Push Notification Scheduling on Smartphones. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (Anchorage, AK, USA) *(KDD '19)*. ACM, New York, NY, USA, 2792–2800. https://doi.org/10.1145/3292500.3330732

[75] Kevin Patrick, Fred Raab, Marc A Adams, Lindsay Dillon, Marian Zabinski, Cheryl L Rock, William G Griswold, and Gregory J Norman. 2009. A Text Message-Based Intervention for Weight Loss: Randomized Controlled Trial. *Journal of Medical Internet Research* 11, 1 (Jan. 2009), e1. https://doi.org/10.2196/jmir.1100

[76] Veljko Pejovic and Mirco Musolesi. 2014. InterruptMe: designing intelligent prompting mechanisms for pervasive applications. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Seattle, Washington) *(UbiComp '14)*. ACM, New York, NY, USA, 897–908. https://doi.org/10.1145/2632048.2632062

[77] Veljko Pejovic, Mirco Musolesi, and Abhinav Mehrotra. 2015. Investigating The Role of Task Engagement in Mobile Interruptibility. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct* (Copenhagen, Denmark) *(MobileHCI '15)*. ACM, New York, NY, USA, 1100–1105. https://doi.org/10.1145/2786567.2794336

[78] Pew Research Center. 2024. Demographics of Mobile Device Ownership and Adoption in the United States. Retrieved January 31, 2024 from https://www.pewresearch.org/internet/fact-sheet/mobile/

[79] Xuan-Lam Pham, Thi-Huyen Nguyen, Wu-Yuin Hwang, and Gwo-Dong Chen. 2016. Effects of Push Notifications on Learner Engagement in a Mobile Learning App. In *2016 IEEE 16th International Conference on Advanced Learning Technologies (ICALT)* (Austin, TX, USA). IEEE, Piscataway, NJ, USA, 90–94. https://doi.org/10.1109/ICALT.2016.50

[80] Martin Pielot, Bruno Cardoso, Kleomenis Katevas, Joan Serrà, Aleksandar Matic, and Nuria Oliver. 2017. Beyond Interruptibility: Predicting Opportune Moments to Engage Mobile Phone Users. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 91 (Sept. 2017), 25 pages. https://doi.org/10.1145/3130956

[81] 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* (Osaka, Japan) *(UbiComp '15)*. ACM, New York, NY, USA, 707–718. https://doi.org/10.1145/2750858.2805840

[82] Tauhidur Rahman, Mary Czerwinski, Ran Gilad-Bachrach, and Paul Johns. 2016. Predicting "About-to-Eat" Moments for Just-in-Time Eating Intervention. In *Proceedings of the 6th International Conference on Digital Health Conference* (Montréal, Québec, Canada) *(DH '16)*. ACM, New York, NY, USA, 141–150. https://doi.org/10.1145/2896338.2896359

[83] Ian Renfree, Daniel Harrison, Paul Marshall, Katarzyna Stawarz, and Anna Cox. 2016. Don't Kick the Habit: The Role of Dependency in Habit Formation Apps. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI EA '16)*. ACM, New York, NY, USA, 2932–2939. https://doi.org/10.1145/2851581.2892495

[84] Stephanie Rosenthal, Anind K Dey, and Manuela Veloso. 2011. Using Decision-Theoretic Experience Sampling to Build Personalized Mobile Phone Interruption Models. In *Pervasive Computing*, Kent Lyons, Jeffrey Hightower, and Elaine M Huang (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 170–187. https://doi.org/10.1007/978-3-642-21726-5_11

[85] Kathryn M Ross, Peihua Qiu, Lu You, and Rena R Wing. 2019. Week-to-week predictors of weight loss and regain. *Health Psychology* 38, 12 (Dec. 2019), 1150–1158. https://doi.org/10.1037/hea0000798

[86] Miroslav Rusin, Eirik Årsand, and Gunnar Hartvigsen. 2013. Functionalities and input methods for recording food intake: A systematic review. *International Journal of Medical Informatics* 82, 8 (Aug. 2013), 653–664. https://doi.org/10.1016/j.ijmedinf.2013.01.007

[87] Alireza Sahami Shirazi, Niels Henze, Tilman Dingler, Martin Pielot, Dominik Weber, and Albrecht Schmidt. 2014. Large-scale assessment of mobile notifications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. ACM, New York, NY, USA, 3055–3064. https://doi.org/10.1145/2556288.2557189

[88] Prasanta Saikia, Ming Cheung, James She, and Soochang Park. 2017. Effectiveness of Mobile Notification Delivery. In *2017 18th IEEE International Conference on Mobile Data Management (MDM)* (Daejeon, South Korea). IEEE Computer Society, Los Alamitos, CA, USA, 21–29. https://doi.org/10.1109/MDM.2017.14

[89] Nazir Saleheen, Amin Ahsan Ali, Syed Monowar Hossain, Hillol Sarker, Soujanya Chatterjee, Benjamin Marlin, Emre Ertin, Mustafa al'Absi, and Santosh Kumar. 2015. puffMarker: a multi-sensor approach for pinpointing the timing of first lapse in smoking cessation. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan) *(UbiComp '15)*. ACM, New York, NY, USA, 999–1010. https://doi.org/10.1145/2750858.2806897

[90] Hillol Sarker, Moushumi Sharmin, Amin Ahsan Ali, Md. Mahbubur Rahman, Rummana Bari, Syed Monowar Hossain, and Santosh Kumar. 2014. Assessing the availability of users to engage in just-in-time intervention in the natural environment. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Seattle, Washington) *(UbiComp '14)*. ACM, New York, NY, USA, 909–920. https://doi.org/10.1145/2632048.2636082

[91] Bruno M C Silva, Joel J P C Rodrigues, Isabel de la Torre Díez, Miguel López-Coronado, and Kashif Saleem. 2015. Mobile-health: A review of current state in 2015. *Journal of Biomedical Informatics* 56 (Aug. 2015), 265–272. https://doi.org/10.1016/j.jbi.2015.06.003

[92] Karlijn Sporrel, Rémi D D De Boer, Shihan Wang, Nicky Nibbeling, Monique Simons, Marije Deutekom, Dick Ettema, Paula C Castro, Victor Zuniga Dourado, and Ben Kröse. 2021. The Design and Development of a Personalized Leisure Time Physical Activity Application Based on Behavior Change Theories, End-User Perceptions, and Principles From Empirical Data Mining. *Frontiers in Public Health* 8 (Feb. 2021), 528472. https://doi.org/10.3389/fpubh.2020.528472

[93] S Shyam Sundar and Sampada S Marathe. 2010. Personalization versus Customization: The Importance of Agency, Privacy, and Power Usage. *Human Communication Research* 36, 3 (June 2010), 298–322. https://doi.org/10.1111/j.1468-2958.2010.01377.x

[94] Arun James Thirunavukarasu, Darren Shu Jeng Ting, Kabilan Elangovan, Laura Gutierrez, Ting Fang Tan, and Daniel Shu Wei Ting. 2023. Large language models in medicine. *Nature Medicine* 29, 8 (July 2023), 1930–1940. https://doi.org/10.1038/s41591-023-02448-8

[95] Roger Tourangeau. 2000. Remembering what happened: Memory errors and survey reports. In *The science of self-report: Implications for research and practice*, Arthur A Stone, Jaylan S Turkkan, Christine A Bachrach, Jared B Jobe, Howard S Kurtzman, and Virginia S Cain (Eds.). Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US, 29–47.

[96] Liam D. Turner, Stuart M. Allen, and Roger M. Whitaker. 2017. Reachable but not receptive: Enhancing smartphone interruptibility prediction by modelling the extent of user engagement with notifications. *Pervasive and Mobile Computing* 40 (Sept. 2017), 480–494. https://doi.org/10.1016/j.pmcj.2017.01.011

[97] Umelo A. Ugwoaba, Andrea N. Brockmann, Kelsey M. Arroyo, Taylor N. Swanson, Kelsey L. Barrett, Meena N. Shankar, Simone Marini, and Kathryn M. Ross. 2023. Behavioral weight loss success: Android vs. iOS. *Annals of Behavioral Medicine* 57 (April 2023), S201. https://doi.org/10.1093/abm/kaad011

[98] Leena Ventä, Minna Isomursu, Aino Ahtinen, and Shruti Ramiah. 2008. "My Phone is a Part of My Soul" - How People Bond with Their Mobile Phones. In *Proceedings of the 2008 The Second International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies (UBICOMM '08)*. IEEE Computer Society, USA, 311–317. https://doi.org/10.1109/UBICOMM.2008.48

[99] Shihan Wang, Karlijn Sporrel, Herke van Hoof, Monique Simons, Rémi D D de Boer, Dick Ettema, Nicky Nibbeling, Marije Deutekom, and Ben Kröse. 2021.

Reinforcement Learning to Send Reminders at Right Moments in Smartphone Exercise Application: A Feasibility Study. *International Journal of Environmental Research and Public Health* 18, 11 (June 2021), 6059. https://doi.org/10.3390/ijerph18116059

[100] Shihan Wang, Chao Zhang, Ben Kröse, and Herke van Hoof. 2021. Optimizing Adaptive Notifications in Mobile Health Interventions Systems: Reinforcement Learning from a Data-driven Behavioral Simulator. *Journal of Medical Systems* 45, 12 (Oct. 2021), 102. https://doi.org/10.1007/s10916-021-01773-0

[101] WHO Global Observatory for eHealth. 2011. *mHealth: new horizons for health through mobile technologies: second global survey on eHealth.* Technical Report. World Health Organization, Geneva, Switzerland. https://iris.who.int/handle/10665/44607

[102] Julia Woodward, Zari McFadden, Nicole Shiver, Amir Ben-hayon, Jason C. Yip, and Lisa Anthony. 2018. Using Co-Design to Examine How Children Conceptualize Intelligent Interfaces. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. ACM, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574.3174149

[103] Longqi Yang, Cheng-Kang Hsieh, Hongjian Yang, John P. Pollak, Nicola Dell, Serge Belongie, Curtis Cole, and Deborah Estrin. 2017. Yum-Me: A Personalized Nutrient-Based Meal Recommender System. *ACM Trans. Inf. Syst.* 36, 1, Article 7 (July 2017), 31 pages. https://doi.org/10.1145/3072614

[104] Elad Yom-Tov, Guy Feraru, Mark Kozdoba, Shie Mannor, Moshe Tennenholtz, and Irit Hochberg. 2017. Encouraging Physical Activity in Patients With Diabetes: Intervention Using a Reinforcement Learning System. *Journal of Medical Internet Research* 19, 10 (Oct. 2017), e338. https://doi.org/10.2196/jmir.7994

[105] Fengpeng Yuan, Xianyi Gao, and Janne Lindqvist. 2017. How Busy Are You? Predicting the Interruptibility Intensity of Mobile Users. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. ACM, New York, NY, USA, 5346–5360. https://doi.org/10.1145/3025453.3025946

[106] Renwen Zhang, Kathryn E. Ringland, Melina Paan, David C. Mohr, and Madhu Reddy. 2021. Designing for Emotional Well-being: Integrating Persuasion and Customization into Mental Health Technologies. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. ACM, New York, NY, USA, Article 542, 13 pages. https://doi.org/10.1145/3411764.3445771

[107] Mo Zhou, Yoshimi Fukuoka, Yonatan Mintz, Ken Goldberg, Philip Kaminsky, Elena Flowers, and Anil Aswani. 2018. Evaluating Machine Learning-Based Automated Personalized Daily Step Goals Delivered Through a Mobile Phone App: Randomized Controlled Trial. *JMIR mHealth and uHealth* 6, 1 (Jan. 2018), e28. https://doi.org/10.2196/mhealth.9117