Is More Always Better? Discovering Incentivized mHealth Intervention Engagement Related to Health Behavior Trends

NABIL ALSHURAFA, Northwestern University, USA
JAYALAKSHMI JAIN, Northwestern University, USA
RAWAN ALHARBI, Northwestern University, USA
GLEB IAKOVLEV, Northwestern University, USA
BONNIE SPRING, Northwestern University, USA
ANGELA PFAMMATTER, Northwestern University, USA

Behavioral medicine is devoting increasing attention to the topic of participant engagement and its role in effective mobile health (mHealth) behavioral interventions. Several definitions of the term "engagement" have been proposed and discussed, especially in the context of digital health behavioral interventions. We consider that engagement refers to specific interaction and use patterns with the mHealth tools such as smartphone applications for intervention, whereas adherence refers to compliance with the directives of the health intervention, independent of the mHealth tools. Through our analysis of participant interaction and self-reported behavioral data in a college student health study with incentives, we demonstrate an example of measuring "effective engagement" as engagement behaviors that can be linked to the goals of the desired intervention. We demonstrate how clustering of one year of weekly health behavior self-reports generate four interpretable clusters related to participants' adherence to the desired health behaviors: healthy and steady, unhealthy and steady, decliners, and improvers. Based on the intervention goals of this study (health promotion and behavioral change), we show that not all app usage metrics are indicative of the desired outcomes that create effective engagement. As such, mHealth intervention design might consider eliciting not just more engagement or use overall, but rather, effective engagement defined by use patterns related to the desired behavioral outcome.

CCS Concepts: • **Human-centered computing** → **Empirical studies in ubiquitous and mobile computing**; • **Information systems** → *Clustering*;

Additional Key Words and Phrases: engagement, adherence, mHealth, longitudinal study, multiple health behaviors, college students, clustering algorithm

Authors' addresses: Nabil Alshurafa, Northwestern University, 680 N Lake Shore Dr, Chicago, IL, 60611, USA, nabil@northwestern.edu; Jayalakshmi Jain, Northwestern University, 680 N Lake Shore Dr, Chicago, IL, 60611, USA, jayalakshmi.jain@northwestern.edu; Rawan Alharbi, Northwestern University, 680 N Lake Shore Dr, Chicago, IL, 60611, USA, rawan.alharbi@northwestern.edu; Gleb Iakovlev, Northwestern University, 680 N Lake Shore Dr, Chicago, IL, 60611, USA, g-iakovlev@northwestern.edu; Bonnie Spring, Northwestern University, 680 N Lake Shore Dr, Chicago, IL, 60611, USA, bspring@northwestern.edu; Angela Pfammatter, Northwestern University, 680 N Lake Shore Dr, Chicago, IL, 60611, USA, angela@northwestern.edu.

1 INTRODUCTION

Smartphone applications (apps) are increasingly used to provide real-time feedback and deliver intervention material to support positive health behavior changes [65]. The ease of access to information on mobile devices has increased the reach of health behavior change interventions and the potential for remote health monitoring [23, 38]. With the advent of moving behavioral interventions on to smartphone apps, patients can now have an intervention available all day and every day. While there are several proposed frameworks for designing interventions that might enhance adherence, few consider either evaluating engagement by the patient with the intervention or more specific to mHealth, evaluating multiple metrics of app use or exposure and dose of the mHealth intervention. These may be important mechanisms in the mHealth treatment pathway [4, 5, 73]. Studies involving participant experiences with technology, not only in mHealth, but in most studies involving participant interaction with technology, generally agree that engagement with the technology is essential in generating health behavior change [7, 12, 16, 17, 28, 44, 51, 52, 60, 63, 64, 77]. As such, there is a growing interest in examining ways to increase engagement in mHealth behavior interventions [24]. However, there is little guidance regarding how much engagement, or dose of intervention, is sufficient to elicit positive health behavior change.

There is great variability in the definition of 'engagement', ranging from a sequence of linked behaviors and usage patterns [6], a focused experience [66], a participant's subjective experience [75], to a willingness to direct attention to a specific activity [55]. Furthermore, engagement is sometimes referred to as the offline experience of the user, that is, a cognitive or reflective type of experience resulting from collection and integration of data that could lead to action to inform change. [3, 19, 20, 40]. In response, Perski et al's. [55] systematic review proposed that in the field of digital health behavior change interventions, engagement should incorporate objective and subjective measures. The objective measure is defined based on the use patterns measured using various tools including smartphones to track number of logins, time spent online, and the amount of content used during the intervention period [13, 17, 25, 42, 47]. It also included physiological measures using wearable sensors such as cardiac and electrodermal activity, as well as eye tracking to determine psychophysical measures. The subjective measure is defined based on self-report questionnaires measuring the levels of engagement with digital games and the intervention [8, 30, 32, 47].

Taki et al. [69] harmonized subjective and objective measures in an engagement index (EI). This index was evaluated using a smartphone app that suffered from technological issues that were experienced by participants in receiving their push notifications, and low satisfaction with the app, limiting its utility. This further expands the confusion regarding the conceptualization and measurement of engagement.

Some researchers have identified that more engagement may not always relate to better health outcomes, particularly when using a deficit technology or attempting to decrease a certain behavior such as smoking [11, 27]. Such evidence points to a need to identify patterns of intervention dose that elicit the desired response, be it a health behavior change or a specific clinical outcome. As it stands, most mHealth apps are built to deliver a behavior change intervention "at will" with little guidance on the optimal time, duration, and frequency of dose to achieve a desired effect [31, 62]. Thus, as mHealth apps proliferate, measuring what might be considered meaningful or effective engagement, that is, engagement that leads to a desired outcome, is necessary [46, 77].

To prevent confusion, we define the following terms as they relate to digital health behavior interventions:

- *Adherence*: Adherence can be defined as the level of compliance with the directives of the health intervention, independent of the mHealth tools measured by analyzing *proximal health outcomes* in hopes of obtaining a *desired distal health outcome* (e.g. weight loss) [36, 39, 43]. For example, in an intervention aimed at achieving the distal health outcome of weight loss, a proximal desired behavior might be achievement of 150 minutes of moderate-intensity activity, or 75 minutes of vigorous-intensity aerobic activity a week (with at-least 10 minutes per bout of activity) [22].
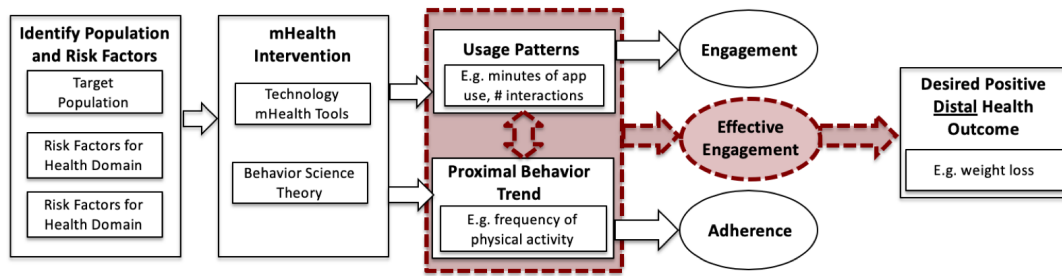
Fig. 1. Pathway to Positive Health Outcomes: To obtain a positive distal health behavior change.

- *Engagement*: Specific interaction and usage patterns with the mHealth tools might define engagement. For example, in an intervention delivered on an app, minutes of app use and number of app interactions per week are features or metrics that make up a proxy we can define as engagement. However, the presence of engagement defined by app use metrics does not ensure that the intervention dose is received as intended or that the individual is attending to the intervention in such a way that produces a desired effect. For example, the number of app uses per hour in a weight loss intervention is not meaningful by itself unless it correlates with the desired outcome. Hence, a definition of what constitutes meaningful engagement could be helpful in evaluating intervention and app effectiveness.
- *Effective Engagement*: We propose that effective engagement is mHealth tool *usage patterns* and behaviors that correspond to desired proximal health trends or outcomes. The reason we focus on proximal outcomes, is that distal measures often involve multiple complex factors beyond the control of studies. For example, in a weight loss intervention that is smartphone-based, effective engagement might be the number of days the participant uses the app (engagement metric), because it is most significantly correlated with adherence to physical activity recommendations (adherence metric). Discovering these metrics will allow interventionists to target usage patterns to encourage desired proximal health outcomes. It will also allow us to identify individuals that might need greater support in an intervention, based on usage patterns. Effective engagement, therefore, is not app use that is meaningless or use that does not correspond in the desired health behavior. By disentangling types of engagement, we may further design interventions that elicit, not just more app use without purpose or attention, but rather, app use that elicit engagement in the intervention in such a way that produces a desired effect.

While studies have sought to improve participant engagement through identifying patterns in usage [58, 70], we investigate how engagement, use patterns, and self-reported health-related behavioral trends are associated in order to identify effective engagement as defined above. Previous studies have measured and defined engagement by using coarse measures [35, 48], but with current technology, a variety of metrics can be used to identify patterns of interacting with an mHealth intervention app to investigate temporal and dynamic relationships between app use and proximal health behavior change. The objective of the current study is to provide a framework for future research to quantitatively and qualitatively build mediators to distal health outcomes, that is, to relate intervention app engagement behaviors with adherence to behavioral recommendations purported to lead to clinically meaningful outcomes. Thus, our primary goal is to use a data driven approach to define what it means to be effectively engaged with an mHealth intervention such that it results in a desired distal health outcome. Figure 1 illustrates the pathway to a distal positive health outcome. Some researchers study engagement with mHealth tools with the underlying assumption that engagement relates to distal positive health outcomes [6, 55, 69], and others study the relationship between mHealth tools and adherence, independent of engagement, under

the assumption that adherence effects distal health outcomes [21, 37, 54, 57, 72, 74]. Yardley et al. promote effective engagement by relaying how studies involving participant experiences with technology generally agree that engagement with technology is essential in generating health behavior change [77]. The goal of our work, comparatively, is to operationalize effective engagement by considering both engagement and adherence, together rather than independently, to build a foundation for effective engagement as a mechanism on a proposed pathway to the desired distal health outcome.

We examine engagement usage patterns and desired proximal health behaviors in the context of an ongoing college student health study where both app use metrics and multiple proximal health behaviors were measured. The NUYou study aimed to study the feasibility and acceptability of a multiple health behavior promotion and change intervention delivered during the first two years of college life. Though a period of the lifespan often neglected for health promotion, the time spent in college represents a unique epoch in that there is evidence of a significant drop in cardiovascular health [26]. Evidence suggests that maintaining a healthy lifestyle during these critical years may protect against cardiovascular disease risk factors [50, 67]. To test a mHealth mediated intervention, the NUYou trial enrolled Northwestern University students during their freshman year and cluster randomized by dormitory to a Whole Health (WH) intervention targeting four non-cardiovascular disease (non-CVD) related behaviors (e.g. hydration, safe sex, sun safety, and vehicular safety) or cardiovascular disease (CVD) health intervention targeting four CVD-related behaviors including diet quality, weight management, physical activity, and smoking. Since college students are relatively healthy compared to the population to start with [26], the study had a goal of both health promotion and health behavior change, to support health behaviors students already had, while changing health risk behaviors they had already acquired, or acquired during the trial. Thus, the desired outcome was two fold and we define the *desired proximal health trend* as college students that are healthy, remain healthy, and students that are unhealthy improve in health over time.

Specifically, we aim to:

- Given an incentivized mHealth intervention, identify clusters of participants with similar change in health behavior patterns over time along with their corresponding app use metrics that produce effective engagement linked to the intervention goals, particularly health promotion and behavioral change as defined by the American Heart Association.

The findings can inform mHealth researchers on what aspects of app use are critical to see desired proximal changes in health behavior during an intervention and provide a preliminary basis for defining effective engagement.

2 METHODS

The purpose of the NUYou study was to develop and test a scalable mHealth intervention to preserve and promote health behaviors in college students [56]. College freshmen were primarily recruited during their first fall on campus via letters, brochures, emails, banners, flyers, Facebook, information tables at events, and word of mouth from staff and faculty. Interested Freshmen were eligible if they owned an iPhone or Android smartphone and if they were living on campus. Participants were cluster randomized by dormitory into two groups: Cardiovascular Health group (CVH) and Whole Health group (WH). The target behaviors promoted in the CVH group were smoking, physical activity, fruit and vegetable consumption, and being in a healthy weight. Behaviors targeted in the WH group were hydration, sun, travel, and sexual safety. All participants were asked to complete a baseline health questionnaire documenting their health behaviors prior to the study, attend an in-person health assessment, and use a free smartphone application to receive intervention material and monitor their health behaviors. Participants were provided with results regarding their health assessment and were instructed on how to use the smartphone application. Participants attended an in-person session where they were asked to download and learn how to use a smartphone application customized to their randomized group, including application

functionality to orient the participant to each feature of the app, and weekly assessments. The primary method of intervention was through the NUYou smartphone application and was tailored to the behaviors targeted by their random assignment. Participants were also added to and asked to participate in a private study Facebook group that contained only members with the same random assignment.

2.1 Smartphone App Components

Questions	Possible Responses				
Q1: "How many fruits and vegetables did you have on average?"	0 - 1 servings	2 servings	3 servings	4 servings	≥ 5 servings
Q2: "How many minutes of physical activity did you do?"	0 - 30 minutes	30 - 60 minutes	60 - 90 minutes	90 - 150 minutes	> 150 minutes
Q3: "How many cigarettes did you have on average?"	≥ 7 cigarettes	5 - 6 cigarettes	3 - 4 cigarettes	1 - 2 cigarettes	0 cigarettes
Q4: "How many pounds are you over your threshold?"	≥ 7 pounds	5 - 6 pounds	3 - 4 pounds	1 - 2 pounds	0 pounds

Non-complaint behavior
Score = 0
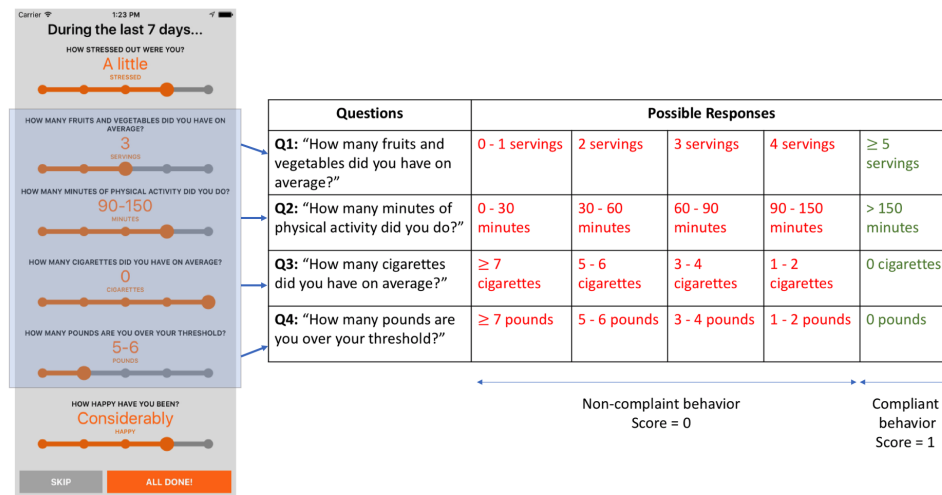
Compliant behavior
Score = 1

Fig. 2. NUYou app behavioral queries interface for CVH group is shown on the left, and table to the right shows the different responses each query can take. The responses in red are considered to be non-compliant behaviors, whereas the responses in green are considered to be compliant behaviors.

Core features of the app included weekly health behavior questions, screens to self-monitor daily health behaviors and review information about their progress in achieving behavior goals, a calendar function, and screens that included Facebook group content. Participants were trained on app use and asked to open and use the app daily.

2.1.1 Health Behavior Queries. On a weekly basis, participants were asked to complete very brief health behavior assessment surveys that determined where they ranked their adherence to their 4 target health behaviors as well as their level of stress and happiness. A positive response label is assigned to any answer that is above a numerical threshold that varies depending on the question, and was set in line with American Heart Association's Life's Simple Seven ideal criteria [1] as part of the NUYou study protocol. Only four of the six highlighted questions shown in Figure 2 and Figure 3 are associated with the health behavior outcome.

The other two questions relate to how stressed and happy the participants were, are not considered to be part of the proximal health behavior outcomes, but rather were used to allow students to track their stress and happiness. From the tables in Figures 2 and 3, we see that the responses in green are considered to be indicative of compliant behaviors whereas the responses in red are indicative of non-compliant behaviors. Responding to a query with a compliant behavior response gets a score of 1, otherwise it is scored 0. Then, the scores for each of the four questions are summed to give the behavioral query a final score between 0 and 4. For example, a participant from the CVH responded to the weekly queries as follows: [Q1: 60-90 minutes, Q2: >=5 servings, Q3: 0 cigarettes, Q4:
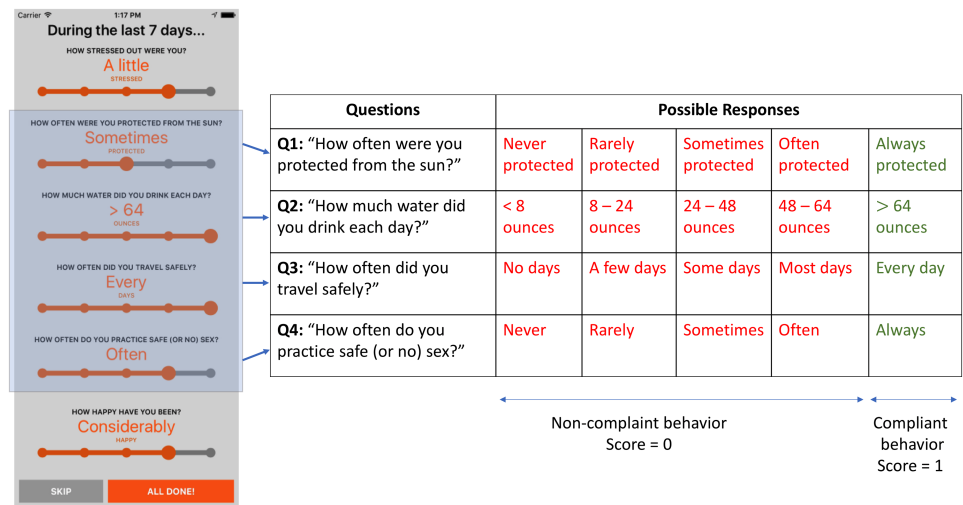
Fig. 3. NUYou app behavioral queries interface for WH group is shown on the left, and table to the right shows the different responses each query can take. The responses in red are considered to be non-compliant behaviors, whereas the responses in green are considered to be compliant behaviors.

1-2 pounds]. From this response set, only 2 responses are considered to be of compliant behavior (i.e Q2 and Q3). These two responses get a score of 1, bringing the score of the response set to 2. Consider another CVH participant who respond as follows: [Q1: >150 minutes, Q2: >=5 servings, Q3: 0 cigarettes, Q4: 0]. All 4 responses from this response set are considered to be of compliant behavior. Each of them get a score of 1, bringing the score of the response set to 4.

2.1.2 Self-tracking and Self-monitoring. For the present investigation, we analyzed data collected from 242 participants during a one year time period, the greatest amount of time when all enrolled participants used the second-generation health promotion app as previously described [56]. Both the second-generation CVH and WH version of the app had the ability to connect to student calendars allowing students to see class information and schedule time to do healthy behaviors to which they were assigned to monitor. Both groups could also track stress and happiness, play a cognitive speed game, and review their levels of behaviors, stress, happiness, and cognitive speed in a single graph at will as previously described [56]. Figure 4 shows some examples of the user interfaces for participants in the CVH group. Focus groups and student app design groups were established to optimize the look and feel of the CVH and WH apps.

While participants in the CVH group were able to track behaviors for which they were deemed to be at risk based on the answers to the weekly questions, WH participants could track water intake using the app module provided at all times. Thus, because the apps functioned somewhat differently for the groups, some analyses of user interactions have been performed separately for the groups. For the period of time included for the purpose of the present study, a loss aversion incentive protocol was used, where incentives are framed as losses instead of gains due to participants increased sensitivity to loss relative to gain [9]. Concern over retaining participants during the study lead to an implementation of the incentive structure which was collaboratively designed with student input [56].
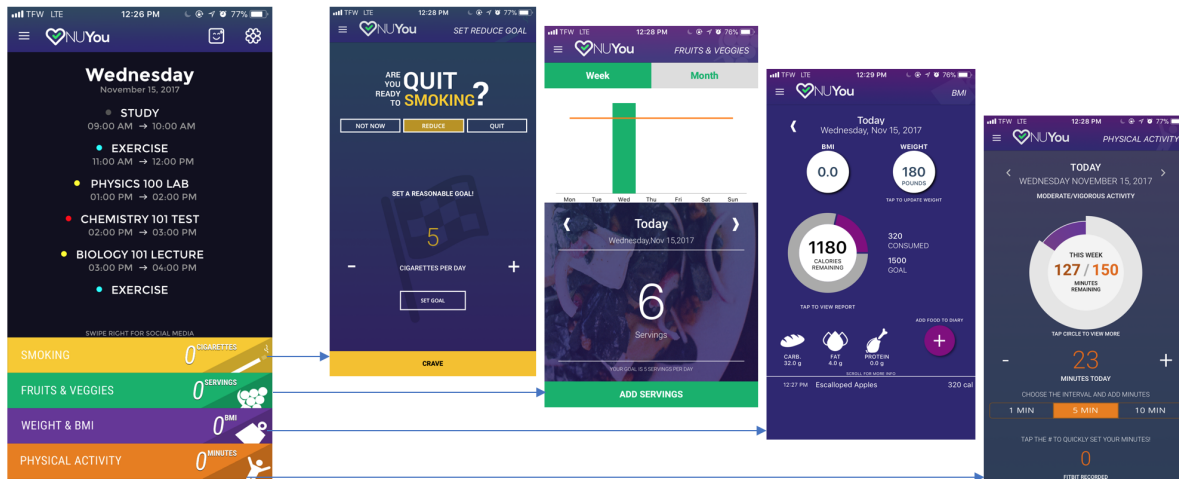
Fig. 4. NUYou app behavioral self-tracking and self-monitoring interface for participants in the CVH group.

The loss aversion incentive structure was implemented such that participants started with a bank of $110 from which $0.10 was removed for each day they did not open the app and an additional $0.10 was removed each day they did not enter some value into the app. Finally, participants had $0.70 removed each week they did not post something in their assigned Facebook group. Participants received a payout from their bank at each in-person assessment, which for the purpose of this investigation, occurred after the one year analysis period.

2.2 Data Pre-Processing

Although participants were prompted (a push notification was sent) to respond to weekly behavioral queries on Sundays, the data shows response times to the query varying throughout the week. From a behavioral standpoint, responses collected on a Sunday, Monday, or Tuesday are reflective of prior week's behavior, rather than the week that they were recorded in, and are matched to the appropriate week's data. Some of the user interactions with the app showed that the app was opened for less than a second, which were treated as a non-use (by setting interaction time to 0 seconds) due to the lack of an intentional interaction (i.e. participant might have accidentally opened the app).

3 ANALYSIS

The following subsections outline the details of how engagement patterns and adherence to health behavior were identified in college students using data collected from the NUYou app, primarily focusing on how we generate the usage patterns as well as the proximal behavioral trend clusters. Figure 5 provides a framework for the approach. We extract features from app usage and participants weekly health behavior query response to generate behavioral trend clusters that represent proximal behavioral trends, and single instance and weekly instance measures that represent usage patterns. The usage patterns are then used to describe and distinguish the different behavioral clusters by examining the contrasts of interest based on goals of the primary study.

The one year time period was chosen as the time frame when enrollment of new participants had ended and all participants were using the same app verison [56]. Additionally, by using the entire year long time frame available, we control for seasonality, and the academic schedule.
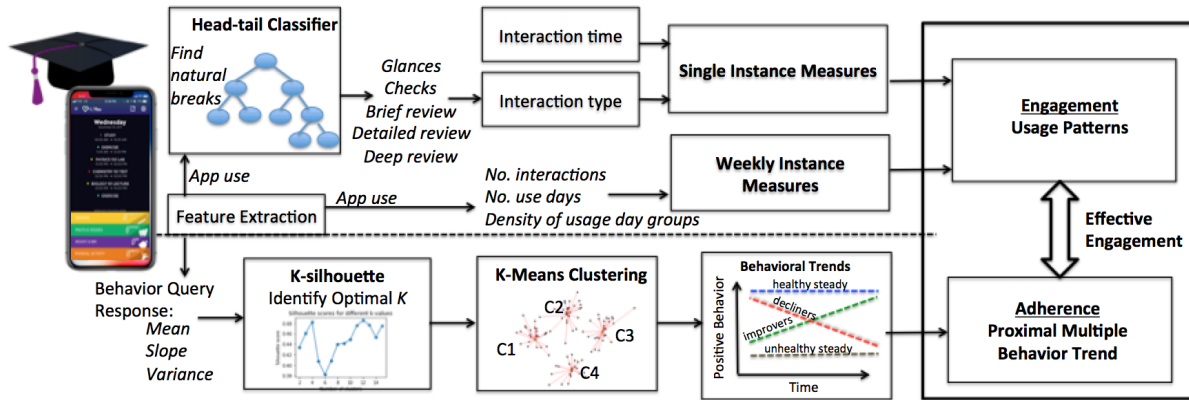
Fig. 5. This figure illustrates the framework for generating the usage patterns and proximal multiple health behavior trends.

Since health behavior queries occurred on a weekly basis, the app interaction data is processed in a weekly context. Hence, the analyses will explore how a particular week's use features are associated with a user's health behaviors for that week. We also define single interaction statistics such as seconds of a single interaction, and type of interaction that are used to distinguish between proximal behavioral trends.

Analysis is divided into two categories: 1) Unsupervised clustering to define behavioral trends that relate to adherence to health behaviors as prescribed by the random group assignment; and 2) Statistical measures related to single instance and weekly instance interactions with the app (usage patterns). As the health behaviors targeted in each group differed, we conduct these analyses first for the CVH group and then the WH group.

3.1 Health Behavior Adherence Trends

Health behavior trend is the time series trajectory of the participant's health behavior query response over a 1 year epoch. By analyzing participants longitudinally, we determine clusters of participants that demonstrate similar behaviors. Using these clusters of people, we can then identify groups of people that demonstrate similarity in the fluctuation in their health behaviors over time.

Three features, namely the slope, variance, and mean are used to quantitatively represent the weekly behavioral response data across 1 year for each participant. In context of the behavior trends, the slope signifies the general change in behavior over time, variance as the consistency of behavior, and the mean as the overall amplitude of the weekly behavior response across the study. These features are calculated and then normalized (between 0 and 1) for each group (CVH and WH) separately, allowing for comparison between the different groups. K-means clustering algorithm [41] is then performed on these 3 features, giving rise to different clusters of participants that demonstrate similar behaviors. The k-means algorithm is computationally efficient and produces tight clusters, given the right value of k. The clusters are analyzed further in order to identify the key characteristics of each cluster.

The optimal k-value for the number of clusters is determined by calculating the silhouette score [61], which is computed for a given range of k-values and evaluates which k value yields clusters that are most representative of the data that comprise them. The silhouette score quantifies (between -1 and +1) how similar a given data point is to its own cluster, and how dissimilar it is to the other clusters. A silhouette score closer to +1 indicates that the data point is well matched to its own cluster and poorly matched to its neighbors.

3.2 Engagement and App Use Patterns

Engagement and app use patterns refer to the type of interaction a participant experiences within a single or weekly interaction with the app and is primarily characterized by the amplitude (in seconds) of the interaction instance, or by frequency of interactions.

While analyzing single and weekly usage interactions provide us with a fine-grained picture of how a particular instance or week's data can affect the participant's response to the behavioral query, analyzing the relationship between these statistics and health behavior trends longitudinally for each participant allows us to identify engagement that is effective (i.e. related to desirable health behavior trends). Table 1 provides descriptions of the single and weekly instance measures used to understand engagement and app use patterns.

3.2.1 Single Instance Measures. To identify the different types of single instance engagement, a head/tail classifier was applied [33] (also used by Gouveia et al. [27] to find natural breaks in app usage) to determine natural breaks in the data, providing threshold cutoffs (in seconds) for varying levels of instance engagement. The head/tail classifier is especially suited for interaction amplitude data, as it is primarily meant for distributions that are prominently skewed toward having few large values (few long app interaction instances) and numerous smaller values (short app interaction instances). We see this in the interaction data from the NUYou app, where there are a large number of shorter interactions, and a fewer number of lengthy interactions. For this analysis, interaction lengths that were more than 3 standard deviations from the mean were considered as outliers and weren't included in the analysis.

3.2.2 Weekly Instance Measures. Weekly instance measures can be defined as the interaction of a participant over the course of an entire week. We use the week as a primary unit of time for engagement measurement due to the weekly nature of the behavioral query prompts for the participants.

Based on prior research [69], we identified variables of interest that have been shown to be effective in predicting engagement. We calculated features on a weekly basis, because our self-reported responses to their behavioral queries were reported on a weekly basis. Participants were reminded every Sunday to provide behavioral query responses. Participants also had the opportunity to self-monitor their data across the varying health behaviors, and could also report information to help them self-monitor their app use. Specifically, we investigated app usage data including: 1) number of interactions with the app per week which is associated with loyalty or frequency of use; 2) number of use days per week (weeks with no use days were not included); and 3) density of usage day groups (i.e. fraction of weekly instances where number of use days are 0-1 days, 2-4 days or 5-7 days). We calculate these features across both the CVH group and the WH group.

Measuring engagement on a weekly basis is beneficial for our analysis in two primary ways. First, we are able to identify factors of engagement over a longer period of time than the instance level, allowing us to begin identifying time series patterns in the data. Also, the week is the shortest amount of time that we are able to relate to a behavioral indicator, as the participants were asked to respond to behavioral queries every week that they were involved in the study.

Referring back to the idea of engagement as a process, analyzing both the trends within a given week and then analyzing the weeks of a participant's interactions with the app allow us to identify intermediate trends of engagement for the participants. In analyzing individual weeks in a participant's app usage, we primarily seek to identify factors that can be associated with the health behavioral query for that week. This allows us to measure engagement patterns that are also associated with the self-reported health behavior outcomes for a given week.

3.3 Contrasts

To determine defining features of effective engagement, we will perform 2 comparisons related to the two primary goals of the intervention. First, related to the *health promotion goal* of the intervention, we will test for differences

Table 1. Features extracted from the app interaction data.

Interaction Type	Features	Explanation
Single instance	Interaction length	The average length (in secs) of single instance interactions with the app.
	Interaction type (5 types based on the length of interaction)	The fraction of interactions labeled as Glances, Checks, Brief review, Detailed review and Deep review.
Weekly instance	Number of weekly interactions with the app	The average number of weekly interactions made by the participants in each behavioral cluster.
	Number of use days per week	The average number of days per week where the app was used by the participants in each behavioral cluster.
	Density of usage day groups divided into 3 categories: 0-1 days, 2-4 days, and 5-7 days	The fraction of weekly instances in a behavioral cluster where no. of use days is either 0-1, 2-4 or 5-7.

between those that had healthy behaviors and those who maintained them (desired outcome) and those that lost them (undesired outcome). Next, related to the *health behavior change goal* of the intervention, we will test for differences between those who had unhealthy behaviors at study start and improved them (desired behavior), and those who continued to report unhealthy behaviors throughout the intervention (undesired outcome). To perform these tests, we examine the 95% bootstrap confidence interval [18] to determine significant mean differences between the different behavioral trends.

4 RESULTS

4.1 Participants

The NUYou study enrolled a total of 303 participants. For this investigation on the second generation version of the intervention app, 242 participants were participating during the one year time period (130 in the CVH group and 112 in the WH group). The remaining 61 participants were excluded because they 1) either did not update to the second version of the app (n=39), 2) they did not answer at least two behavioral queries over the one year time period (n = 15), preventing the calculation of statistical features, or 3) they were considered significant outliers (more than 3 standard deviations from the mean). There were a total of 29,882 app interactions (all of them longer than 1 second) recorded during the one year time period, where all study participants were enrolled and using the same app. Total minutes of app use was 16338.40 minutes (Mean minutes per participant ± Std dev: 67.52 ± 61.08, Median: 40.43). Total number of weeks where participants interacted with the app at least once a week was 9155 (Mean Weeks per Participant ± Std dev: 37.83 ± 24.95, Median: 47). Total number of weekly behavioral queries completed were 7819 (Mean behavioral queries completed per participant ± Std dev: 32.31 ± 16.26, Median: 38 weeks). Table 2 summarizes this data and also provides statistics for the CVH and WH group. The mean interaction time per participant for the CVH group for the entire year is 83.85 minutes, which is significantly different than that of the WH group (48.55 minutes). There is a significant difference between the two group apps in terms of the health behaviors participants are managing, which further justifies the need to separate the analysis between the two groups.

Table 2. Statistics on participant interaction with the app, N=242

		Mean per participant (95% CI)	Median	Std dev	Total
All	App usage (in minutes)	67.52 (50.9, 87.2)	40.43	61.08	16338.40
	No. of weeks with at least 1 interaction	37.83 (30.9, 43.5)	47	24.95	9155
	No. of weekly behavioral queries completed	32.31 (26.0, 32.2)	38	16.26	7819
CVH	App usage (in minutes)	83.85*(67.1, 96.5)	58.94	75.53	10900.59
	No. of weeks with at least 1 interaction	41.52 (31.4, 49.9)	46	24.99	5397
	No. of weekly behavioral queries completed	32.98 (25.9, 33.9)	38	15.18	4287
WH	App usage (in minutes)	48.55*(39.3, 63.4)	33.51	54.75	5434.81
	No. of weeks with at least 1 interaction	33.55 (23.7, 44.0)	47	24.89	3758
	No. of weekly behavioral queries completed	31.53 (25.9, 31.9)	38	17.31	3532

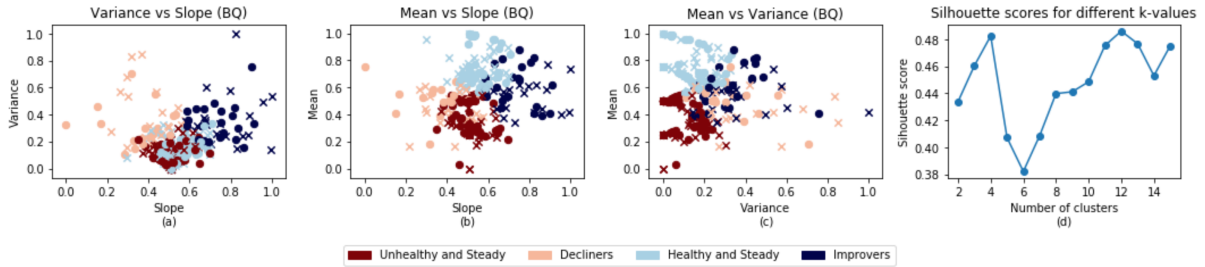* Average app usage time between the CVH and WH groups is significantly different.

Fig. 6. Clusters for the behavioral query response data obtained using the k-means algorithm. The circles indicate participants from the CVH group and crosses indicate participants from the WH group. a) Shows the general change in behavior over time (slope) vs the consistency (variance) of behavior over time. b) Shows the general change in behavior over time (slope) vs the overall amplitude (mean) of behaviors. c) Shows the consistency (variance) of behavior over time vs the overall amplitude (mean) of behavior. d) Silhouette plot showing the local optimal k value (k=4).

4.2 Trends in Health Behavior Adherence

Figure 6 illustrates the results of clustering the health behavior query responses. The clusters are colored according to the labels assigned by the k-means algorithm, and the circle or cross markers indicate whether the participant was in the CVH (circle) or WH (cross) group for the study respectively.

We identify four primary clusters of health behavior query trends. For ease of visualization, we flatten the three-dimensional clusters of participants into three subplots that show the participant clusters two-dimensionally. These three subplots are labeled and shown in Figure 6, along with a fourth subplot showing the silhouette plot. A peak in the silhouette plot indicates an optimal number of clusters where the data points within a cluster are close in proximity to each other and distant from neighboring clusters. It should be noted that higher values of k can create too many sparsely populated clusters. Since we are attempting to identify health behavior patterns that exist across entire groups of participants, we use local maximums (weighting lower values of k higher) as the best value for k.

Table 3. Characteristics of health behavior trend clusters. The observed avg ± Std dev of features used for clustering are indicated.

Cluster Color	Description	Mean	Variance	Slope [*]	Percent of Participants		
					CVH	WH	Total
	Decliners	Medium	High	Negative	14.62%	20.53%	17.35%
		0.47 ± 0.13	0.37 ± 0.19	0.38 ± 0.19			
	Healthy and Steady	High	Low	Constant	29.23%	34.82%	31.82%
		0.75 ± 0.12	0.14 ± 0.09	0.55 ± 0.08			
	Improvers	Medium	High	Positive	21.54%	17.86%	19.83%
		0.58 ± 0.14	0.36 ± 0.17	0.75 ± 0.11			
	Unhealthy and Steady	Low	Low	Constant	35.38%	25.89%	30.99%
		0.36 ± 0.13	0.1 ± 0.08	0.52 ± 0.07			

[*] For the slope, a normalized value lesser than 0.5 is considered negative slope, a normalized value greater than 0.5 is considered positive slope and a normalized value of approximately 0.5 is considered constant slope.

For the health behavior clusters, we identify the light blue cluster as primarily consisting of participants who demonstrate little variance or slope, indicating little change for them over the course of the study, but high mean, indicating these are people that begin with mostly positive health behaviors and maintain them over the course of the study. We later reference this group as the 'healthy and steady' group for their consistently high positive health behaviors. The maroon cluster demonstrates similar stability, but lower mean, indicating participants that begin with mostly negative health behaviors and maintain them, and we refer to this group as the 'unhealthy and steady' group. The dark blue cluster demonstrates higher variance, positive slope, and relatively high mean, indicating participants that trend upwards in their positive health behaviors. Because the slope is increasing and the mean is relatively high, we assign this cluster of participants as the 'improvers' group. The pink cluster comprises participants that are generally trending down in slope, possibly indicating the participants that are getting worse in terms of positive health behaviors over the course of the study, and are the 'decliners' group. Table 3 provides a summary of the characteristics for each behavioral cluster. Since all three variables have been normalized between 0 and 1, interpreting their average values is different, especially for the slope variable. A normalized value smaller than 0.5 is considered to have a negative slope, and a normalized value greater than 0.5 is considered to have a positive slope, while a normalized value of approximately 0.5 is considered to have a constant slope.

4.3 Loss of Incentives across the Different Health Behavior Clusters

Table 4 provides statistics on the incentives lost by participants on a weekly basis and across 1 year for not opening the app, not entering values and not posting on their assigned Facebook groups on a regular basis.

- Health Promotion: The average weekly loss of incentives by healthy-steady groups ($0.52) is slightly lower than that observed across the decliners ($0.57).
- Health Behavior Change: The improvers show a lower average loss of incentives on a weekly basis ($0.32) as compared to the unhealthy-steady group ($0.57).

4.4 CVH Group

4.4.1 CVH Health Behavior Cluster Overview. The unhealthy and steady cluster comprises participants with consistently negative health behaviors. While this group can be identified primarily based on the low variance,

Table 4. Statistics on the loss of incentives (in $) by the participants in each health behavior cluster.

	Weekly incentive loss		Incentive loss over 1 year	
	Mean (95% CI)	Std dev	Mean	Std dev
Decliners	-0.57 (-0.9, -0.3)	0.60	-29.71	31.23
Healthy and Steady	-0.52 (-0.7, -0.4)	0.57	-27.02	29.85
Improvers	-0.32 (-0.5, -0.1)	0.54	-16.51	28.32
Unhealthy and Steady	-0.57 (-0.8, -0.4)	0.63	-29.70	32.92

Table 5. Density of positive health behavior responses along with statistical features (0 means no positive health behaviors, and 4 means all positive health behaviors are met) for each behavioral trend cluster group in the **CVH** group.

	0	1	2	3	4	Mean (95% CI)	Std dev
Decliners	0.05	0.21	0.46	0.24	0.04	2.01 (1.9, 2.1)	0.89
Healthy and Steady	NaN	NaN	0.26	0.41	0.33	3.06 (3.0, 3.1)	0.77
Improvers	0.02	0.10	0.43	0.24	0.22	2.54 (2.4, 2.6)	1.01
Unhealthy and Steady	0.04	0.40	0.53	0.03	NaN	1.57 (1.5, 1.6)	0.63

slope, and mean of the response collected for the health behavior queries, intermediate engagement patterns can be identified that uniquely sets these participants apart. 97% of the responses to the health behavior queries completed by the participants in this cluster were 0, 1 or 2, which is significantly higher than what was measured in the other clusters as shown in Table 5. No participant from this group responded with a 4 (a perfect score) in the health behavior queries. This group has the lowest mean score (1.57) amongst all clusters.

The healthy and steady cluster show the highest mean score (3.06), with 74% of the responses recorded as either a 3 or 4, and no participant from the group responded with a 0 or 1 in the behavioral queries. The clusters which show changing behaviors, i.e. decliners and improvers, show a wider spread of scores in the behavioral queries and hence their mean scores are higher than the unhealthy-steady but lesser than the healthy-steady groups.

4.4.2 Single Instance Measures of Engagement. Table 7 provides some key usage pattern statistics (mean, median and standard deviation) on the single instance interaction length for participants in different behavioral clusters. There are no significant differences in the number of health behavior queries between the different CVH clusters, suggesting good representation of each of the clusters in terms of proximal behavioral trends (see Table 6). Applying the head/tail classifier on the interaction lengths after removing outliers result in five cutoff points. Based on the cutoff, we identify 5 types of single instance engagement, namely, glances (<7 sec), checks (7-21.7 sec), brief review (21.7-43 sec), detailed review (43-64 sec), and deep review (>64 sec). Table 8 shows how the participants of different behavioral clusters are distributed amongst different types of single instance engagement.

- Health Promotion: The decliners show higher average single interaction lengths (20.81 secs) as compared to the healthy-steady group (18.22 secs), as shown in Table 7. This distinction is significant. The head/tail classifier results show an interesting distinction between the decliners and healthy-steady groups in Table 8, in that the percent of glances observed in the decliners are higher than that observed in the healthy-steady groups (34% and 25% respectively). We also notice more time spent in brief-review in the healthy-steady group than the decliners (25% compared to 12%).

Table 6. Statistics on number of health behavior queries completed by the participants in each behavioral query cluster for the **CVH** group.

	Mean (95% CI)	Median	Std dev
Decliners	35.94 (25.0, 45.1)	39	14.23
Healthy and Steady	38.19 (32.4, 44.5)	42	15.34
Improvers	35.43 (25.8, 45.0)	35	15.27
Unhealthy and Steady	33.85 (27.4, 40.1)	35	15.08

Table 7. Statistics on single instance interaction length (in secs) per health behavior cluster for the **CVH** group.

	Mean (95% CI)	Median	Std dev
Decliners	20.81 (19.9, 21.6)	14.39	18.54
Healthy and Steady	18.22 (17.6, 18.8)	12.39	16.18
Improvers	20.79 (19.9, 21.5)	15.86	17.76
Unhealthy and Steady	19.14 (18.7, 19.9)	13.34	16.51

Table 8. Distribution of single instance interactions in each threshold group for each health behavior trend cluster in the **CVH** group.

	Glances < 7 sec	Checks 7 - 21.7 sec	Brief review 21.7 - 43 sec	Detailed review 43 - 64 sec	Deep review > 64 sec
Decliners	0.34	0.39	0.12	0.09	0.06
Healthy and Steady	0.25	0.36	0.25	0.09	0.05
Improvers	0.27	0.41	0.20	0.08	0.03
Unhealthy and Steady	0.38	0.35	0.12	0.10	0.05

- Health Behavior Change: The improvers show a higher average single instance interaction length (20.79 secs) as compared to the unhealthy-steady group (19.14 secs), as indicated in Table 7. This difference is significant. The percent of glances observed in the unhealthy-steady groups is higher than that observed in the improvers, 38% and 27%, respectively (as shown in Table 8). We also notice more brief review type of app interaction in improvers, 20% compared to 12% in the unhealthy-steady.

4.4.3 Weekly Instance Measures of Engagement. Figure 7 provides results on the weekly app usage data described in Table 1, which we further elaborate on below.

- Health Promotion: There is a significant difference in the average number of weekly app usages between the decliners and healthy-steady group (6.2 and 8.2 interactions per week on average respectively). Also, on comparing the number of use days per week between the 2 groups (4.0 and 4.6 days on average per week for the decliners and healthy-steady respectively), the difference was found to be significant. In addition, there is a significantly higher percent of weekly instances where the healthy and steady interacted with the app for 5-7 days (58%), compared to the decliners (44%).
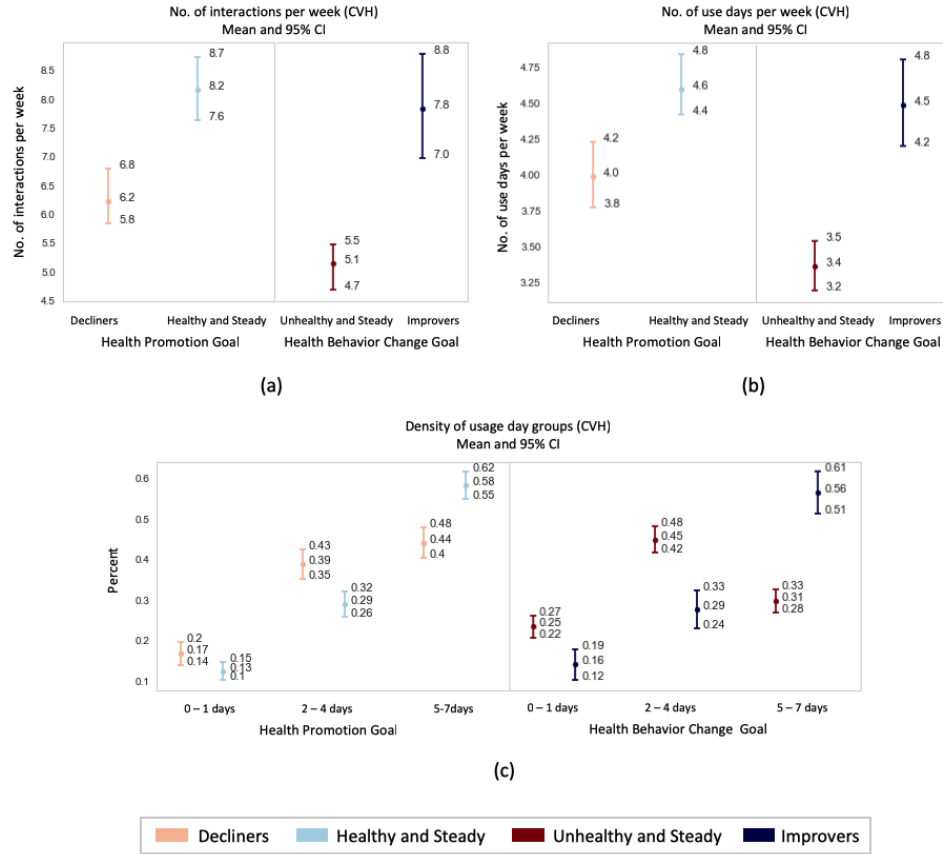
Fig. 7. Weekly app usage statistics for each behavioral trend clusters in the **CVH** group dividing the results between the health promotion and health behavior change goals. a) Number of interactions per week. b) Number of use days per week. c) Density of usage days for given thresholds.

- Health Behavior Change: A significant difference can also be observed in the average number of weekly app usages between the improvers and unhealthy-steady group (7.8 and 5.1 interactions per week respectively). Also, on comparing the number of use days per week between the 2 groups (4.5 and 3.4 days on average per week for the improvers and unhealthy-steady respectively), the difference was found to be significant. In addition, there is a significantly higher percent of weekly instances where the improvers interacted with the app for 5-7 days (56%), compared to the unhealthy-steady (31%).

4.5 WH Group

4.5.1 WH Health Behavior Cluster Overview. The unhealthy and steady cluster in the WH group shows similar intermediate engagement patterns that were observed in the CVH group. 98% of the responses to the behavioral queries completed by the participants in this cluster were 0, 1 or 2, which is significantly higher than what was

Table 9. Density of positive health behavior query responses (0 means no positive health behaviors, and 4 means all positive health behaviors are met) for each behavioral trend cluster group in the **WH** group

	0	1	2	3	4	Mean (95% CI)	Std dev
Decliners	0.10	0.36	0.33	0.18	0.04	1.71 (1.6, 1.8)	0.99
Healthy and Steady	NaN	0.02	0.17	0.59	0.21	3.00 (2.9, 3.1)	0.68
Improvers	0.02	0.25	0.34	0.34	0.05	2.15 (2.1, 2.3)	0.93
Unhealthy and Steady	0.07	0.56	0.35	0.02	NaN	1.32 (1.3, 1.4)	0.63

Table 10. Statistics on number of health behavior queries completed by the participants in each health behavior cluster for the **WH** group

	Mean (95% CI)	Median	Std dev
Decliners	33.37 (25.0, 46.1)	35	17.23
Healthy and Steady	37.51 (30.6, 44.0)	42	15.45
Improvers	28.19 (19.3, 39.3)	36	15.23
Unhealthy and Steady	29.81 (20.5, 37.4)	30	19.53

measured in the other clusters as shown in Table 9. None of the participants from this cluster responded with a 4 in the behavioral queries. This group also has the lowest mean score (1.32).

A mean score of 3.0 (highest amongst the 4 clusters) was observed for the healthy and steady clusters. 80% of the responses were a 3 or 4 and none of the participants from this group responded with a 0.

The clusters which show changing behaviors i.e. decliners and improvers show a wider range of responses in the behavioral queries and hence their mean scores are higher than the unhealthy-steady but not as high as the healthy-steady group. It is consistent with the definitions that the improvers and healthy-steady groups would have higher behavioral query responses than the decliners and unhealthy-steady groups.

4.5.2 Single Instance Measures of Engagement. There are no significant differences in the number of health behavior queries between the different WH clusters, suggesting good representation of each of the clusters in terms of proximal behavioral trends (see Table 10). Table 11 provides some key statistics (mean, median and standard deviation) on the single instance interaction length for participants in different behavioral clusters. Applying the head/tail classifier on the interaction lengths, Table 12 shows how the participants of different behavioral clusters are distributed amongst different types of single instance engagement. We elaborate on the two tables below as they relate to each of the two health behavior goals.

- Health Promotion: The decliners show lower average single interaction length (19.26 secs) as compared to the healthy-steady group (20.92 secs). This distinction is significant (see Table 11). The head/tail classifier results show an interesting distinction between the decliners and healthy-steady (see Table 12), in that the percent of glances observed in the decliners is higher than that observed in the healthy-steady groups (36% and 28% respectively). We notice more time spent in brief-review in the healthy-steady group than the decliners (29% compared to 14%).
- Health Behavior Change: The improvers show a higher average single instance interaction length (21.92 secs) as compared to the unhealthy-steady group (16.66 secs). This difference is significant. The percent of glances observed in the unhealthy-steady groups is higher than that observed in the improvers (37% and

Table 11. Statistics on single instance interaction length (in secs) per health behavior cluster for the **WH** group.

	Mean (95% CI)	Median	Std dev
Decliners	19.26 (18.4, 20.2)	15.34	15.80
Healthy and Steady	20.92 (20.1, 21.6)	15.99	17.01
Improvers	21.92 (20.7, 23.4)	18.04	18.15
Unhealthy and Steady	16.66 (15.8, 17.4)	12.93	14.41

Table 12. Distribution of single instance interactions in each threshold group for each health behavior trend cluster in the **WH** group.

	Glances < 7 sec	Checks 7 - 21.7 sec	Brief review 21.7 - 43 sec	Detailed review 43 - 64 sec	Deep review > 64 sec
Decliners	0.36	0.39	0.14	0.08	0.03
Healthy and Steady	0.28	0.27	0.29	0.11	0.04
Improvers	0.26	0.32	0.28	0.10	0.03
Unhealthy and Steady	0.37	0.42	0.12	0.07	0.03

26% respectively). We also notice more brief review type of app interaction in improvers, 28% compared to 12% in the unhealthy-steady.

4.5.3 Weekly Instance Measures of Engagement. Figure 8 provides results on the weekly app usage data described in Table 1, which we further elaborate on below.

- Health Promotion: Figure 8 shows a significant difference in the average number of weekly app usages between the decliners and healthy-steady group (3.7 and 5.3 interactions per week on average respectively). Also, on comparing the number of use days per week between the 2 groups (2.6 and 3.8 days on average per week for the decliners and healthy-steady respectively), the difference was found to be significant. Also, there is a higher percentage of weekly instances where the healthy-steady interacted with the app for 5-7 days (41%), compared to the decliners (36%).
- Health Behavior Change: There is a higher percentage of weekly instances where the improvers interacted with the app for 5-7 days (47%), compared to the unhealthy-steady (18%). This difference was significant. A similar pattern can be observed with the number of interactions and number of use days per week but the differences aren't significant.

5 DISCUSSION AND FUTURE WORK

The purpose of the current investigation was to use a data driven approach to identify engagement in app delivered interventions that was meaningfully associated with health behavior. By doing so, mHealth researchers and interventionists may find insight into empirically supported mHealth design choices and dose recommendations that drive effects [68]. We sought to elucidate app use patterns that were related to health behavior patterns desired by the intervention to disentangle app use that was not resulting in the desired effects. We used data from a unique multiple health behavior promotion and change intervention in a college student population. The study provided intensive, daily and weekly level data on app use and health behaviors for a year where 242 students were using the same app version. With this intensity of data, we were able to cluster participants into categories
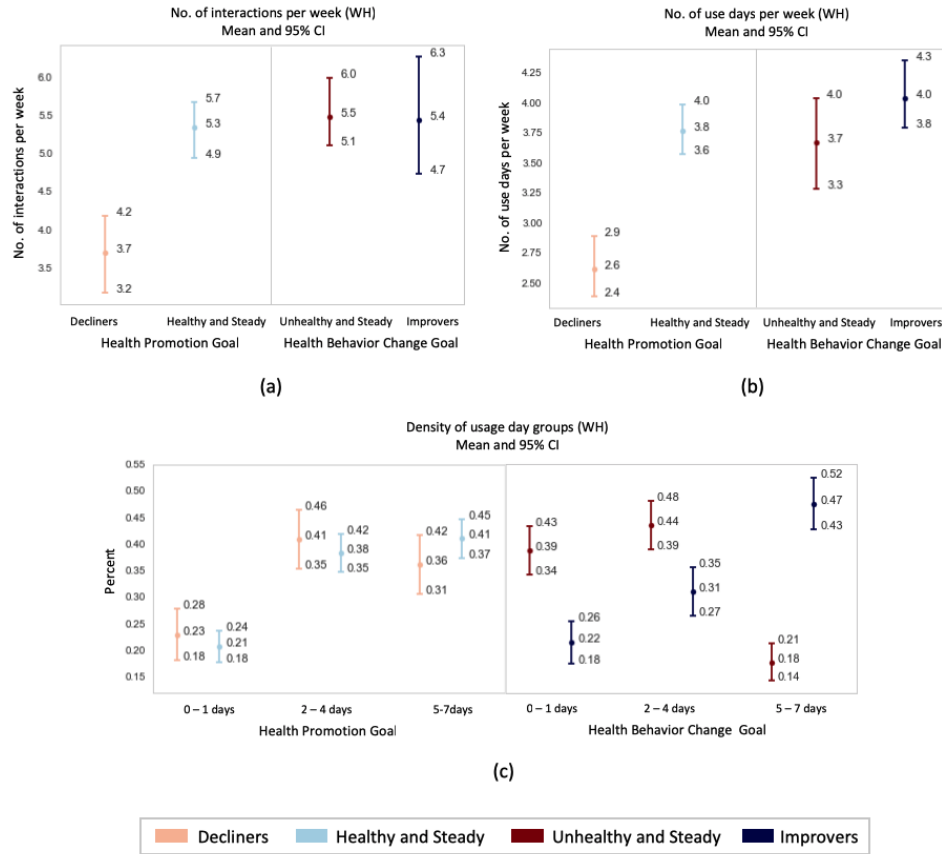
Fig. 8. Weekly app usage statistics for each behavioral trend clusters in the **WH** group dividing the results between the health promotion and health behavior change goals. a) Number of interactions per week. b) Number of use days per week. c) Density of usage days for given thresholds.

based on health behavior adherence patterns that occurred over time. Indeed, 4 patterns arose: healthy-steady, improvers, decliners, and unhealthy-steady. As students tend to start college with healthier behaviors than the general population, the goal of the study intervention delivered by smartphone app was both promotion and change of health behaviors. Hence, we investigated contrasts that would provide differentiating features of app use to support health promotion and multiple health behavior change. As the study randomized participants by dormitory to two groups that targeted 4 different health behaviors, we had the opportunity to replicate our findings from one group in the other.

5.1 Effective Engagement for Health Promotion

Our clusters in the CVH group revealed two groups that would primarily benefit from health promotion strategies; essentially those participants who began the study in a healthy state. The two clusters, healthy-steady and decliners, showed some significant differences that could identify when a participant is effectively engaged in

a health promotion intervention. First, single interaction time was higher in the decliners, yet showed more instances of glancing (short app interaction type, less than 7 seconds) when compared to the reviewing (longer interaction types) that was occurring in the healthy-steady cluster. Such a pattern could be indicative of a reflective process occurring in the healthy-steady participant that others have purported to support health behaviors [40]. Li et al. describes a five stage model that emphasizes the goal of knowledge supporting behavior change including preparation, collection, integration, reflection, and action [40]. Using such a model, we might conclude that while the decliners might have been collecting data, it may not have lead to a reflective process that leads to the type of effective action that supports positive health behavior maintenance. Furthermore, our data suggest that decliners were using the app a fewer number of times per week, for fewer days and are less likely to show 5-7 days of use leading us to conclude that not only is app interaction length important to distinguish those likely to benefit, but also that the number of interactions is an important indicator as well to maintain positive health behavior. Our clustering algorithm resulted in a greater percentage of healthy-steady (31.82%) than decliners (17.35%). Collectively, these findings are inconsistent with others suggesting that those who have positive health behaviors are more unlikely to spend more time in an intervention app [27]. However, in our studies we do incentivize participants, which may be one reason that encourages health promotion, and a greater number of healthy-steady participants than decliners. Our clusters in the WH group were not wholly consistent with the findings in the CVH group, suggesting that patterns indicative of effective engagement might depend on the health behaviors targeted in the intervention. In WH, decliners did not demonstrate higher interaction time. These findings suggest that for health promotion of cardiovascular health behaviors, the number of uses might be more indicative of effective engagement than interaction duration, but that for health promotion of other behaviors like hydration and sun, sex, and travel safety, the interaction duration and consistency of use of the intervention app might be most indicative of effective engagement.

5.2 Effective Engagement for Multiple Health Behavior Change

For individuals who need to actively change a behavior, we might expect to see a different pattern of interaction with the intervention. Individuals that needed to change behaviors can reasonably be found in the unhealthy-steady and improver groups because both clusters by virtue of their behavioral clusters, started out with some unhealthy behaviors. In the CVH group, it is evident from the distribution of single instance interactions, that the unhealthy-steady exhibit more glances than the improvers. The improvers also have longer single instance interactions than the unhealthy-steady (20.79 secs and 19.14 secs respectively). The improvers are also characterized by relatively greater number of weekly interactions, weekly use days, and more instances where the participants used the app for 5-7 days in a week than the unhealthy-steady group. Our clusters in the WH group were not wholly consistent with the findings in the CVH group. The WH group noted no significant difference in the number of weekly interactions and use days between the improvers and unhealthy-steady groups. This indicates that for behavior change, single interaction time and consistent 5-7 day app use a week are important and effective engagement may present similarly no matter the intervention target health behavior. Based on this, we can postulate that the engagement patterns present in the improver cluster are characteristic of participants that are trying to develop healthy behavior patterns, by having long interactions on a frequent basis to track their progress.

5.3 Measuring Engagement

While not all studies necessarily collect the intermediate behavioral self-report data that NUYou did [6, 53], the interaction data from the app can still allow other studies to potentially identify participant trajectories of overall health behavior. With the ultimate goal of identifying engagement patterns in college students that lead to positive or negative health behaviors and intervening appropriately, in the future, we aim to predict the

behavioral outcomes for a given week based on features obtained from the interaction data for the associated week in real time. Such predictors could empower interventionists to support engagement patterns that are associated with positive proximal and distal health behavior and clinical outcomes. With the results of our analysis, we can determine patterns of behaviors and app use that are indicative of effective engagement with the app, or engagement that is also related to health behavior change. As such, we can use patterns of app use as a proxy for an offline process of cognition, reflection, and/or learning that takes place to induce a process of behavior change. However, this offline process and it's connection to app use will need to be further investigated in future studies. Nevertheless, using a definition of effective engagement that is defined by the effect it has on health behaviors of interest may provide future research additional targets of intervention to incentivize proper amounts and patterns of engagement that is hypothesized to be supportive of the desired outcome. Our results may not deviate dramatically from traditional conceptualization of engagement, where more frequent interactions indicate better engagement, which is hypothesized to lead to better health behaviors. We show here that positive health behavior change can be characterized by relatively lengthy and numerous interactions that occur over 5-7 days in a given week. In this sense, more use is associated with more effective engagement in the context of participants that improve over time. However, the results demonstrating that improvers were uniquely using the app, not just more frequently, but more consistently over the days of a week, leads us to conclude that, at least for an intervention delivered like our CVH app, consistent use, not just more use, should be recommended.

5.4 Clustering

k-means clustering, used to generate our 4 health behavior adherence patterns, is known to be computationally efficient and produces tight clusters, given the right value of k, which we optimize. However, in other techniques like hierarchical clustering, which we aim to test in the future, k does not need to be predefined in advance [49]. We also deploy a hard clustering technique, where each participant belongs to only one of four clusters. However, a participant may actually belong to several clusters where they may exhibit healthy-steady patterns for some time, but partially a decliner pattern for another. This may be captured through a more soft-clustering technique that calculates the likelihood of belonging to different clusters. Such a technique could benefit our ability to understand dynamic behaviors and patterns that occur within a participant. Moreover, participants may belong to different clusters at different time points, throughout the year, where at one time point they are decliners, but then improvers, and then they are healthy-steady. Future work will analyze these transitions more closely which may uncover engagement patterns that are effective more proximally. Moreover, we mainly cluster based on variance, mean and slope. Other variables may be equally important in distinguishing participants, such as their max and min value (to separate participants in different stages of initial health).

5.5 Loss Aversion Incentives

Loss aversion is based on the assumption that an individual is motivated more by losses or disadvantages than by gains and advantages [71]. Using a loss aversion strategy as a way to encourage participation in a longitudinal research trial and to change health behaviors has been used successfully in previous research [29, 34]. In the case of the present study, dollars to devote to participant incentives were scarce. However, the behavior of opening the app daily, entering a single data point daily, and posting to the social group weekly were small frequent behaviors that we could construct an incentive structure around whereby earning money may not seem valuable, but losing a small amount of money might be aversive. While we would expect this strategy to enhance app engagement, we expected it to be equal across all participants. Indeed, despite, the incentive structure, we saw differential app use patterns between the health behavior change clusters indicating that while incentives may have minimally kept participants engaged in the study, there was enough variability and heterogeneity in the patterns that cannot

be explained by the incentive alone. However, future research might compare various incentive structures to determine if there is an effect of incentive on app use and also on the behavioral outcome differentially. Combined with further insight into the optimal use patterns of a behavior change intervention app, a suggested dose with a corresponding incentive might be particularly potent.

5.6 How Engagement and Adherence Are Related to Reflection and Learning

Recent human computer interaction research has proposed a process of reflection that an individual experiences when tracking and reviewing data regarding health behaviors. The premise is that reflection leads to an awareness of the behavior or self knowledge which is hypothesized to lead to behavioral action. [14]. Indeed, the patterns of engagement we identify could be indicative of a process of reflection, and as such, the differential patterns of app use and interaction between the improvers and healthy steady clusters in particular could correspond to the discovery and maintenance phases of reflection. However, whereas the HCI literature has focused on samples that are interested or already engaged in their own personal data collection, we might consider a social cognitive perspective for understanding the process by which an individual engages in an intervention to promote or change health as "pushed" from an institution, healthcare entity or the like. Using theories of self-regulation, the app was built to facilitate processes that bolster self-regulation such that participants could change or maintain healthy behaviors [2, 10]. This approach is not wholly inconsistent with conceptualizations of reflection, and could be seen as a way to answer the question as posed by Baumer [3] which is, "in what ways are people self-reflecting?" The app provides tools that present health behavior goals, display discrepancies to those goals, and allow the participant to self-monitor and track their behavior. These processes are also proported to facilitate an increase in self-efficacy to perform the desired behavior and adhere to the behavioral recommendations.

Indeed, while some apps have reflection as the end goal [76], in health-related interventions, behavior change is a desired goal. Reconciling theories of reflection from the perspective of self-initiated engagement with behavior change techniques and theories that target those who necessarily need to change behavior may be the next step in the future of this work [45]. We might also consider the role of learning in a longitudinal study. Learning can be facilitated by a feedback loop created by an mHealth application. Unfortunately, many mHealth apps do not directly test whether the use of an app translates into significant learning about the behaviors it targets. However, it stands to reason that health tools that leverage real time, personally relevant feedback loops could support the individual in a learning process that results in greater behavior change over time. Both self tracking and behavior change is dynamic and the ways in which individuals engage or disengage in these activities is an important area of consideration if we want to fully leverage technology to facilitate processes of behavior change to improve the health of the population [59].

5.7 Limitations

The present investigation is not without limitations. First, it will be critical to test the effects on downstream outcomes of interest, in this case, the primary outcome of AHA's Life's Simple Seven Scores at the annual assessment. It should be noted that the outcome of interest in the analyses presented within are comprised of self-reported health behaviors, known to be biased, inaccurate, and at some points missing. The reporting differences, however, did not seem to differ between groups, and the response rate to the weekly behavioral queries were equitable across groups and clusters. The study also did not look at variation in the health behaviors that did not meet AHA's Life's Simple Seven Scores (e.g. variation in number of cigarettes smoked, gaining 5 pounds compared to 10 pounds). This paper focuses on aggregate analyses, and in the future we aim to consider temporal relationships to improve our understanding of factors that correspond to proximal behavior changes. This study was chosen as a testbed in which to investigate the phenomenon of effective engagement primarily because there were multiple app use features collected and information about health behaviors were collected

weekly over a one year time period. The glaring caveat is that the NUYou study was conducted on college students and the findings may not generalize to the general population. Despite such shortcomings, the framework for analyzing such data could be of value to other mHealth researchers. Finally, participants that provided a single time point were excluded from the present analyses and as such, we cannot draw conclusions about those that fail to engage with our mHealth interventions from the start. Due to this limitation we are unable to confirm prior work demonstrating that low and sporadic engagement can actually result in improved health outcomes [15]. Further investigations are necessary to conceptualize for whom this is a problem.

5.8 Future Work

In future work, we aim to analyze the the relationship between the behavioral trends and positive health outcomes at the study endpoints. If these behavioral trends align with positive health outcomes, then they may act as proximal outcomes, or could help inform future just in-time interventions to identify when participants need further support. Given the somewhat discrepant findings between the CVH and the WH apps, future work will focus on whether our findings generalize to other smartphone-based health behavior interventions. Overall, this line of investigation could provide mHealth researchers with necessary information on which to base dose recommendations for multiple health behavior change and promotion interventions.

6 CONCLUSION

Our analysis provides insight into the intermediate engagement patterns that are related to longitudinal behavioral trends. Relating patterns of engagement to quantifiable behavioral trends identifies participants that are engaged in a way that is both quantitatively verifiable with app interaction data and is shown to be associated with behavioral trajectories. For the purpose of identifying the different behavioral trends exhibited by the participants, clustering was performed using the mean, slope, and variance of the responses to the weekly behavioral queries. Four clusters of participants emerged: healthy and steady; unhealthy and steady; decliners, and improvers. In the CVH group, clusters that would benefit from health promotion strategies (i.e decliners and healthy-steady) showed some significant differences in the app usage metrics. The healthy-steady group showed higher number of weekly interaction, number of use days and percent of weekly instances of app use for 5-7 days than the decliners. The decliners showed higher percent of glances and longer interaction times with the app. However with the WH group, the decliners exhibited shorter interactions than the healthy-steady. In the CVH group, for the clusters where individuals needed to change behaviors (i.e. unhealthy-steady and improver), significant differences were observed in the app usage metrics. The improvers showed longer single instance interaction, lower percent of glances, higher number of weekly interactions, weekly use days and percent of weekly instances of app use for 5-7 days than the unhealthy-steady group. With the WH group though, the number of weekly interactions and weekly use days do not show a significant difference between the 2 clusters. Thus, there could be a greater need for investigations such as the current one, to be carried out during the development and pilot phase of mHealth interventions. These results suggest analyzing a single user interaction variable or summary variable such as total use time across an entire study, may not be the most valuable approach in predicting interim health behavior change in the context of an mHealth intervention. Rather, results from this study suggest a variety of app interaction features could be more useful in discriminating between health behavior patterns during the course of an intervention. As such, considerations for mHealth design that supports effective engagement for a particular desired health behavior outcome could be warranted. Future studies should empirically test whether mHealth intervention designs can change engagement patterns to be more adaptive and supportive of proximal health behavior change.

ACKNOWLEDGMENTS

REFERENCES

[1] American Heart Association. 2018. *My Life Check - Life's Simple 7.* http://www.heart.org/en/healthy-living/healthy-lifestyle/my-life-check--lifes-simple-7 [Accessed: 15-Nov-2018].

[2] Albert Bandura. 1989. Human agency in social cognitive theory. *American psychologist* 44, 9 (1989), 1175.

[3] Eric P.S. Baumer, Vera Khovanskaya, Mark Matthews, Lindsay Reynolds, Victoria Schwanda Sosik, and Geri Gay. 2014. Reviewing Reflection: On the Use of Reflection in Interactive System Design. In *Proceedings of the 2014 Conference on Designing Interactive Systems (DIS '14).* ACM, New York, NY, USA, 93–102. https://doi.org/10.1145/2598510.2598598

[4] Marshall H Becker. 1990. Theoretical models of adherence and strategies for improving adherence. (1990).

[5] Albert J Bellg, Belinda Borrelli, Barbara Resnick, Jacki Hecht, Daryl Sharp Minicucci, Marcia Ory, Gbenga Ogedegbe, Denise Orwig, Denise Ernst, and Susan Czajkowski. 2004. Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the NIH Behavior Change Consortium. *Health Psychology* 23, 5 (2004), 443.

[6] Dror Ben-Zeev, Emily A Scherer, Jennifer D Gottlieb, Armando J Rotondi, Mary F Brunette, Eric D Achtyes, Kim T Mueser, Susan Gingerich, Christopher J Brenner, Mark Begale, et al. 2016. mHealth for schizophrenia: patient engagement with a mobile phone intervention following hospital discharge. *JMIR mental health* 3, 3 (2016).

[7] Felicity A. S. Bright, Nicola M. Kayes, Linda Worrall, and Kathryn M. McPherson. 2015. A conceptual review of engagement in healthcare and rehabilitation. *Disability and Rehabilitation* 37, 8 (2015), 643–654. https://doi.org/10.3109/09638288.2014.933899 arXiv:http://dx.doi.org/10.3109/09638288.2014.933899 PMID: 24969698.

[8] Christopher G Burns and Stephen H Fairclough. 2015. Use of auditory event-related potentials to measure immersion during a computer game. *International Journal of Human-Computer Studies* 73 (2015), 107–114.

[9] Colin Camerer. 2005. Three Cheers–Psychological, Theoretical, Empirical–For Loss Aversion. *Journal of Marketing Research* 42, 2 (2005), 129–133. http://www.jstor.org/stable/30164010

[10] Charles S Carver and Michael F Scheier. 1981. Self-consciousness and reactance. *Journal of Research in Personality* 15, 1 (1981), 16–29.

[11] Raquel Paz Castro, Severin Haug, Andreas Filler, Tobias Kowatsch, and Michael P Schaub. 2017. Engagement Within a Mobile Phone–Based Smoking Cessation Intervention for Adolescents and its Association With Participant Characteristics and Outcomes. *Journal of Medical Internet Research* 19, 11 (2017).

[12] Kate Cavanagh et al. 2010. Turn on, tune in and (don't) drop out: engagement, adherence, attrition, and alliance with internet-based interventions. *Oxford guide to low intensity CBT interventions* (2010), 227–233.

[13] Zhenghao Chen, Pang Wei Koh, Philip L Ritter, Kate Lorig, Erin O'Carroll Bantum, and Suchi Saria. 2015. Dissecting an online intervention for cancer survivors: four exploratory analyses of internet engagement and its effects on health status and health behaviors. *Health Education & Behavior* 42, 1 (2015), 32–45.

[14] Eun Kyoung Choe, Bongshin Lee, Haining Zhu, Nathalie Henry Riche, and Dominikus Baur. 2017. Understanding self-reflection: how people reflect on personal data through visual data exploration. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare.* ACM, 173–182.

[15] Kay Connelly, Katie A Siek, Beenish Chaudry, Josette Jones, Kim Astroth, and Janet L Welch. 2012. An offline mobile nutrition monitoring intervention for varying-literacy patients receiving hemodialysis: a pilot study examining usage and usability. *Journal of the American Medical Informatics Association* 19, 5 (2012), 705–712.

[16] Mick P Couper, Gwen L Alexander, Nanhua Zhang, Roderick JA Little, Noel Maddy, Michael A Nowak, Jennifer B McClure, Josephine J Calvi, Sharon J Rolnick, Melanie A Stopponi, et al. 2010. Engagement and retention: measuring breadth and depth of participant use of an online intervention. *Journal of medical Internet research* 12, 4 (2010).

[17] Brian G Danaher, Shawn M Boles, Laura Akers, Judith S Gordon, and Herbert H Severson. 2006. Defining participant exposure measures in Web-based health behavior change programs. *Journal of Medical Internet Research* 8, 3 (2006).

[18] Jay Devore. 2006. A Modern Introduction to Probability and Statistics: Understanding Why and How. F. M. Dekking, C. Kraaikamp, H. P. Lopuhaa, and L. E. Meester. *J. Amer. Statist. Assoc.* 101 (02 2006), 393–394. https://doi.org/10.2307/30047473

[19] Amy Wenxuan Ding, Shibo Li, and Patrali Chatterjee. 2015. Learning user real-time intent for optimal dynamic web page transformation. *Information Systems Research* 26, 2 (2015), 339–359.

[20] Daniel A. Epstein, An Ping, James Fogarty, and Sean A. Munson. 2015. A Lived Informatics Model of Personal Informatics. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15).* ACM, New York, NY, USA, 731–742.

https://doi.org/10.1145/2750858.2804250

[21] Ronghua Fang and Xia Li. 2016. Electronic messaging support service programs improve adherence to lipid-lowering therapy among outpatients with coronary artery disease: an exploratory randomised control study. *Journal of clinical nursing* 25, 5-6 (2016), 664–671.

[22] Center for Disease Control and Prevention. 2018. *How much physical activity do adults need?* https://www.cdc.gov/physicalactivity/basics/adults/ [Accessed: 15-Nov-2018].

[23] Caroline Free, Gemma Phillips, Louise Watson, Leandro Galli, Lambert Felix, Phil Edwards, Vikram Patel, and Andy Haines. 2013. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. *PLoS medicine* 10, 1 (2013), e1001363.

[24] Claire Garnett, David Crane, Robert West, Jamie Brown, and Susan Michie. 2015. Identification of behavior change techniques and engagement strategies to design a smartphone app to reduce alcohol consumption using a formal consensus method. *JMIR mHealth and uHealth* 3, 2 (2015).

[25] Russell E Glasgow, Steven M Christiansen, Deanna Kurz, Diane K King, Tim Woolley, Andrew J Faber, Paul A Estabrooks, Lisa Strycker, Deborah Toobert, and Jennifer Dickman. 2011. Engagement in a diabetes self-management website: usage patterns and generalizability of program use. *Journal of medical Internet research* 13, 1 (2011).

[26] H. C. Gooding, C. Milliren, C. M. Shay, T. K. Richmond, A. E. Field, and M. W. Gillman. 2016. Achieving Cardiovascular Health in Young Adulthood-Which Adolescent Factors Matter? *J Adolesc Health* 58, 1 (Jan 2016), 119–121.

[27] Ben Gouveia, Evangelos Karapanos, and Marc Hassenzahl. 2015. How do we engage with activity trackers?: a longitudinal study of Habito. *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (2015), 1305–1316. https://doi.org/10.1145/2750858.2804290

[28] Rebecca Gulotta, Jodi Forlizzi, Rayoung Yang, and Mark Wah Newman. 2016. Fostering Engagement with Personal Informatics Systems. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems (DIS '16)*. ACM, New York, NY, USA, 286–300. https://doi.org/10.1145/2901790.2901803

[29] Scott D. Halpern, Benjamin French, Dylan S. Small, Kathryn Saulsgiver, Michael O. Harhay, Janet Audrain-McGovern, George Loewenstein, Troyen A. Brennan, David A. Asch, and Kevin G. Volpp. 2015. Randomized Trial of Four Financial-Incentive Programs for Smoking Cessation. *New England Journal of Medicine* 372, 22 (2015), 2108–2117. https://doi.org/10.1056/NEJMoa1414293 arXiv:https://doi.org/10.1056/NEJMoa1414293 PMID: 25970009.

[30] László Harmat, Örjan de Manzano, Töres Theorell, Lennart Högman, Håkan Fischer, and Fredrik Ullén. 2015. Physiological correlates of the flow experience during computer game playing. *International Journal of Psychophysiology* 97, 1 (2015), 1–7.

[31] Onaedo Ilozumba, Ibukun-Oluwa Omolade Abejirinde, Marjolein Dieleman, Azucena Bardají, Jacqueline EW Broerse, and Sara Van Belle. 2018. Targeting strategies of mHealth interventions for maternal health in low and middle-income countries: a systematic review protocol. *BMJ open* 8, 2 (2018), e019345.

[32] Charlene Jennett, Anna L Cox, Paul Cairns, Samira Dhoparee, Andrew Epps, Tim Tijs, and Alison Walton. 2008. Measuring and defining the experience of immersion in games. *International journal of human-computer studies* 66, 9 (2008), 641–661.

[33] Bin Jiang. 2013. Head/Tail Breaks: A New Classification Scheme for Data with a Heavy-Tailed Distribution. *The Professional Geographer* 65, 3 (2013), 482–494. https://doi.org/10.1080/00330124.2012.700499

[34] Leslie K. John, George Loewenstein, Andrea B. Troxel, Laurie Norton, Jennifer E. Fassbender, and Kevin G. Volpp. 2011. Financial Incentives for Extended Weight Loss: A Randomized, Controlled Trial. *Journal of General Internal Medicine* 26, 6 (01 Jun 2011), 621–626. https://doi.org/10.1007/s11606-010-1628-y

[35] Katherine R Jones, Nirmala Lekhak, and Napatsawan Kaewluang. 2014. Using Mobile Phones and Short Message Service to Deliver Self-Management Interventions for Chronic Conditions: A Meta-Review. *Worldviews on Evidence-Based Nursing* 11, 2 (2014), 81–88.

[36] Sunanda V Kane and Andrew Robinson. 2010. understanding adherence to medication in ulcerative colitis–innovative thinking and evolving concepts. *Alimentary pharmacology & therapeutics* 32, 9 (2010), 1051–1058.

[37] Sahar Khonsari, Pathmawathi Subramanian, Karuthan Chinna, Lydia A Latif, Lee W Ling, and Omid Gholami. 2015. Effect of a reminder system using an automated short message service on medication adherence following acute coronary syndrome. *European Journal of Cardiovascular Nursing* 14, 2 (2015), 170–179.

[38] Alain B. Labrique, Lavanya Vasudevan, Erica Kochi, Robert Fabricant, and Garrett Mehl. 2013. mHealth innovations as health system strengthening tools: 12 common applications and a visual framework. *Global Health, Science and Practice* 1, 2 (2013), 160–171. https://doi.org/10.9745/GHSP-D-13-00031

[39] Mark Lemstra, Yelena Bird, Chijioke Nwankwo, Marla Rogers, and John Moraros. 2016. Weight loss intervention adherence and factors promoting adherence: a meta-analysis. *Patient preference and adherence* 10 (2016), 1547.

[40] Ian Li, Anind K. Dey, and Jodi Forlizzi. 2011. Understanding My Data, Myself: Supporting Self-reflection with Ubicomp Technologies. In *Proceedings of the 13th International Conference on Ubiquitous Computing (UbiComp '11)*. ACM, New York, NY, USA, 405–414. https://doi.org/10.1145/2030112.2030166

[41] S. Lloyd. 1982. Least squares quantization in PCM. *IEEE Transactions on Information Theory* 28, 2 (March 1982), 129–137. https://doi.org/10.1109/TIT.1982.1056489

[42] Jamie L Manwaring, Susan W Bryson, Andrea B Goldschmidt, Andrew J Winzelberg, Kristine H Luce, Darby Cunning, Denise E Wilfley, and C Barr Taylor. 2008. Do adherence variables predict outcome in an online program for the prevention of eating disorders? *Journal of Consulting and Clinical Psychology* 76, 2 (2008), 341.

[43] Corby K Martin, L Anne Gilmore, John W Apolzan, Candice A Myers, Diana M Thomas, and Leanne M Redman. 2016. Smartloss: a personalized mobile health intervention for weight management and health promotion. *JMIR mHealth and uHealth* 4, 1 (2016).

[44] Jennifer B McClure, Susan M Shortreed, Andy Bogart, Holly Derry, Karin Riggs, Jackie St John, Vijay Nair, and Larry An. 2013. The effect of program design on engagement with an internet-based smoking intervention: randomized factorial trial. *Journal of medical Internet research* 15, 3 (2013).

[45] Susan Michie, Charles Abraham, Craig Whittington, John McAteer, and Sunjai Gupta. 2009. Effective techniques in healthy eating and physical activity interventions: a meta-regression. *Health Psychology* 28, 6 (2009), 690.

[46] S Miller, B Ainsworth, L Yardley, A Milton, M Weal, P Smith, and L Morrison. 2018. Supporting Systematic Assessment of Digital Interventions: A Framework for Analysing and Measuring Usage and Engagement Data (AMUsED). *Journal of Medical Internet Research* (2018). https://doi.org/10.2196/preprints.10966

[47] Leanne Morrison, Rona Moss-Morris, Susan Michie, and Lucy Yardley. 2014. Optimizing engagement with internet-based health behaviour change interventions: Comparison of self-assessment with and without tailored feedback using a mixed methods approach. *British journal of health psychology* 19, 4 (2014), 839–855.

[48] Adrià Muntaner, Josep Vidal-Conti, and Pere Palou. 2016. Increasing physical activity through mobile device interventions: A systematic review. *Health informatics journal* 22, 3 (2016), 451–469.

[49] Fionn Murtagh and Pedro Contreras. 2012. Algorithms for hierarchical clustering: an overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 2, 1 (2012), 86–97.

[50] Venkatesh L. Murthy, Siddique A. Abbasi, Juned Siddique, Laura A. Colangelo, Jared Reis, Bharath A. Venkatesh, J. Jeffrey Carr, James G. Terry, Sarah M. Camhi, Michael Jerosch-Herold, Sarah de Ferranti, Saumya Das, Jane Freedman, Mercedes R. Carnethon, Cora E. Lewis, Joao A. C. Lima, and Ravi V. Shah. 2016. Transitions in Metabolic Risk and Long-Term Cardiovascular Health: Coronary Artery Risk Development in Young Adults (CARDIA) Study. *Journal of the American Heart Association* 5, 10 (2016). https://doi.org/10.1161/JAHA.116.003934 arXiv:http://jaha.ahajournals.org/content/5/10/e003934.full.pdf

[51] Heather O'Brien and Paul Cairns. 2016. *Why Engagement Matters: Cross-Disciplinary Perspectives of User Engagement in Digital Media.* Springer.

[52] Heather L O'Brien and Elaine G Toms. 2008. What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the Association for Information Science and Technology* 59, 6 (2008), 938–955.

[53] Mary Paone, Sandy R Whitehouse, and Dewey Evans. 2015. JUST TRAC it!: Assessing a Mobile-Health (mhealth) Intervention for Youth With Chronic Health Conditions and/or Disabilities (CHC/Ds) to Encourage Engagement and Readiness Behaviors for Transition. *Journal of Adolescent Health* 56, 2 (2015), S27–S28.

[54] Linda G Park, Jill Howie-Esquivel, Misook L Chung, and Kathleen Dracup. 2014. A text messaging intervention to promote medication adherence for patients with coronary heart disease: a randomized controlled trial. *Patient education and counseling* 94, 2 (2014), 261–268.

[55] Olga Perski, Ann Blandford, Robert West, and Susan Michie. 2016. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. *Translational Behavioral Medicine* (2016). https://doi.org/10.1007/s13142-016-0453-1

[56] Angela Fidler Pfammatter, Alexa Mitsos, Shirlene Wang, Susan Hammett Hood, and Bonnie Spring. 2017. Evaluating and improving recruitment and retention in an mHealth clinical trial: an example of iterating methods during a trial. *mHealth* 3, 11 (2017). http://mhealth.amegroups.com/article/view/17200

[57] Jacques Quilici, Lionel Fugon, Shirley Beguin, Pierre Emmanuel Morange, Jean-Louis Bonnet, Marie-Christine Alessi, Patrizia Carrieri, and Thomas Cuisset. 2013. Effect of motivational mobile phone short message service on aspirin adherence after coronary stenting for acute coronary syndrome. *International journal of cardiology* 168, 1 (2013), 568–569.

[58] Quazi Abidur Rahman, Tahir Janmohamed, Meysam Pirbaglou, Paul Ritvo, Jane M Heffernan, Hance Clarke, and Joel Katz. 2017. Patterns of User Engagement With the Mobile App, Manage My Pain: Results of a Data Mining Investigation. *JMIR Mhealth Uhealth* 5, 7 (2017), e96. https://doi.org/10.2196/mhealth.7871

[59] William T Riley, Daniel E Rivera, Audie A Atienza, Wendy Nilsen, Susannah M Allison, and Robin Mermelstein. 2011. Health behavior models in the age of mobile interventions: are our theories up to the task? *Translational behavioral medicine* 1, 1 (2011), 53–71.

[60] Lee M Ritterband, Frances P Thorndike, Daniel J Cox, Boris P Kovatchev, and Linda A Gonder-Frederick. 2009. A behavior change model for internet interventions. *Annals of Behavioral Medicine* 38, 1 (2009), 18.

[61] Peter J Rousseeuw. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics* 20 (1987), 53–65.

[62] Ryan J Shaw, Dori M Steinberg, Leah L Zullig, Hayden B Bosworth, Constance M Johnson, and Linda L Davis. 2014. mHealth interventions for weight loss: a guide for achieving treatment fidelity. *Journal of the American Medical Informatics Association* 21, 6 (2014), 959–963.

[63] Camille E Short, Amanda L Rebar, Ronald C Plotnikoff, and Corneel Vandelanotte. 2015. Designing engaging online behaviour change interventions: A proposed model of user engagement. *European Health Psychologist* 17, 1 (2015), 32–38.

[64] Camille E Short, Corneel Vandelanotte, Marcus W Dixon, Richard Rosenkranz, Cristina Caperchione, Cindy Hooker, Mohan Karunanithi, Gregory S Kolt, Anthony Maeder, Hang Ding, et al. 2014. Examining participant engagement in an information technology-based physical activity and nutrition intervention for men: the manup randomized controlled trial. *JMIR research protocols* 3, 1 (2014).

[65] Karandeep Singh, Kaitlin Drouin, Lisa P Newmark, Ronen Rozenblum, Jaeho Lee, Adam Landman, Erika Pabo, Elissa V Klinger, and David W Bates. 2016. Developing a framework for evaluating the patient engagement, quality, and safety of mobile health applications. *Issue Brief (Commonw Fund)* 5, 1 (2016), 11.

[66] Alan Sosa, Nathan Heineman, Kimberly Thomas, Kai Tang, Marie Feinstein, Michelle Y Martin, Baran Sumer, and David L Schwartz. 2017. Improving patient health engagement with mobile texting: A pilot study in the head and neck postoperative setting. *Head & neck* 39, 5 (2017), 988–995.

[67] B. Spring, A. C. Moller, L. A. Colangelo, J. Siddique, M. Roehrig, M. L. Daviglus, J. F. Polak, J. P. Reis, S. Sidney, and K. Liu. 2014. Healthy lifestyle change and subclinical atherosclerosis in young adults: Coronary Artery Risk Development in Young Adults (CARDIA) study. *Circulation* 130, 1 (Jul 2014), 10–17.

[68] Stoyan R Stoyanov, Leanne Hides, David J Kavanagh, Oksana Zelenko, Dian Tjondronegoro, and Madhavan Mani. 2015. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. *JMIR mHealth and uHealth* 3, 1 (2015).

[69] S. Taki, S. Lymer, C. G. Russell, K. Campbell, R. Laws, K. L. Ong, R. Elliott, and E. Denney-Wilson. 2017. Assessing User Engagement of an mHealth Intervention: Development and Implementation of the Growing Healthy App Engagement Index. *JMIR Mhealth Uhealth* 5, 6 (Jun 2017), e89.

[70] Sarah Taki, Sharyn Lymer, Catherine Georgina Russell, Karen Campbell, Rachel Laws, Kok-Leong Ong, Rosalind Elliott, and Elizabeth Denney-Wilson. 2017. Assessing User Engagement of an mHealth Intervention: Development and Implementation of the Growing Healthy App Engagement Index. *JMIR Mhealth Uhealth* 5, 6 (2017), e89. https://doi.org/10.2196/mhealth.7236

[71] Amos Tversky and Daniel Kahneman. 1991. Loss Aversion in Riskless Choice: A Reference-Dependent Model*. *The Quarterly Journal of Economics* 106, 4 (1991), 1039–1061. https://doi.org/10.2307/2937956

[72] William M Vollmer, Ashil A Owen-Smith, Jeffrey O Tom, Reesa Laws, Diane G Ditmer, David H Smith, Amy C Waterbury, Jennifer L Schneider, Cyndee H Yonehara, Andrew Williams, et al. 2014. Improving adherence to cardiovascular disease medications with information technology. *The American journal of managed care* 20, 11 Spec No. 17 (2014), SP502–10.

[73] S Wagner, Thomas Skjødeberg Toftegaard, Olav W Bertelsen, et al. 2013. Introducing the Adherence Strategy Engineering Framework (ASEF). *Methods of information in medicine* 52, 3 (2013), 220–230.

[74] David S Wald, Jonathan P Bestwick, Lewis Raiman, Rebecca Brendell, and Nicholas J Wald. 2014. Randomised trial of text messaging on adherence to cardiovascular preventive treatment (INTERACT trial). *PloS one* 9, 12 (2014), e114268.

[75] Joshua H West, Lindsay M Belvedere, Rebecca Andreasen, Christine Frandsen, P Cougar Hall, and Benjamin T Crookston. 2017. Controlling your "App" etite: how diet and nutrition-related mobile apps lead to behavior change. *JMIR mHealth and uHealth* 5, 7 (2017).

[76] Mark Whooley, Bernd Ploderer, and Kathleen Gray. 2014. On the Integration of Self-tracking Data Amongst Quantified Self Members. In *Proceedings of the 28th International BCS Human Computer Interaction Conference on HCI 2014 - Sand, Sea and Sky - Holiday HCI (BCS-HCI '14)*. BCS, UK, 151–160. https://doi.org/10.14236/ewic/hci2014.16

[77] Lucy Yardley, Bonnie J. Spring, Heleen Riper, Leanne G. Morrison, David H. Crane, Kristina Curtis, Gina C. Merchant, Felix Naughton, and Ann Blandford. 2016. Understanding and Promoting Effective Engagement With Digital Behavior Change Interventions. *American Journal of Preventive Medicine* 51, 5 (2016), 833 – 842. https://doi.org/10.1016/j.amepre.2016.06.015