A Multi-Faceted Approach to Characterizing User Behavior and Experience in a Digital Mental Health Intervention

Annie T. Chen[1], **Shuyang Wu**[1,2], **Kathryn N. Tomasino**[3,4], **Emily G. Lattie**[3], **David C. Mohr**[3]

[1]Biomedical Informatics and Medical Education, UW Medicine South Lake Union, 850 Republican Street, Box 358047, Seattle, WA 98109

[2]Veracyte, Inc., 6000 Shoreline Ct., Suite 300, South San Francisco, CA 94080

[3]Center for Behavioral Intervention Technologies, Northwestern University, 750 N Lake Shore Dr, 10th Fl, Chicago, IL 60611

[4]Department of Medicine, Northwestern University Feinberg School of Medicine, 420 E Superior St, Chicago, IL 60611

Abstract

Digital interventions offer great promise for supporting health-related behavior change. However, there is much that we have yet to learn about how people respond to them. In this study, we present a novel mixed-methods approach to analysis of the complex and rich data that digital interventions collect. We perform secondary analysis of IntelliCare, an intervention in which participants are able to try 14 different mental health apps over the course of eight weeks. The goal of our analysis is to characterize users' app use behavior and experiences, and is rooted in theoretical conceptualizations of engagement as both usage and user experience. In the first aim, we employ cluster analysis to identify subgroups of participants that share similarities in terms of the frequency of their usage of particular apps, and then employ other engagement measures to compare the clusters. We identified four clusters with different app usage patterns: Low Usage, High Usage, Daily Feats Users, and Day to Day users. Each cluster was distinguished by its overall frequency of app use, or the main app that participants used. In the second aim, we developed a computer-assisted text analysis and visualization method – message highlighting – to facilitate comparison of the clusters. Last, we performed a qualitative analysis using participant messages to better understand the mechanisms of change and usability of salient apps from the cluster analysis. Our novel approach, integrating text and visual analytics with more traditional qualitative analysis techniques, can be used to generate insights concerning the behavior and experience of users in digital health contexts, for subsequent personalization and to identify areas for improvement of intervention technologies.

Keywords

Corresponding author: Annie T. Chen, Biomedical Informatics and Medical Education, UW Medicine South Lake Union, 850 Republican Street, Box 358047, Seattle, WA 98109 atchen@uw.edu.

1 Introduction

In recent years, we have seen a proliferation of health-related applications, delivered through mobile and desktop devices, to support the management of chronic health conditions[1]. The increased popularity and functionality of mobile devices offers great promise for enhancing the delivery of mental health services including: increasing access to and use of evidence-based care, informing and engaging users more actively in treatment, and enhancing care after treatment has concluded[2]. However, many challenges also exist for effective deployment, including high attrition rates, the digital divide, and intellectual capabilities of users[3]. More broadly, there are also issues and gaps in our knowledge of digital mental health interventions, such as the need to identify predictors of therapeutic success or failure[4], reasons for attrition and dropout[5], and determining active intervention components[6]. Though extant literature has shown that digital interventions work, we are not yet able to explain how, why, and for whom[7–9].

As such, there has been great interest in engagement with digital interventions and many definitions of engagement have emerged. There are differences in conceptualization of engagement across disciplines, with research from the behavioral science literature tending to focus on "engagement as usage", and from the computer science and human-computer interaction (HCI) literatures, on "engagement as flow"[10]. The term flow arises from the work of Csikszentmihalyi, who first employed it to refer to a state of positive affect and focused attention, in which people are optimally engaged in an experience[11]. HCI literature has defined engagement as a state arising out of system use, a quality of user experience that is associated with challenge, positive affect, endurability, aesthetic and sensory appeal, attention, feedback, variety/novelty, interactivity, and perceived user control[12]. In this study, we incorporate both perspectives, considering engagement as both usage and user experience.

A number of conceptual frameworks to understand engagement and associated factors, such as mechanisms of change and usability, have been proposed (e.g. [1,10,13]). These models often elucidate factors that influence engagement, such as user characteristics (e.g., demographics; Internet self-efficacy, current and past health behavior), environmental factors that facilitate or impede intervention use (e.g. personal, professional, healthcare system), and aspects of interventions such as design, usability, content, and delivery. However, there are also aspects which the models do not have in common.

Current methods of analyzing engagement and factors that influence engagement tend to rely on simple metrics, such as raw counts of usage, which do not capture the complexity and richness of these processes. Engagement has typically been characterized using a variety of types of usage data, including logins, accesses, sessions, duration, messages sent, posts and comments made, and other patterns of use over time[14–18]. Usage data is also often conceptualized as a measure of adherence[19,20] or exposure to intervention content[14,21]. Other studies have examined usage trends to better characterize users in terms of needs and user preferences[15,22]. Still other research has employed log data to derive insights into users' support needs, develop recommendations for incorporating new content, and identify points of attrition[23,24]. Visualization of user interactions can also temporally illustrate

how participants journey through an application and facilitate comparison of users[25]. However, these types of analyses of log data still do not enable us to understand how participants *experience* an intervention.

Interventions involving multiple and diverse potentially impactful components pose both a particular challenge and opportunity to understand user engagement and factors that influence engagement. The mode of delivery of digital interventions means that there is usually a rich set of log data including the user's actions, time of day taken, and the intervention components used. Moreover, there are usually multiple intervention components, which make analyses increasingly complex, but could also enable us to understand user experiences more holistically, as well as lead to insights concerning the aspects of interventions that lead to behavioral change.

In this study, we examine engagement, defined as usage and user experience, and factors that may influence engagement, such as mechanisms of change and usability, through log data collected in the context of treatment delivery. To achieve this goal, we take a novel approach, integrating text and visual analytics, to facilitate analysis of a rich and complex dataset. This secondary analysis was based on data collected from individuals who used IntelliCare, an eclectic, skills-based app suite comprised of 14 apps involving different treatment strategies for depression and anxiety[26,27]. Because participants were able to start and stop using apps whenever they wanted, they could be using any number of apps at any given time, and the challenge of characterizing user interactions with different apps is particularly difficult. Previous work has examined the relationship between app use and clinical outcomes[28]; however, that worked relied primarily on fairly standard use metrics such as number of app use sessions and did not explore richer methods of understanding user engagement.

In this study, we use a three-pronged approach to combine the rich set of quantitative usage data along with qualitative data, participants' SMS conversations with coaches, to better understand users' app usage patterns and the nature of their user experiences. First, we characterize user engagement as usage, through cluster analysis. Second, we characterize participants' engagement holistically, considering engagement both as usage and as user experience, using computer-assisted textual analysis and visualization of participants' text messages. Third, we focus on two important constructs in the extant literature on engagement of digital interventions, mechanisms of change and usability, to derive insights to inform the future re-design of the apps in the IntelliCare suite.

2   Methods

2.1   Intervention and Dataset

The IntelliCare suite of apps was comprised of 13 clinical apps designed to improve symptoms of depression and anxiety, and a "Hub" app, which coordinated participants' experiences with the other apps by consolidating app notifications and providing recommendations of apps to try[27]. The tools were designed to be intuitive and require few instructions. This suite of apps is available on Google Play, and users were able to select the apps that they wanted to use. Those who enrolled in the trial also received eight weeks of coaching on the use of IntelliCare and automated recommendations of apps to try. During

the trial, participants received weekly recommendations to download and try two new apps each week. Recommendation sequences were generated randomly.

The coaching protocol[29] was based on the Efficiency Model of Support [30] and supportive accountability[31]. Coaches encouraged participants to try the apps recommended to them through the Hub app, answered questions about the tools found in the apps, encouraged application of skills in daily life, and provided technical support. They positively reinforced participants for app use, inquired about problems and preferences when usage was low, and engaged in problem-solving around low usage as needed. Coaching began with an initial 30- to 45-minute phone call to establish goals for mood and anxiety management, ensure that participants could download the Hub app from the Google Play store, introduce the suite of apps, build rapport, and set expectations for the coach-participant relationship. During the trial, participants received 2–3 texts per week from their coach, and were able to reach out to their coach via text message whenever they had questions about the use of the apps or application of the learned skills. Coaches would respond to questions within one business day.

2.2 Aim 1 Characterizing User Engagement as Usage

In this aim, we first consider engagement as usage and cluster participants based on their usage of the apps in the IntelliCare suite, to identify individuals that shared similarities in their engagement. To enrich our characterization of participants' engagement behaviors, we then employed cross-cluster comparisons on both overall and app-specific measures of engagement, and outcomes. Cluster analysis is a method of finding similar groups in data which has been employed in many disciplines, for diverse tasks ranging from segmentation of consumers by food choice[32,33], analyzing learner subpopulations in online educational environments[34,35], developing personas to inform the design of eHealth technologies [36,37], and identifying patterns of user engagement with a mobile app intervention[38].

With this dataset, one of the key challenges we faced concerned the need to understand both a users' *overall engagement*, as well as *app-specific engagement*, interaction tendencies that were specific to particular apps, and the numerous ways engagement could thus be considered. We employed a two-step user characterization process, in which we first clustered participants based on their app usage patterns, and then compared the clusters based on other overall and app-specific measures of engagement, thereby facilitating a richer characterization of participant engagement.

We employed k-means, a common clustering method in which the squared error between the empirical mean of a cluster and the points in the cluster is minimized[39]. We clustered participants based on the number of times they accessed each of the 14 apps. By doing so, emergent clusters would be comprised of individuals who exhibited similar usage frequencies for the same apps, where frequency was defined by the number of sessions of usage of each app.

One of the issues that one encounters using k-means cluster analysis is the selection of the appropriate number of clusters. We employed two methods, the Elbow method and visual comparison, together. With the Elbow method, the variance is plotted against the number of

clusters and the point at which increasing the number of clusters does not lead to substantial decreases in variance is selected[40,41]. Because there was some ambiguity, we visually compared solutions to make the final determination of the solution with the most meaningful separation in terms of app usage.

The k-means clustering method can be susceptible to the values that were used to initialize the clustering[42]. To avoid bias, we repeated the clustering with different initialization values and observed that the defining characteristics of the clustering solutions remained the same in the repetitions.

After identifying the individuals with similar app usage patterns in terms of frequency, we compared the clusters based on multiple measures of *overall engagement*: the total number of app usage sessions across all apps (where an app use session was defined as a single access of an app), the total messages sent, and proportion of recommendations followed. The number of app usage sessions and messages sent are common measures of engagement in digital interventions. An app recommendation was considered followed if an individual accessed an app within seven days of receiving a recommendation to try it. We might consider the proportion of recommendations followed to gauge participants' receptivity to suggestions, a potential indicator of engagement. To compare the clusters based on the first two measures, we employed the Kruskal-Wallis one-way analysis-of-variance-by-ranks test (H test), which does not assume that the variables are normally distributed[43]. We compared the clusters in terms of the extent to which they followed app recommendations using a Chi-square test.

We also compared the clusters based on measures of *app-specific engagement*: app start day, or the number of days since enrollment that participants began using each app, and app usage duration, where app use duration was defined as the number of days between the last and the first day that an app was used, using T-tests. The rationale for considering app start day was that it might indicate a user's interest in a given app. For example, an earlier start date might be indicative of a greater need for the service provided by the app or a greater interest in the concept of the app. App usage duration might be associated with personal relevance of the app, e.g. immediate abandonment might suggest that the app was not perceived to be useful at all, whereas sustained use might suggest that the user found the app useful. In both cases, we recognize that at this stage, these are suppositions, and thus the need to consider this an exploratory analysis.

Lastly, we rendered line charts depicting the outcomes of the participants of the study measured using the GAD-7[44] and PHQ-9[45]. The measures employed in the cluster analysis and subsequent comparisons are presented in Table 2. K-means clustering was performed using the machine learning toolkit, scikit-learn (http://scikit-learn.org/stable/index.html), and all visualizations were rendered using the D3 library (https://d3js.org/).

2.3   Aim 2 Message Highlighting: Computer-Assisted Discovery of Salient User Experience Characteristics

We developed a computer-assisted text analysis and visualization method to extract elements of participants' text messages to coaches that characterize engagement in terms of usage and

user experience, as well as illustrate factors affecting engagement. This process involved developing a lexicon-based visualization to identify salient concepts from the text messages and then mapping these concepts to constructs in the engagement literature.

We used a keyword-matching method to identify concepts relating to behavior change and engagement with digital interventions and used these features to render an interactive visualization enabling viewers to compare clusters in terms of the extent to which users' intervention experiences were reflected in the messages. To facilitate a process of "discovery" of salient concepts from the text data, we employed a sentiment analysis lexicon comprised of over 6,000 words[46]. We calculated the extent to which each concept was represented in the participants' communications, at the participant level. In other words, we calculated the proportion of participants whose communications included each word at least once.

Given the large number of concepts, there was a need to reduce dimensionality. We did so using several criteria. First, we eliminated all concepts that were not mentioned by at least half of the sample. Then, we eliminated shorter words that were likely to be ambiguous or lack substance. This included all three-letter words. We manually curated 4- and 5- letter words to remove those that lacked substance, were embedded in another word, or were part of one of the app names. For example, "rough" and "sever" were eliminated as being part of "through" and "several", respectively.

The concepts were further grouped into high-level categories based on constructs that are used to explain engagement and behavior change in digital interventions. As explained in the Introduction, owing to the different disciplines in which engagement has been studied, different facets have been studied. To be able to more richly represent concepts relating to engagement that are expressed in the text data, we incorporated constructs from multiple models and other literature. In particular, we consider Short et al.'s model[1] and Ritterband's model of Internet Interventions[13]. Short et al.'s model draws upon research on engagement from multiple disciplines and includes four primary constructs: environment, individual, intervention (design, usability, personal relevance), and engagement[1]. The environment is comprised of external factors that impede or facilitate intervention use, such as the length of time available to the user. The design of the intervention and individual characteristics affect perceived usability, and individual characteristics affect personal relevance. The design, usability, and personal relevance in turn affect engagement. Ritterband's model, which is highly cited in behavioral science literature, argues that website use leads to behavior change and symptom improvement, and that this process can be mediated by mechanisms of change[13]. We define and describe mechanisms of change in more detail in the next aim. As none of the models account for all aspects of engagement and associated factors, we based our assignment of categories on multiple conceptualizations and report these in in the Results section. The assignment of the concept to high-level categories was performed by the first author using the interactive visualization, by studying the messages retrieved by the keyword and assigning the suitable category. This process of computer-assisted discovery is depicted in Figure 1.

2.4 Aim 3 Qualitative Analysis of User Feedback

In this last aim, we performed qualitative analysis of participants' messages to better understand their use experiences with the most salient apps identified in Aim 1. Though up to this point, we focused on considering differences in subgroups of users, in this last aim, we focused on characterizing users' needs specific to the apps that participants used most, earliest, or longest according to the analysis of usage data in Aim 1. We employed manual coding of the data to better understand the benefits and issues that participants experienced with each app. As the basis for the qualitative analysis, we first filtered the dataset to include only the messages that explicitly mentioned one of the apps. During the filtering process, we also included variant spellings of app names, such as "melocate" for "ME Locate" and "Move Me" for "MoveMe"; and "sleep diary" as a reference to Slumber Time. Then we created separate documents for each of the 14 apps, with one document containing all of the messages relating to a particular app, and imported these into Atlas.ti, a qualitative data analysis software, for analysis.

We employed a general inductive qualitative coding procedure based on the Grounded Theory Methodology[47], in which the researcher engages in line-by-line coding of the data and develops conceptual codes[48,49]. This involves an initial reading of the text data, identifying text segments relating to the topics of interest, labeling the segments with "codes", iteratively comparing the codes to reduce redundancy and overlap, and creating a conceptual model comprised of the categories of greatest interest[49]. To ensure rigor in the codes and their conceptual linkages, we considered the criteria for empirical grounding of findings outlined by Corbin and Strauss[50]. These criteria include concerns such as: whether the concepts are systematically related, well developed, and have conceptual density. We coded the entire dataset, as doing so enabled us to increase conceptual density and develop conceptual categories that are relevant not only for the findings that we report, but for a greater breadth of apps. We report the findings that are most relevant to apps that were salient in Aim 1. We report the themes identified that pertain to two primary topics relating to engagement with digital interventions: mechanisms of change, and usability perceptions. Mechanisms of change, which mediate the effect of engagement with the intervention and intervention effectiveness, include knowledge/information, motivation, beliefs and attitudes, and skill building, account for behavior change and subsequent symptom improvement[10,13]. The qualitative analysis was performed by the first author.

3 Results

3.1 Dataset

Ninety-nine participants enrolled in the study. One participant withdrew after the study was over, and thus we report data based on 98 participants over the course of the eight week intervention (Table 3). Overall, participants and coaches together sent 10,227 messages. Participants alone sent 3,427 messages. The most commonly used apps were Day to Day, Daily Feats, and Slumber Time (Table 4).

3.2   Aim 1 Cluster Analysis of App Usage

We performed k-means cluster analysis and identified a 4-cluster solution as described in the Methods section. Each cluster had one or more defining characteristics based on their overall app usage, as well as usage of specific apps. We depict the mean usage for each app in Figure 2. Given that the Hub app served a coordinating function within the suite, its usage was understandably high in all groups. Hereafter, we will refer to these clusters as Low Usage, High Usage, Daily Feats Users, and Day to Day Users, based on their most distinguishing characteristic. We now briefly describe each cluster.

The Low Usage group was the largest of the groups (n=45), and exhibited low usage of all apps, relative to other users. The High Usage group was the smallest of the groups (n=6), and engaged in comparatively high usage. These two groups were characterized primarily by their overall amount of usage, rather than usage of particular apps. In contrast, Daily Feats Users (n=26) and Day to Day Users (n=21) were characterized by their usage of these two apps in particular.

We compared the clusters in terms of several measures of overall engagement: the number of app sessions, messages sent, and recommendations followed (Table 5). Not surprisingly, the clusters exhibited overall app usage patterns consistent with their cluster characteristics, with the High Usage cluster engaging in the most app usage sessions, the Low Usage cluster least, and the other two clusters exhibiting app usage session totals somewhere in between. The number of participant messages sent did not differ significantly between all of the clusters, suggesting that participants responded to coaches consistently regardless of the level of engagement with the apps. Participants received two randomly selected app recommendations per week. All clusters followed recommendations to some extent, and the clusters did not differ significantly in the extent to which they followed recommendations.

We examined whether there were significant differences in two app-specific engagement measures: average start day or duration of usage in the clusters, as the start dates might be considered to reflect some aspect of cluster members' tendencies to find particular apps personally relevant. We rendered visualizations for each cluster, with the average start day on the x-axis and app name on the y-axis (Figure 3). We depict the average start day of each cluster using a circular indicator. A colored bar is sustained to the average end day. Asterisks on the left indicate statistical significance in terms of the start day, and asterisks on the right, statistical significance in terms of the duration.

In considering these figures, we see that the Low Usage group exhibited earlier start days for Aspire and Slumber Time as compared to the rest of the sample, and shorter app usage durations for a number of the apps. The High Usage did not exhibit any significant associations with regard to start days, but they did sustain use of a number of apps longer, as compared to the rest of the sample.

Aside from starting usage of Daily Feats earlier than the rest of the sample, Daily Feats Users also began to use Day to Day, Slumber Time and Social Force at a later time. Day to Day Users' most prominent distinguishing characteristic was, not surprisingly, an earlier start and more sustained use of Day to Day.

We have characterized the clusters in terms of multiple quantitative measures of app-specific and overall measures of engagement. To conclude, we briefly compare the clusters in terms of mental health outcomes (Figure 3). There were no significant differences among the clusters in terms of either the GAD-7 or the PHQ-9, though the Daily Feats and High Usage clusters exhibit a greater trend towards continued progress in symptom improvement, as compared to the other clusters, which exhibit a tendency towards reaching an improvement plateau.

3.3 Aim 2 Message Highlighting: Computer-Assisted Discovery of Salient User Experience Characteristics

We identified salient keywords in the text by calculating the proportion of individuals using keywords from a lexicon and developed a visualization called, "Message Highlight Map" to examine the expression of these concepts. In the Message Highlight Map, the horizontal axis represents the clusters, and the vertical axis, concepts (Figure 4). The color of each cell is based on the proportion of individuals within a given cluster mentioning the focal concept at least once in their messages, such that the darker the color, the more prominent the concept was within the sample, and the lighter, the less prominent.

Each keyword was mapped to a construct that related to engagement and behavior change in digital interventions: engagement, environment, improvement, mechanisms, and usability (Table 6). We consider engagement in terms of two sub-categories, to account for differences in the ways that engagement has been conceptualized. First, the Engagement (experience) category represents the users' experience of parts of the intervention, including affective reactions. This construct is rooted in a conceptualization of engagement as including a user's thoughts and feelings during system use[12,51]. The Engagement (usage) category was used to represent the conceptualization of engagement as participants' efforts to complete the intervention as prescribed, which is often called adherence in the behavioral science literature on digital interventions (e.g. [52,53]). The Environment represents external factors that may affect an individual's engagement with an intervention, such as being busy or their interaction with others. The Mechanisms category included intervention content, skills, and mechanisms of change[10,13]. Usability is often conceptualized in terms of the effectiveness, efficiency, and satisfaction with which products can be used by users, to achieve particular goals within a specific context of use[54,55]. In this study, we adopted a broader conceptualization of usability which included characteristics such as helpfulness and usefulness, which have been seen as antecedents of satisfaction[56], and have been included in arguments and taxonomies for usability[57,58].

We now describe the main features that were commonly expressed regardless of cluster (Table 7). Participants expressed a substantial amount of apologetic sentiment for not being available or not responding quickly to the coaches. They also shared many of the reasons why they were busy. They made statements about what they needed to do (e.g. spend more time working with the apps). In the Improvement category, participants shared ways in which they felt that their symptom management had improved, e.g. their thought patterns.

Participants also used language that provided insight into their interactions with the mechanisms of the intervention. For example, all of the clusters mentioned using the apps to

deal with stress. They also asked coaches for suggestions and recommendations. Last, participants expressed positive sentiment towards the apps, finding them helpful.

We now consider differences in the clusters. As the consistently lighter shades in the Low Usage column demonstrate, we have less information about the user experience of this cluster. Other than the few themes that were prominent across all the clusters, the Low Usage cluster as a whole did not express sentiments that were not prominent in other clusters. However, they did appear to engage with coaches in terms of asking for their suggestions and following app recommendations. The High Usage cluster, on the other hand, tended to communicate both problems that they were having with the apps and in life, as well as positive impressions. Their communication thus paralleled their high engagement through app usage and recommendations.

The two remaining clusters, Daily Feats Users and Day to Day Users, were in between Low Usage and High Usage in terms of the extent to which the message content reflected aspects of user experience. Daily Feats Users made more comments about the skills were learning (e.g. thinking differently, and how to deal with stress and anxiety) and the differences that they observed. Day to Day participants were more likely to comment that they were making progress, and that the apps were helping them to think positively. Like the High Usage cluster, they tended to be appreciative of the support and encouragement that coaches provided.

The High Usage, Daily Feats and Day to Day Users all mentioned usability issues, such as the apps not opening when they were trying to open them, or inadvertently opening when they were not expecting it. Participants also talked about what they 'wished' the apps would do. There were a few participants that became overwhelmed dealing with the large number of apps or messages.

3.4 Aim 3 Qualitative Analysis of User Feedback

To enrich our understanding of participants' experiences with specific intervention components and how these particular apps may have affected participants' engagement with treatment, we performed qualitative analysis of text messages mentioning each of the applications in the IntelliCare suite. Altogether, we analyzed 777 mentions of the 14 apps. A table showing the number of mentions of each app is available in Appendix 1. We consider the results of four applications, Daily Feats, Day to Day, Slumber Time, and MoveMe. The first two apps, Daily Feats and Day to Day, were selected because there was one cluster that used the app most frequently, and for a longer duration, than the other clusters. We also selected two apps that featured prominently in the cluster analysis in terms of start dates and duration of usage: Slumber Time and MoveMe.

3.4.1 Daily Feats and Day to Day—Day to Day and Daily Feats both include features that encouraged daily user engagement. Daily Feats relies on checklists to encourage the user to incorporate worthwhile and productive activities in their day, and to give them credit for activities completed. The user is provided a list of activities to complete and check off each day, based on their level of distress, to which they can add their own personalized

activities. Consistent use is encouraged using streaks, in which consistent use over a number of days levels the user up to another set of activities[26].

In terms of mechanisms of change, Daily Feats assisted users to develop a routine (Table 8). It provided structure, the act of checking off tasks throughout the day led to a sense of accomplishment. This action also facilitated participants' self-awareness of their own actions. Participants found Daily Feats helpful and easy to use, and overall, strong usability concerns were not expressed.

Day to Day also encouraged daily interaction, but in a different manner. The Day to Day app provided people with a brief lesson each week, and then delivered a daily stream of tips and other information throughout the day to help the person implement the strategies described in the lesson[26]. Weekly themes included how to effectively cultivate gratitude, behavioral activation, increase social connectedness, problem solving, and challenge one's thinking.

Overall, many participants liked Day to Day. Mechanisms of change included instilling motivation and changing thought patterns. However, participants also reported issues with regard to the app's usability, which also resulted in confusion and irritation.

In comparison, both Daily Feats and Day to Day encouraged daily interactions, and their mechanisms of **change** resulted in daily reinforcement of the behavioral changes that they were designed to promote. As a checklist, Daily Feats involves a form of interaction in which the user perceives they have requested the support provided by the application, whereas Day to Day may provide support and/or information at times when participants perceive support is not needed or inconvenient. Additionally, navigation of the application seemed to be a source of confusion. Allowing the user to customize when they receive messages could be helpful, as well as providing more explicit instructions and/or more intuitive navigational controls. Lastly, Daily Feats instilled a sense of accomplishment that kept individuals engaged over time; one might also wonder whether the sense of accomplishment instilled by Daily Feats may be the primary intervention component influencing the steady improvement in outcomes witnessed in that cluster.

3.4.2   Slumber Time and MoveMe—We also consider participants' experiences with Slumber Time and MoveMe, two apps that featured prominently in the cluster analysis in terms of start dates and duration of usage. With both apps, overall there was little content in the text messages reflecting mechanisms of change. Rather, we observed greater reporting of problems and clarifying questions on usage.

Slumber Time provides functionality that is intended to improve participants' sleep hygiene. These include a sleep diary, a bedtime checklist, audio recordings to facilitate relaxation, and an alarm clock feature[27]. We did not observe discussion of mechanisms of change, though one of the participants did mention that their sleep habits had improved: "…my sleep habits have certainly improved from using the Slumber Time app. Specifically the leaving the phone and TV off at bedtime, using better background sounds, and trying to get up at about same time each morning. I do still need to work on trying to get to bed at more normal times."[1537]

Overall participants' messages concerning Slumber Time were primarily comprised of usability issues and the need for clarification on feature usage. With regard to the former, issues included errors such as receiving the same feedback or not receiving feedback. Other participants found certain aspects of Slumber Time inconvenient, such as the necessity of entering large amounts of information instead of having the information automatically detected. There were also many questions to clarify how to use the application. For example, participant 1363 asked what the sleep efficiency numbers indicated in terms of what they should be aiming for, and participant 1759 said that they thought it "would be nice to nice to see what causes me to sleep good vs bad."

MoveMe provides the user guidance in the selection of exercises to improve mood, including exercise videos, motivational lessons, and a function for scheduling exercises[27]. As with Slumber Time, there was little discussion of mechanisms of change. One participant did mention that they liked the coaching videos and found them helpful for learning how to do the exercises correctly (an example of skill building), but there were many others who felt that the exercises were too difficult. Participants also experienced usability issues relating to the scheduling function, such as difficulty getting reminders to pop up when they wanted them to. There were clarifying questions on usage, particularly relating to the scheduling function.

Returning to the app usage clusters, we observe that the Daily Feats Users used two apps for a longer duration: Daily Feats and MoveMe. As the apps both scheduling activities, it may be that Daily Feats Users appreciate this similarity in terms of the mechanisms by which they operate. The Low Usage cluster started Slumber Time earlier as compared to the rest of the sample, and one might wonder if, had Slumber Time fewer usability issues, engagement might have been higher with some members of that cluster. Overall, addressing usability concerns and providing clearer instructions for app use can be helpful to retain interest and engagement with the apps.

4 Discussion

In this study, we employed a three-pronged approach to study user engagement with an eclectic, skills-based app suite for the treatment of depression and anxiety. In Aim 1, considering engagement as usage, we identified four clusters of different types of users based on their app usage patterns: Low Usage, High Usage, Daily Feats Users, and Day to Day Users. In Aim 2, we rendered an interactive visualization, Message Highlight Map, to facilitate compare clusters in terms of engagement, conceptualized as both usage and user experience, and associated factors. Lastly, through qualitative analysis of participants' text messages, we considered the mechanisms of change and usability perceptions of the apps that were most salient in the cluster analysis.

We now consider the findings in terms of the potential for design recommendations for each cluster (Table 10). Daily Feats Users focused on the skills that they were learning and differences in their experiences. These participants appreciated concreteness, spoke of routines, and continuing to use the apps, so recommending apps that help develop structure could work well. Day to Day Users were particularly appreciative of the app's

encouragement to stay positive and to change thought patterns, as well as the support from the coaches. Thus, recommending other apps that stimulate thinking and provide inspiration can perhaps be helpful.

The Low Usage cluster offered less qualitative feedback in terms of their experience, but they appeared receptive to coaches. Developing ways for individuals to work in shorter activity bursts and/or assisting coaches in supporting these individuals might be particularly important. This might include improving the recommendations that are provided. One way to do so might be to recommend apps that are more personally relevant to participants (e.g. needing to improve sleep hygiene, to learn stress management techniques, etc.) by learning their needs earlier on, either through surveys or through inference from passively collected data. Another might be to learn mechanisms of change that might resonate with the participants, and offering recommendations that employ these mechanisms. There were multiple participants in this study that perceived Aspire and Daily Feats to be similar (as referenced in Table 8), suggesting that app recommendation through affinity of mechanism could be effective. Recommending Daily Feats and Day to Day earlier on, since they were the apps that participants seemed to engage with and derive the most benefit from, is perhaps another strategy.

The High Usage cluster used the apps most frequently and, overall, were enthusiastic about the intervention. While they appear to be highly engaged, it is perhaps important to ensure that the amount of time and effort they are spending is useful and effective, to ensure continued engagement and benefit. It might also be useful to consider whether there are potential adverse effects from too much use of the apps.

All of the clusters except Low Usage provided substantial feedback concerning both positive and negative aspects of usability. In our last aim, we conducted a deeper exploration of mechanisms of change and usability, specific to the apps, to inform future intervention design. Though our qualitative analyses provided evidence of intervention mechanisms at work, they also provided examples where usability issues may have interfered with these mechanisms of change.

4.1 Limitations and Future Directions

This study has a few limitations. First, there may be various reasons for users' app preferences, including predispositions towards app mechanisms, need for improvement in the area addressed by the app, and life context. In a future study, obtaining more detail on participants' life situations and needs would permit investigation of whether there are systematic associations these characteristics and their app usage patterns. This would allow for greater tailoring of app recommendations to user needs, preferences, and circumstances. Though in this study, we did not have access to daily self-assessments, it would also be interesting to examine associations between participants' affinities and usage for particular IntelliCare apps in relation to their daily states.

Second, there are various ways in which the Message Highlight Map might be improved. For example, our initial derivation of features was based on one lexicon. Other lexicons might also contain concepts of interest, and these would have been missed by our approach.

Additionally, we used proportions to colorize the cells of the Message Highlight Map and address the issue of differing cluster sizes. While this does facilitate the comparison, the smaller clusters are perhaps more susceptible to fluctuations in expressed density.

Lastly, the Message Highlight Map, as it currently exists, provides a coarse differentiation of features. For example, the dimension, "Mechanisms", does not provide a nuanced differentiation of mechanisms of change, which has been recognized as a need in the study of engagement of digital interventions[59]. One possible way to extend this work might to incorporate an ontological framework with natural language processing techniques to be able to identify and characterize mechanisms of change more precisely. However, we hope this visualization may serve as a basis to build upon for future work studying the relationship between constructs that are related to engagement.

4.2   Methodological Reflection

In this study, we presented a novel, mixed-methods approach to accomplish the overarching aim of understanding engagement, as usage and user experience, and potentially associated factors, particularly mechanisms of change and usability, through use and messaging data that are commonly acquired in digital mental health interventions.

The first step is use clustering, which identifies and segments subpopulations of users based on intervention usage behaviors. In our study, this was performed based on participants' usage of different apps. Traditionally, digital mental health interventions have used one application, however, these web and mobile applications tend to have multiple modules that could be similarly clustered. This first step recognizes the diversity in user preferences and patterns in the use of digital mental health tools.

In this case, we clustered users based on sessions of app use. Measuring engagement through logins or sessions is common in digital interventions (e.g. [15,21]). An alternative method could have been to cluster users based on characteristics other than engagement, or different types of engagement characteristics. For example, previous research that employed cluster analysis to study user engagement with a pain management app using variables relating to sociodemographic characteristics, medical history and medication regimen/use[38]. Cluster analysis using variables of different types could potentially lead to a tighter association between the features of interest. In this case, we regarded usage of specific apps as the primary characteristic for differentiation, and thus these were employed as the specific set of variables for analysis, with subsequent statistical comparisons on other demographic variables to expose other differences between the clusters.

The next step, message highlighting, identified salient characteristics of the messages sent by users. We developed a computer-assisted text analysis and visualization pipeline which grouped messages into categories based on constructs from engagement literature. These can then be compared across the app use clusters in a Message Highlight Map. This more granular understanding of messaging types in the context of app use patterns could be useful for a variety of purposes, such as understanding the needs or problems different types of users are having, or anticipating the types of messages coaches will have to respond to across different categories of users. An interactive visualization such as this might be used as

part of quality assurance or quality improvement. For example, the messages in the Usability section of the Message Highlight Map could be particularly helpful for identifying usability issues that need to be addressed, as well as opportunities to assist participants to adjust app behavior to suit their preferences. For example, a few participants wished the notifications were more obvious or that features would pop up on their own, whereas others felt that there were too many notifications.

The final step is qualitative analysis of data filtered based on insights from the previous two steps. In this case, we brought the analysis back to user interactions with specific intervention components to maximize insights for their effectiveness with respect to therapeutic outcomes, and improving the usability of these components. Messaging can provide a rich source of data in understanding the user through his or her interactions with the coach. This analysis provides a richer understanding of the user's experience, which is potentially useful for improving the tools or coaching protocol.

The method that we have demonstrated begins with the conceptualization of engagement as usage, segmenting users and characterizing them through their use behaviors. Previous research has observed that a critical challenge to the analysis of digital interventions is the organization of log data for analysis[25]. Prior work has often analyzed engagement using module accesses and website utilization (e.g. [14,25]). An interesting aspect of this dataset was that participants would not necessarily use intervention components in a certain sequence or time, nor were they recommended to use apps a specific amount. The visualizations that we employed were thus designed to expose differences in usage frequency, along usage dimensions of potential interest, such as the use of start date and duration of app use as potential indicators of interest and/or resonance with an application.

To understand engagement more holistically, we developed an interactive visualization that affords us a picture of users' engagement along dimensions of relevance to digital interventions. Previous methods to analyze text messages from digital interventions have often employed quantitative content analysis (e.g. [60,61]). Doing so often limits the analysis to a predefined set of categories. Our method involved a hybrid inductive-deductive approach, in which a lexicon was used to identify salient concepts, and then subsequent manual review was employed to assign relevant extracted concepts to categories of relevance. This approach afforded a systematic search of a concept space.

In our last aim, we employed a qualitative assessment approach focused on mechanisms of change and usability, which can serve to confirm that mechanisms of change are working as intended, and/or identifying opportunities to improve existing app designs. The findings concerning Daily Feats and Day to Day clearly illustrated differences in mechanisms, whereas the findings for all apps other than Daily Feats showed the need to address usability concerns, as well as provide additional clarification, often around those concerns. Iterative analysis and refinement cycles, incorporating both passively collected data and elicited user feedback, can facilitate continued improvement of the apps to suit participants' use patterns and preferences.

Recent literature has observed that there is an increased need to identify, characterize, and assess health-related apps[62,63]. In this study, we have demonstrated how theory-informed analysis of log data can be used to characterize engagement, factors associated with it, and implications for the design of digital tools with a therapeutic purpose. These methods have considerable utility in digital mental health research. They provide detailed information about specific groups of users that can be used to optimize both the applications as well as the coaching protocols, without having to subject participants to lengthy user feedback interviews. As the data are collected through the normal use of a digital mental health intervention, the data do not rely on memory or recollection; rather, they are generated in the context of treatment itself. Because it requires no additional effort on the part of the user, this method can be employed in a variety of contexts, including the course of ongoing services outside of a research setting.

5   Conclusion

In this study, we presented a novel mixed-methods approach for analysis of the complex and rich data that is routinely collected in digital mental health interventions. We performed a secondary analysis of IntelliCare, an intervention in which participants were able to try 14 different mental health apps over the course of eight weeks, using a three-pronged approach: use clustering, message highlighting, and qualitative analysis. This study contributes to extant methods for understanding users of digital interventions in multiple ways. First, the study illustrates the diversity and richness of the data that are being collected in digital interventions, and how these data might be used complementarily to characterize engagement. We were able to see that there can be differences within a sample in terms of overall engagement, as well as engagement with specific aspects of an intervention. Distinguishing among these enables us to explore whether there are aspects of an intervention that resonate particularly well with particular groups. This information might in turn be used to personalize intervention treatment and perform quality assurance/ improvement. Last, this study serves as an example of how passively collected data can be used to derive insights about behavior and experience in digital health contexts.

Acknowledgements

Appendix 1.: App Mentions

App	Mentions
Aspire	78
Boost Me	41
Daily Feats	76
Day to Day	85
Hub	35
Icope	43

App	Mentions
MeLocate	33
MoveMe	57
My Mantra	56
Purple Chill	76
Slumber Time	48
Social Force	32
Thought Challenger	65
Worry Knot	52
Total	777

7 References

1. Short CE, Rebar AL, Plotnikoff RC, Vandelanotte C. Designing engaging online behaviour change interventions: A proposed model of user engagement. Eur Health Psychol 2015;17(1):32–38.

2. Price M, Yuen EK, Goetter EM, Herbert JD, Forman EM, Acierno R, Ruggiero KJ. mHealth: A Mechanism to Deliver More Accessible, More Effective Mental Health Care. Clin Psychol Psychother 2014 9;21(5):427–436. [PubMed: 23918764]

3. Becker S, Miron-Shatz T, Schumacher N, Krocza J, Diamantidis C, Albrecht U-V. mHealth 2.0: Experiences, Possibilities, and Perspectives. JMIR MHealth UHealth [Internet] 2014 5 16 [cited 2016 Jan 3];2(2).

4. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on Internet-based mental health interventions — A systematic review. Internet Interv 2014 10;1(4):205–215. [doi: 10.1016/j.invent.2014.08.003]

5. Riper H, Andersson G, Christensen H, Cuijpers P, Lange A, Eysenbach G. Theme Issue on E-Mental Health: A Growing Field in Internet Research. J Med Internet Res [Internet] 2010 12 19;12(5).

6. Cunningham J, Gulliver A, Farrer L, Bennett K, Carron-Arthur B. Internet Interventions for Mental Health and Addictions: Current Findings and Future Directions. Curr Psychiatry Rep 2014;16(12): 1–5. [doi: 10.1007/s11920-014-0521-5]

7. Klasnja P, Consolvo S, Pratt W. How to evaluate technologies for health behavior change in HCI research Proc SIGCHI Conf Hum Factors Comput Syst [Internet] ACM; 2011 [cited 2017 Aug 7]. p. 3063–3072. Available from: http://dl.acm.org/citation.cfm?id=1979396

8. Schueller SM, Muñoz RF, Mohr DC. Realizing the Potential of Behavioral Intervention Technologies. Curr Dir Psychol Sci 2013 12;22(6):478–483. [doi: 10.1177/0963721413495872]

9. Andersson G, Carlbring P, Berger T, Almlöv J, Cuijpers P. What Makes Internet Therapy Work? Cogn Behav Ther 2009 1;38(sup1):55–60. [doi: 10.1080/16506070902916400] [PubMed: 19675956]

10. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 6;7(2):254–267. [doi: 10.1007/s13142-016-0453-1] [PubMed: 27966189]

11. Csikszentmihalyi M Flow: The Psychology of Optimal Experience. 1 edition New York: Harper Perennial Modern Classics; 1990 ISBN:978-0-06-133920-2

12. O'Brien HL, Toms EG. What is user engagement? A conceptual framework for defining user engagement with technology. J Am Soc Inf Sci Technol 2008 4;59(6):938–955. [doi: 10.1002/asi.20801]

13. Ritterband LM, Thorndike FP, Cox DJ, Kovatchev BP, Gonder-Frederick LA. A Behavior Change Model for Internet Interventions. Ann Behav Med Publ Soc Behav Med 2009 8;38(1):18–27.

14. Danaher BG, Boles SM, Akers L, Gordon JS, Severson HH. Defining Participant Exposure Measures in Web-Based Health Behavior Change Programs. J Med Internet Res [Internet] 2006 8 30;8(3).

15. Doherty G, Coyle D, Sharry J. Engagement with Online Mental Health Interventions: An Exploratory Clinical Study of a Treatment for Depression Proc SIGCHI Conf Hum Factors Comput Syst [Internet] New York, NY, USA: ACM; 2012 p. 1421–1430. [doi: 10.1145/2207676.2208602]

16. Baltierra NB, Muessig KE, Pike EC, LeGrand S, Bull SS, Hightow-Weidman LB. More than just tracking time: Complex measures of user engagement with an internet-based health promotion intervention. J Biomed Inform 2016 2 1;59:299–307. [doi: 10.1016/j.jbi.2015.12.015] [PubMed: 26732997]

17. Couper MP, Alexander GL, Maddy N, Zhang N, Nowak MA, McClure JB, Calvi JJ, Rolnick SJ, Stopponi MA, Little RJ, Johnson CC. Engagement and Retention: Measuring Breadth and Depth of Participant Use of an Online Intervention. J Med Internet Res 2010;12(4):e52. [doi: 10.2196/jmir.1430] [PubMed: 21087922]

18. Lehmann J, Lalmas M, Yom-Tov E, Dupret G. Models of User Engagement In: Masthoff J, Mobasher B, Desmarais MC, Nkambou R, editors. User Model Adapt Pers [Internet] Berlin, Heidelberg: Springer Berlin Heidelberg; 2012 [cited 2017 Oct 24]. p. 164–175. [doi: 10.1007/978-3-642-31454-4_14]

19. Kelders SM, Bohlmeijer ET, Van Gemert-Pijnen JE. Participants, Usage, and Use Patterns of a Web-Based Intervention for the Prevention of Depression Within a Randomized Controlled Trial. J Med Internet Res [Internet] 2013 8 20;15(8).

20. Neil A, Batterham P, Christensen H, Bennett K, Griffiths K. Predictors of Adherence by Adolescents to a Cognitive Behavior Therapy Website in School and Community-Based Settings. J Med Internet Res 2009;11(1):e6. [doi: 10.2196/jmir.1050] [PubMed: 19275982]

21. Strecher V, McClure J, Alexander G, Chakraborty B, Nair V, Konkel J, Greene S, Couper M, Carlier C, Wiese C, Little R, Pomerleau C, Pomerleau O. The Role of Engagement in a Tailored Web-Based Smoking Cessation Program: Randomized Controlled Trial. J Med Internet Res 2008;10(5):e36. [doi: 10.2196/jmir.1002] [PubMed: 18984557]

22. Nijland N, van Gemert-Pijnen JE, Kelders SM, Brandenburg BJ, Seydel ER. Factors Influencing the Use of a Web-Based Application for Supporting the Self-Care of Patients with Type 2 Diabetes: A Longitudinal Study. J Med Internet Res [Internet] 2011 9 30;13(3).

23. Owen JE, Jaworski BK, Kuhn E, Makin-Byrd KN, Ramsey KM, Hoffman JE. mHealth in the Wild: Using Novel Data to Examine the Reach, Use, and Impact of PTSD Coach. JMIR Ment Health [Internet] 2015 3 25;2(1).

24. Van Gemert-Pijnen JE, Kelders SM, Bohlmeijer ET. Understanding the Usage of Content in a Mental Health Intervention for Depression: An Analysis of Log Data. J Med Internet Res 2014 1 31;16(1):e27. [doi: 10.2196/jmir.2991] [PubMed: 24486914]

25. Morrison C, Doherty G. Analyzing Engagement in a Web-Based Intervention Platform Through Visualizing Log-Data. J Med Internet Res [Internet] 2014 11 13;16(11).

26. Lattie EG, Schueller SM, Sargent E, Stiles-Shields C, Tomasino KN, Corden ME, Begale M, Karr CJ, Mohr DC. Uptake and usage of IntelliCare: A publicly available suite of mental health and well-being apps. Internet Interv 2016 5;4, Part 2:152–158. [doi: 10.1016/j.invent.2016.06.003]

27. Mohr DC, Tomasino KN, Lattie EG, Palac HL, Kwasny MJ, Weingardt K, Karr CJ, Kaiser SM, Rossom RC, Bardsley LR, Caccamo L, Stiles-Shields C, Schueller SM. IntelliCare: An Eclectic, Skills-Based App Suite for the Treatment of Depression and Anxiety. J Med Internet Res [Internet] 2017 1 5;19(1).

28. Kwasny MJ, Schueller SM, Lattie EG, Gray E, Mohr DC. IntelliCare: Towards managing a platform of mental health apps.

29. Kaiser SM. IntelliCare Study Coaching Manual [Internet]. 2016 Available from: https://digitalhub.northwestern.edu/files/00fa4294-5b9f-4afc-897a-7fffceae8f3f

30. Schueller SM, Tomasino KN, Lattie EG, Mohr DC. Human Support for Behavioral Intervention Technologies for Mental Health: The Efficiency Model. :4.

31. Mohr D, Cuijpers P, Lehman K. Supportive Accountability: A Model for Providing Human Support to Enhance Adherence to eHealth Interventions. J Med Internet Res 2011;13(1):e30. [doi: 10.2196/jmir.1602] [PubMed: 21393123]

32. Honkanen P, Frewer L. Russian consumers' motives for food choice. Appetite 2009;52(2):363–371. [PubMed: 19073227]

33. Honkanen P, Frewer L. Russian consumers' motives for food choice. Appetite 2009 4;52(2):363–371. [doi: 10.1016/j.appet.2008.11.009] [PubMed: 19073227]

34. Kizilcec RF, Piech C, Schneider E. Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. Proc Third Int Conf Learn Anal Knowl ACM; 2013 p. 170–179.

35. Wise AF, Marbouti F, Speer J, Hsiao Y-T. Towards an understanding of 'listening'in online discussions: A cluster analysis of learners' interaction patterns. Connect Comput Support Collab Learn Policy Pract CSCL2011 Conf Proceeding 2011 p. 88–95.

36. Holden RJ, Kulanthaivel A, Purkayastha S, Goggins KM, Kripalani S. Know thy eHealth user: Development of biopsychosocial personas from a study of older adults with heart failure. Int J Med Inf 2017 12 1;108:158–167. [doi: 10.1016/j.ijmedinf.2017.10.006]

37. Vosbergen S, Mulder-Wiggers JMR, Lacroix JP, Kemps HMC, Kraaijenhagen RA, Jaspers MWM, Peek N. Using personas to tailor educational messages to the preferences of coronary heart disease patients. J Biomed Inform 2015 2;53:100–112. [PubMed: 25239261]

38. Rahman QA, Janmohamed T, Pirbaglou M, Ritvo P, Heffernan JM, Clarke H, Katz J. Patterns of User Engagement With the Mobile App, Manage My Pain: Results of a Data Mining Investigation. JMIR MHealth UHealth 2017;5(7):e96. [doi: 10.2196/mhealth.7871] [PubMed: 28701291]

39. Jain AK. Data clustering: 50 years beyond K-means. Pattern Recognit Lett 2010 6;31(8):651–666. [doi: 10.1016/j.patrec.2009.09.011]

40. Thorndike R Who belongs in the family? Psychometrika 1953;18(4):267–276.

41. Yang J, Ning C, Deb C, Zhang F, Cheong D, Lee SE, Sekhar C, Tham KW. k-Shape clustering algorithm for building energy usage patterns analysis and forecasting model accuracy improvement. Energy Build 2017 7;146:27–37. [doi: 10.1016/j.enbuild.2017.03.071]

42. Khan SS, Ahmad A. Cluster center initialization algorithm for K-means clustering. Pattern Recognit Lett 2004 8;25(11):1293–1302. [doi: 10.1016/j.patrec.2004.04.007]

43. Chan Y, Walmsley RP. Learning and Understanding the Kruskal-Wallis One-Way Analysis-of-Variance-by-Ranks Test for Differences Among Three or More Independent Groups. Phys Ther 1997 12 1;77(12):1755–1761. [doi: 10.1093/ptj/77.12.1755] [PubMed: 9413454]

44. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7. Arch Intern Med 2006 5 22;166(10):1092–1097. [doi: 10.1001/archinte.166.10.1092] [PubMed: 16717171]

45. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9. J Gen Intern Med 2001 9 1;16(9):606–613. [doi: 10.1046/j.1525-1497.2001.016009606.x] [PubMed: 11556941]

46. Riloff E, Wiebe J. Learning extraction patterns for subjective expressions. Proc 2003 Conf Empir Methods Nat Lang Process Association for Computational Linguistics; 2003 p. 105–112.

47. Charmaz K Constructing grounded theory. 2nd ed London ; Thousand Oaks, Calif: Sage; 2014 ISBN:978-0-85702-913-3

48. Saldaña J The coding manual for qualitative researchers. Los Angeles, Calif: Sage; 2009 ISBN: 978-1-84787-548-8

49. Thomas DR. A General Inductive Approach for Analyzing Qualitative Evaluation Data. Am J Eval 2006 6 1;27(2):237–246. [doi: 10.1177/1098214005283748]

50. Corbin JM, Strauss A. Grounded theory research: Procedures, canons, and evaluative criteria. Qual Sociol 1990;13(1):3–21.

51. O'Brien HL, Toms EG. Examining the Generalizability of the User Engagement Scale (UES) in Exploratory Search [Internet]. University of British Columbia; 2013 [cited 2017 May 28]. Available from: https://open.library.ubc.ca/cIRcle/collections/facultyresearchandpublications/52383/items/1.0107442

52. Bennett-Levy J, Richards D, Farrand P, Christensen H, Griffiths K, Kavanagh D, Klein B, Lau MA, Proudfoot J, Ritterband L, White J, Williams C. Oxford Guide to Low Intensity CBT Interventions. OUP Oxford; 2010 ISBN:978-0-19-102935-6

53. Schubart JR, Stuckey HL, Ganeshamoorthy A, Sciamanna CN. Chronic Health Conditions and Internet Behavioral Interventions: A Review of Factors to Enhance User Engagement. CIN Comput Inform Nurs 2011 2;29(2):81–92. [doi: 10.1097/NCN.0b013e3182065eed] [PubMed: 21164337]

54. Zapata BC, Fernández-Alemán JL, Idri A, Toval A. Empirical Studies on Usability of mHealth Apps: A Systematic Literature Review. J Med Syst 2015 1 20;39(2):1. [doi: 10.1007/s10916-014-0182-2] [PubMed: 25600193]

55. Bevan N, Carter J, Earthy J, Geis T, Harker S. New ISO Standards for Usability, Usability Reports and Usability Measures In: Kurosu M, editor. Hum-Comput Interact Theory Des Dev Pract [Internet] Cham: Springer International Publishing; 2016 [cited 2018 Mar 8]. p. 268–278. [doi: 10.1007/978-3-319-39510-4_25]

56. Calisir F, Calisir F. The relation of interface usability characteristics, perceived usefulness, and perceived ease of use to end-user satisfaction with enterprise resource planning (ERP) systems. Comput Hum Behav 2004 7 1;20(4):505–515. [doi: 10.1016/j.chb.2003.10.004]

57. Simmons WM, Zoetewey MW. Productive Usability: Fostering Civic Engagement and Creating More Useful Online Spaces for Public Deliberation. Tech Commun Q 2012 7;21(3):251–276. [doi: 10.1080/10572252.2012.673953]

58. Dumas J, McGee M, Rich A. Understanding the Usability Construct: User-Perceived Usbility. Th Annu Meet :5.

59. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and Evaluating Digital Interventions to Promote Behavior Change in Health and Health Care: Recommendations Resulting From an International Workshop. J Med Internet Res [Internet] 2017 6 29;19(6). PMID: 28663162

60. Abroms LC, Padmanabhan N, Thaweethai L, Phillips T. iPhone Apps for Smoking Cessation: A Content Analysis. Am J Prev Med 2011 3 1;40(3):279–285. [doi: 10.1016/j.amepre.2010.10.032] [PubMed: 21335258]

61. Paxling B, Lundgren S, Norman A, Almlöv J, Carlbring P, Cuijpers P, Andersson G. Therapist Behaviours in Internet-Delivered Cognitive Behaviour Therapy: Analyses of E-Mail Correspondence in the Treatment of Generalized Anxiety Disorder. Behav Cogn Psychother 2013 5;41(03):280–289. [doi: 10.1017/S1352465812000240] [PubMed: 22717145]

62. Grundy QH, Wang Z, Bero LA. Challenges in Assessing Mobile Health App Quality. Am J Prev Med 2016 12;51(6):1051–1059. [doi: 10.1016/j.amepre.2016.07.009] [PubMed: 27659122]

63. Paglialonga A, Lugo A, Santoro E. An overview on the emerging area of identification, characterization, and assessment of health apps. J Biomed Inform 2018 7 1;83:97–102. [doi: 10.1016/j.jbi.2018.05.017] [PubMed: 29852318]
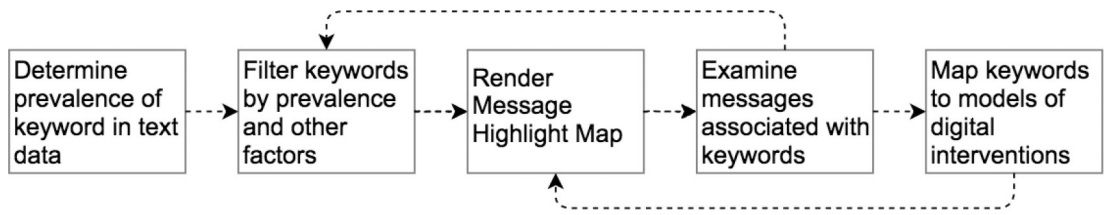
Figure 1.
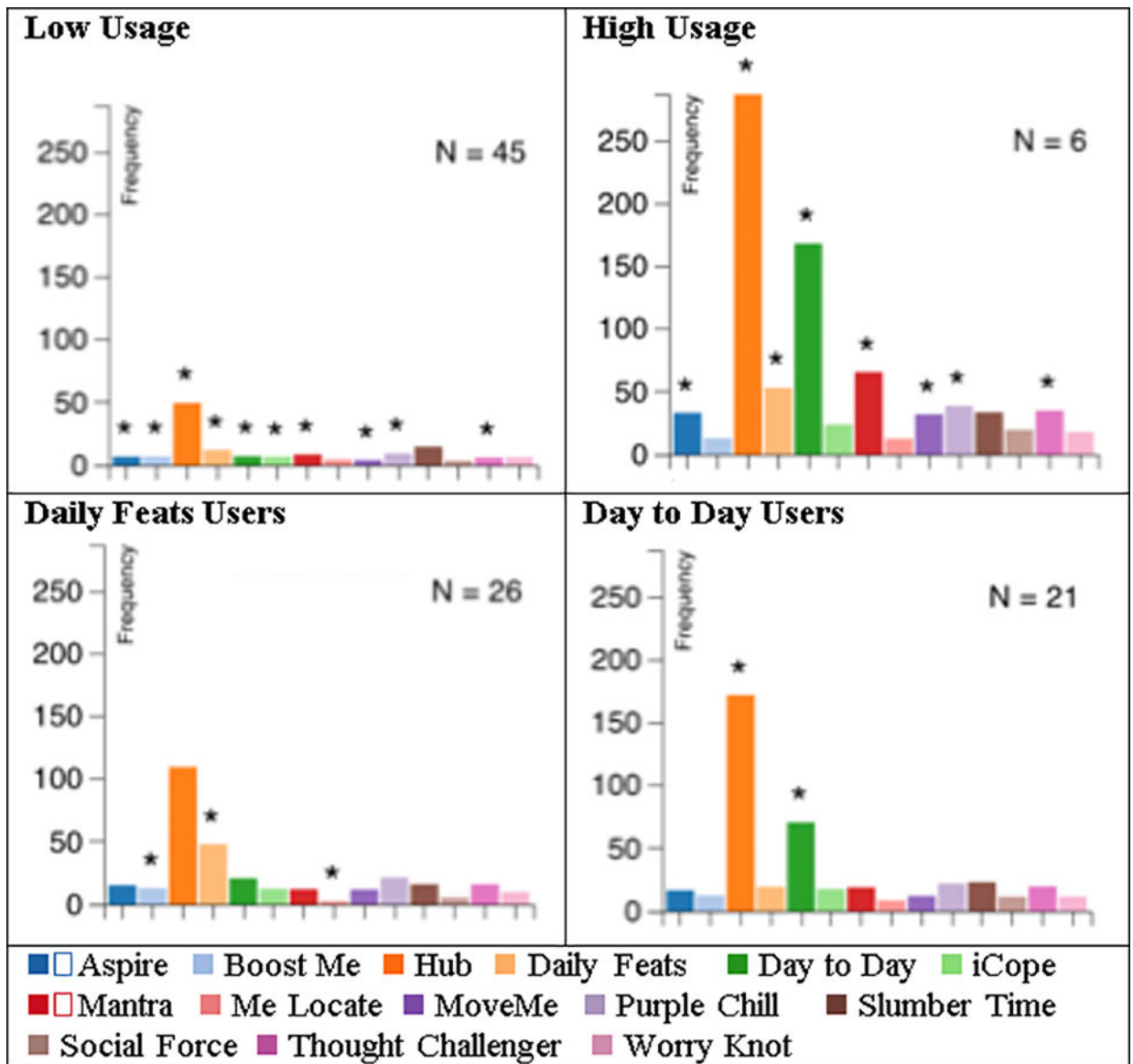Message Highlighting: Computer-Assisted Discovery of Salient User Experience Characteristics

Figure 2.
App Usage Clusters
Note: Asterisks indicate statistical significance at the .05 level between the cluster and the rest of the sample.
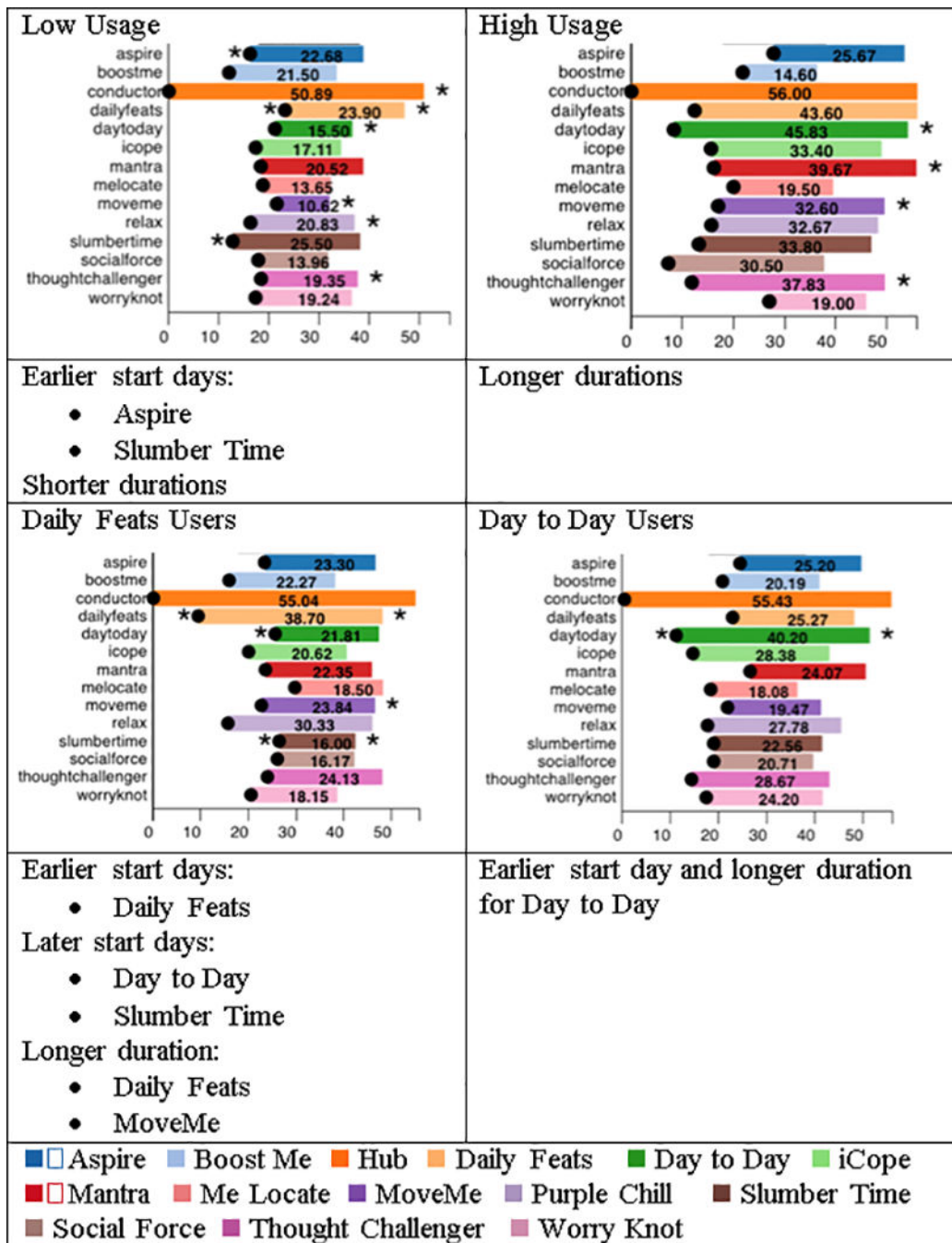
Figure 3.

Average Start Day and App Duration for Cluster Groups

Note: Asterisks indicate statistical significance at the .05 level between the cluster and the rest of the sample.
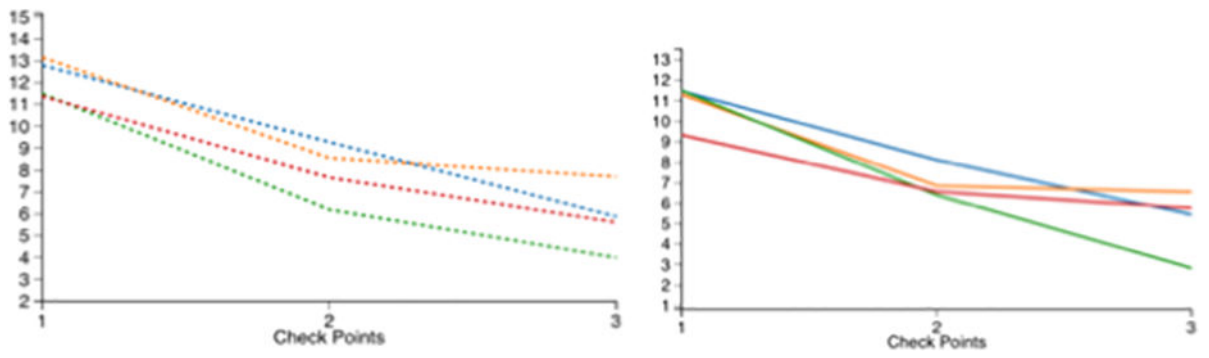
Figure 4.
Mean outcomes of each cluster, PHQ-9 (left), ■ GAD-7 (right). ■ Daily Feats Users, ■ Low Usage, High Usage, ■ Day to Day Users.
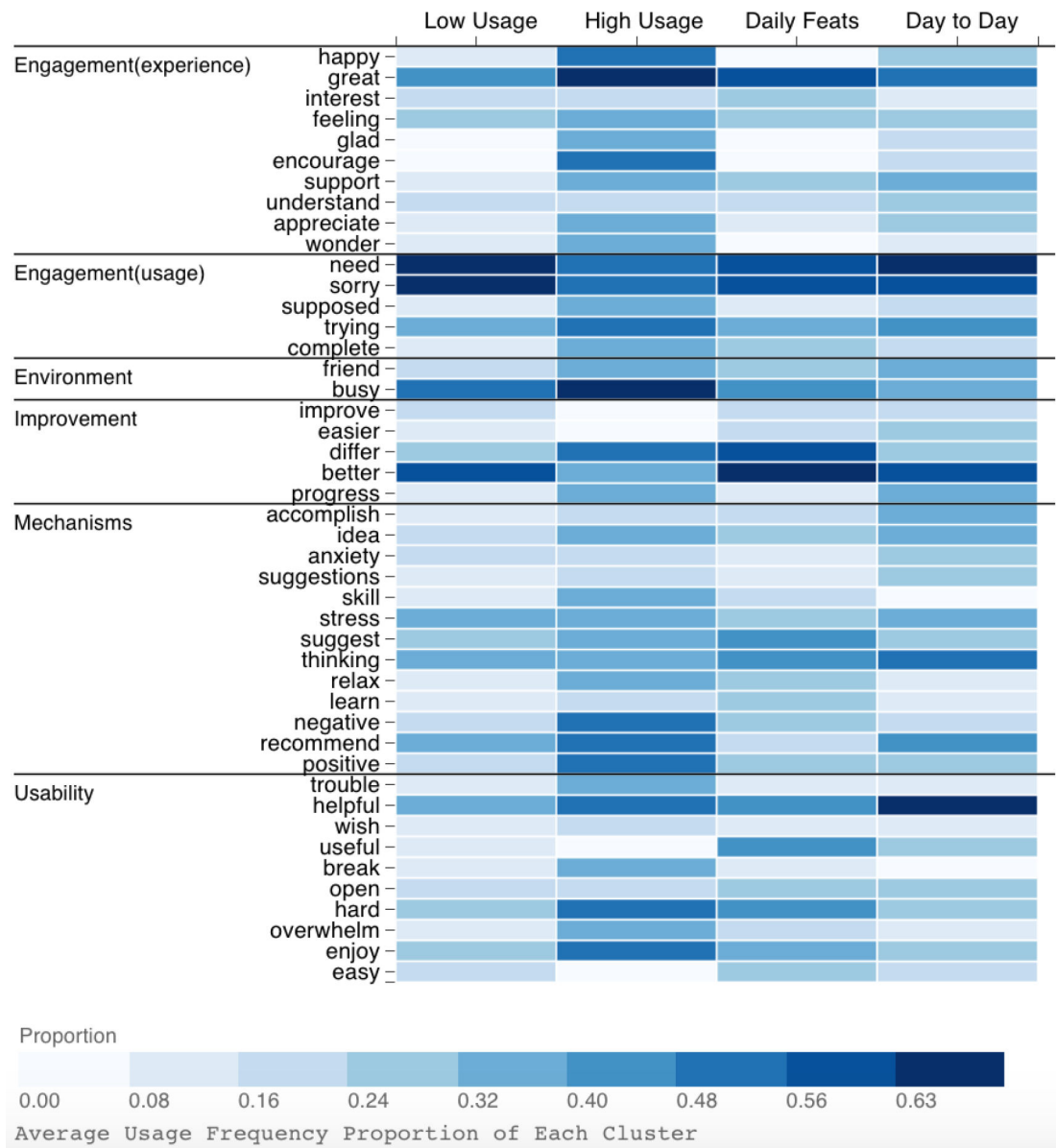
Figure 5.
Message Highlight Map

Table 1.

Description and Behavioral Strategy of the IntelliCare Apps. Reprinted from [27].

App	Behavioral strategy	Description
IntelliCare Hub		Manages messages and notifications from the other apps within the IntelliCare collection.
Aspire	Personal values and goal setting	Guides user to identify the values that guide one's life and the actions (or "paths") that one does to live that value. Helps keep track of those actions throughout the day and supports the user in living a more purpose-driven and satisfying life.
Day to Day	Psychoeducation and prompts	Delivers a daily stream of tips, tricks, and other information throughout the day to boost the user's mood. Prompts the user to work on a particular theme each day and every week; learn more about how to effectively cultivate gratitude, activate pleasure, increase connectedness, solve problems, and challenge one's thinking.
Daily Feats	Goal setting	Encourages the user to incorporate worthwhile and productive activities into the day. Users add accomplishments to the Feats calendar, where they can track their positive activity streaks and level up by completing more tasks. Helps motivate users to spend their days in more meaningful, rewarding ways to increase overall satisfaction in life.
Worry Knot	Emotional regulation and exposure	Teaches the user to manage worry with lessons, distractions, and a worry management tool. Provides a guided tool to address specific problems that a user cannot stop thinking about and provides written text about how to cope with "tangled thinking." Presents statistics about progress as the user practices coping with worry, gives daily tips and tricks about managing worry, and provides customizable suggestions for ways to distract oneself.
ME Locate	Behavioral activation	Provides a personal map for finding and saving user's mood-boosting locations. Assists the user in finding and remembering these places to help them make plans, maintain a positive mood, and stay on top of responsibilities.
Social Force	Social support	Prompts the user to identify supportive people in their lives, and provides encouragement for the user to get back in touch with those positive people.
My Mantra	Self-affirmations and positive reminiscence	Prompts the user to create mantras (or repeatable phrases that highlight personal photo strengths and values and can motivate one to do and feel good) and construct virtual albums to serve as encouragement and reminders of these mantras.
Thought Challenger	Cognitive reframing	Guides the user through an interactive cognitive restructuring tool to examine thoughts that might exaggerate negative experiences, lead one to be overcritical, and bring down one's mood. Teaches the user to get into the habit of changing perspective and moving toward a more balanced outlook on life.
iCope	Proactive coping	Allows the user to send oneself inspirational messages and reassuring statements, written in their own words, to help the user get through tough spots or challenging situations.
Purple Chill	Relaxation	Provides users with a library of audio recordings to relax and unwind. Teaches a variety of relaxation and mindfulness practices to destress and worry less.
MoveMe	Exercise for mood	Helps the user select exercises to improve mood. Provides access to curated exercise videos and to written lessons about staying motivated to exercise. Allows the user to schedule motivational exercise time for oneself throughout the week.
Slumber Time	Sleep hygiene	Prompts the user to complete sleep diaries to track sleep. Provides a bedtime checklist intended to clear one's mind before going to sleep. Provides audio recordings to facilitate rest and relaxation. Features an alarm clock function.
Boost Me	Behavioral activation	Encourages users to select and schedule positive activities ("boosts") when they notice a drop in mood and to track positive activities they note positively impacting their mood. Includes animated mood tracking for pre or post positive activities, calendar integration, and suggested activities that are auto-populated based on past mood improvement.

Table 2.

Cluster Analysis Variables and Comparison Dimensions

Variable Type	Measure
Cluster analysis variables	
Cluster analysis variables	Number of app usage sessions for each app
Comparison dimensions	
Overall engagement	Total number of app usage sessions
	Total number of messages sent
	Proportion of recommendations followed
App-specific engagement	App start date
	App usage duration
Outcomes	GAD-7
	PHQ-9

Table 3.

Sample Characteristics

Sample Characteristic	Statistic
# Participants	98
# Coaches	4
# Messages	10,227
Sender	
Coach	6,800
Participant	3,427
Messages Per Study ID	M(SD)
Messages	104.36(37.32)
Messages sent by coach	69.39(21.30)
Messages sent by participant	34.97(18.70)

Table 4.

App Usage (# Sessions)

App	Mean	SD
Hub	106.5	71.2
Day to Day	34.4	48.5
Daily Feats	25.9	32.2
Slumber Time	18.2	21.4
Purple Chill	17.2	16.9
Mantra	15.4	21.9
Thought Challenger	13.5	16.3
Aspire	12.9	15.2
iCope	11.7	20.7
Boost Me	10.2	11.1
MoveMe	9.8	12.9
Worry Knot	9.2	10.7
Social Force	6.9	11.2
Me Locate	5.6	8.5

Table 5.

Cluster-Wise Comparison of Overall Engagement Metrics

	Low Usage	High Usage	Daily Feats	Day to Day	Statistic
# App Sessions	148.8 (79.9)	836.0 (204.9)	316.7 (76.4)	438.8 (98.4)	$X^2(3)= 72.9$, $p= .000$
# Messages Sent	96.9 (41.6)	103.3 (26.2)	118.5 (36.7)	103.1 (26.4)	$X^2(3)= 4.6$, $p= 0.20$
% Recs Followed	19.0	14.9	23.4	15.9	$X^2(3)= 5.7$, $p= .13$

Table 6.

Key Constructs and Definitions Relating to Engagement of Digital Interventions

Construct	Definition	Source
Engagement (experience)	Users' experience of the program, including the support from coaches	[12,51]
Engagement (usage)	Factors relating participants' efforts to complete elements of the intervention as prescribed	[52,53]
Environment	External factors affecting engagement	[1]
Improvement	Symptom improvement	[13]
Mechanisms	Intervention mechanisms, content and skills	[13]
Usability	Usability perceptions	[1]

Table 7.

Examples of the Salient Concepts from Participants' Text Messages

Category/Keyword	Description	Example
Engagement (experience)		
Appreciate	Expressed appreciation for the support of the coaches	I really appreciate the help, support and guidance that you have given over the past 2 months.[1537]
Great	Expressed enthusiasm for the program and the coaches	It was great having you in my side helping me with this… [1669]
Engagement (usage)		
Trying	Demonstrative of efforts to engage	I've really been trying to use them consistently! I think it's going ok! [1012]
Need	Participants stating or asking what they need to do	I am getting used to them and need to spend more time on the apps. [1362]
Sorry	Apologetic sentiment for not responding to the coaches.	Thank you for all your help. I'm sorry, I've had a long day at work today so I didn't have a chance to get back to you. Take care! [1778]
Environment		
Busy	Explanation of participants' lack of time because of their schedules.	Sorry for the late response! It has been a crazy busy day… [1012]
Improvement		
Better	Improvement in symptoms or a suggestion regarding usability.	My mood is a lot better…[1257] I feel like mostly I have been able to identify triggers in a better way than before using some of the apps, as well as having more coping skills. [1577]
Mechanisms		
Relax	Learning relaxation techniques.	I have learned to reach out more to others, I have been taking the time to actually relax, energy im still trying to figure out. I have started to cope with things in different ways [1668]
Thinking	Talking about counteracting negative thinking or what participants think that they need to do.	I get stuck in negative thinking less often, and it's easier to counter them with more reasonable self talk. It's easier to reassure myself when I need to [1317]
Usability		
Easy	An app is easy to use	It's pretty easy to use, so I don't have questions about it. I know you'll recommend to start using another throughout each week, but could I do so if I'm curious about an app? [1722]
Open	Apps open when users do not want them to, or do not open when they are supposed to	The hub app won't open when I hit the icon my screen goes white for a second then goes back to my home screen [1263]
Helpful	What is helpful about the apps.	Yes, they are helpful in reinforcing the good things I did that day (even if they are small things) and that little steps can add up. [1376]

Table 8.

Daily Feats and Day to Day: Mechanisms of Change and Usability Perceptions

Mechanism of Change/Usability	Examples
Daily Feats	
Development of routine	• I think it's the structure and tangibility of the lists in Daily Feats that I'm motivated to accomplish and checkoff everyday. [1767] •The daily feats give me goals through out the day I can check off and feel good about myself [1257]
Sense of accomplishment	• I feel better about my day when I can see that I actually have accomplished something, like the list in daily feats and aspire. [1355] • I find it helpful to be encouraged to accomplish the goals I set. [1187]
Self-awareness	• Daily Feats helps me notice when I make healthy choices. [1114]
Easy to use	• I like aspire and daily feats lately. They are easy to use…
Day to Day	
Instilling motivation	• I liked 'day to day' it was nice to make me think about working on things. [1365]
Changing thought patterns	• I'm really finding the Day To Day app beneficial, it's making me think in different ways I didn't think of before. [1739] • i just wanted to tell u th[a]t the day to day app has really been helpful this week. Talking about negative thinking and ways to change it. [1718]
Interaction patterns led to confusion and irritation	• Day to day is a little confusing in terms of how you get to the next level. i see the messages that are sent but unsure how to get to the next item. [1362] • The day to day app can be a bit overbearing with the notifications and suggestions. [1433]

Table 9.

Slumber Time and MoveMe: Mechanisms of Change and Usability Perceptions

Mechanism of Change/Usability	Examples
Slumber Time	
Usability concerns	• The alarm clock feature doesn't seem very convenient on the slumber time app. My phone doesn't naturally lay on its side. Otherwise no problems. [1365]
The need for clarification	• Is there a feedback mode for slumber time, or is it just the sleep efficiency numbers? Should I be striving for 100%? [1363] • just wondering if there is a way to put notes in on sleep diary (such as weird dreams) in Slumber Time. [1537]
MoveMe	
Skill building	• Great coaching videos with the MoveMe app. It helps to learn how to do a specific exercise right.
App could be better suited to participants' needs	moveme it would be cool if move me now had easier things like walk around your office once instead of things like lunges and crunches and planks… [1074]
Usability concerns	• I'm getting a notification that there is an update to move me but when I go to the store there isn't an update. So the notification won't go away. [1074]
The need for clarification	• I have found the activity scheduling but am wondering if there is a way to make it a recurring event like on my calendar. [1537]

Table 10.

Sample and Cluster Characteristics with Design Recommendations

	Qualitative Feedback	**Recommendations**
All	• Found the apps helpful, but also encountered usability issues • Apologetic sentiment for being busy	• Address usability issues
Low Usage	• Comparatively less qualitative feedback, but they still engaged with intervention mechanisms	• Develop ways for individuals to work in short activities • Find ways to help coaches support these individuals • Recommend Daily Feats and Day to Day earlier on
High Usage	• Enthusiastic about the intervention	• Improve the quality of the recommendations • Investigate effective intervention components
Daily Feats	• Focused on skill-building	• Apps that provide structure and instill a sense of accomplishment work well
Day to Day	• Felt that the apps were helping them to think positively • Appreciative of the support and encouragement from coaches	• Recommend apps that emphasize thinking about problems in new ways