Assessing Digital Phenotyping for App Recommendations and Sustained Engagement: A Pilot Study

Bridget Dwyer, Matthew Flathers, James Burns, Jane Mikkelson, Elana Perlmutter, Kelly Chen, Nanik Ram, John Torous

Submitted to: JMIR Formative Research
on: May 29, 2024

Table of Contents

Assessing Digital Phenotyping for App Recommendations and Sustained Engagement: A Pilot Study

Bridget Dwyer[1] BS; Matthew Flathers[1] BA; James Burns[1] BS; Jane Mikkelson[1] BA; Elana Perlmutter[1] BS; Kelly Chen[1] MS; Nanik Ram[1] MD; John Torous[1] MD

[1]Division of Digital Psychiatry Beth Israel Deaconess Medical Center Harvard Medical School Boston US

Corresponding Author:
John Torous MD
Division of Digital Psychiatry
Beth Israel Deaconess Medical Center
Harvard Medical School
330 Brookline Ave
Boston
US

Abstract

Background: Low engagement with mental health apps continues to limit their impact. New approaches to helping match patients to the right app may increase engagement by ensuring the app they are using is best suited to their mental health needs.

Objective: This study piloted how digital phenotyping, using data from smartphone sensors to infer symptom, behavioral and functional outcomes, could be used to match people to mental health apps and potentially increase engagement

Methods: After one week of collecting digital phenotyping data with the mindLAMP app, participants were randomized to feedback and recommendations based on that data for selecting one of four predetermined mental health apps (related to mood, anxiety, sleep, fitness) or in the control arm, selecting the same apps but without any feedback or recommendations. All participants used their selected app for 4 weeks with numerous metrics of engagement recorded including objective screentime measures, self reported engagement measures, and Digital Working Alliance Inventory scores. The study offered participants no compensation.

Results: 82 enrolled in the study and 17 dropped out of the digital phenotyping arm and 18 from the control arm. Across both groups, few participants chose or were recommended the insomnia or fitness app. The majority 78.7% used a depression or anxiety app. Engagement as measured by objective screen time and Digital Working Alliance Inventory scores were higher in the digital phenotyping arm. There was no correlation between self-reported and objective metrics of app use. Qualitative results highlighted the importance of habit formation in sustained app use.

Conclusions: Results suggest that digital phenotyping app recommendation is feasible and may increase engagement. This approach is generalizable to other apps beyond the four selected for use in this pilot, and practical for real-world use given the study was conducted without any compensation or external incentives that may have biased results. Advances in digital phenotyping will likely make this method of app recommendation more personalized and thus of even greater interest.

(JMIR Preprints 29/05/2024:62725)
DOI: https://doi.org/10.2196/preprints.62725

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
✔ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.
2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?
✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Assessing Digital Phenotyping for App Recommendations and Sustained Engagement: A Pilot Study

Bridget Dwyer[1] BS, Matthew Flathers[1] BA, James Burns[1] BS, Jane Mikkelson[1] BA, Kelly Chen[1] MS, Elana Perlmutte[1] BS, Nanik Ram[1] MD, John Torous[1] MD MBI

1. Beth Israel Deaconess Medical Center, Department of Psychiatry, Harvard Medical School, Boston, MA, United States

ABSTRACT

Background: Low engagement with mental health apps continues to limit their impact. New approaches to helping match patients to the right app may increase engagement by ensuring the app they are using is best suited to their mental health needs. This study piloted how digital phenotyping, using data from smartphone sensors to infer symptom, behavioral and functional outcomes, could be used to match people to mental health apps and potentially increase engagement

Methods: After one week of collecting digital phenotyping data with the mindLAMP app, participants were randomized to feedback and recommendations based on that data for selecting one of four predetermined mental health apps (related to mood, anxiety, sleep, fitness) or in the control arm, selecting the same apps but without any feedback or recommendations. All participants used their selected app for 4 weeks with numerous metrics of engagement recorded including objective screentime measures, self reported engagement measures, and Digital Working Alliance Inventory scores. The study offered participants no compensation.

Results: 82 enrolled in the study and 17 dropped out of the digital phenotyping arm and 18 from the control arm. Across both groups, few participants chose or were recommended the insomnia or fitness app. The majority 78.7% used a depression or anxiety app. Engagement as measured by objective screen time and Digital Working Alliance Inventory scores were higher in the digital phenotyping arm. There was no correlation between self-reported and objective metrics of app use. Qualitative results highlighted the importance of habit formation in sustained app use.

Discussion: Results suggest that digital phenotyping app recommendation is feasible and may increase engagement. This approach is generalizable to other apps beyond the four selected for use in this pilot, and practical for real-world use given the study was conducted without any compensation or external incentives that may have biased results. Advances in digital phenotyping will likely make this method of app recommendation more personalized and thus of even greater interest.

INTRODUCTION

While COVID-19 accelerated interest in mental health smartphone apps, limited patient use and engagement with these apps has emerged as a primary barrier for successful uptake [1, 2]. While the challenge of limited engagement has already been well documented and ascribed to numerous causes ranging from individual patient preferences to healthcare system barriers, there have been fewer efforts seeking to actually improve engagement [3, 4]. This study pilots one approach, digital navigator guided app recommendation, to increase engagement and seeks to address methodological challenges with prior studies through objectively assessing app usage.

The challenges of low engagement with mental health apps have been well known for nearly a decade. A landmark 2019 study of 93 mental health related smartphone apps found that the median 15-day retention rate was 3.9% [5]. Numerous other studies confirm exponential decay in app engagement, regardless of health condition, or age/gender/race of users [6, 7]. These low engagement numbers are further exacerbated by low initial utilization of mental health apps. A 2023 survey of US veterans noted that while up to 76% reported apps are important for their mental health, only 5% ever reported having tried an app at least once [8].

Yet appreciating the challenge of app engagement does not in itself offer actionable solutions. Recent reviews have covered broad reasons why people often do not download mental health apps as well as why they rapidly stop using them if they do download them [3]. Common themes raised to boost engagement often include the need for personalized app experiences, social and/or therapeutic support, customization, in-app guidance, and use of sensors to offer users real time feedback [3]. Yet awareness of such themes raises the question: will implementing these themes actually increase engagement? And if such a solution can work, will that method of increasing engagement be generalizable given over 10,000 mental health apps exist today and research specific to each unique app is not practical or feasible [9].

One promising approach towards increasing engagement is the impact of digital navigator guided app recommendation. A digital navigator is a member of the care team trained to perform digital health roles related to equity, digital literacy, app selection, and engagement [10]. There is strong research data to suggest that patients would like guidance from clinical teams around selecting an app [11, 12, 13]. Yet clinical teams are not aware of where to find evidence based mental health apps and even fewer how to evaluate them [14]. Indeed, in numerous surveys clinical teams note that they actually want education around app evaluation [13, 15, 16]. While several large healthcare systems have begun to offer app toolkits for their clinical teams to use with patients [17, 18], efforts to help clinical teams recommend apps remain limited. One prior study found little impact of guidance on sustained engagement, but in this study participants were limited to picking exercises within a single app platform [19] that subsequently was shown to itself suffer from low uptake/engagement [20]. A prior study by our own team found that guided app recommendation did increase engagement with apps [21] as compared to national rates.

However, no prior study has examined the impact of digital phenotyping on app recommendation. This involves accessing sensors on a patient’s smartphone to capture data related to behaviors (eg sleep, mobility), cognition (eg memory), and self reported symptoms to better understand a patient’s state and use that information to match them to the best app for that state. For example, a patient who reports depression while at home may benefit from a cognitive behavioral therapy focused app and another who reports anxiety at work may benefit from a different app offering brief mindfulness exercises. Digital phenotyping methods can also be used to predict changes in anxiety and depression [22], meaning that it may be possible to suggest mental health app use early and as a preventive approach.

In piloting how digital phenotyping may help improve app recommendation, there are many metrics to consider. The most important may be engagement, as without engagement even the most effective app will not be impactful. Unfortunately, recent reviews confirm that there is no standard approach towards measuring engagement, with the most common method to measure percentage of patients who complete available modules [13]. This is problematic as not all apps have modules and the completion of modules may not always signify clinically meaningful engagement. Alternative means to measure engagement include time spent in the app and subjective reports of engagement. Yet other means to assess engagement include newer metrics like the Digital Working Alliance Inventory (D-WAI), which assess the degree of alliance a user has to an app and has previously been shown to predict app engagement and outcomes [23]. Thus in this study, we focused on multiple means to measure engagement with the secondary aim of assessing how the measurement of engagement itself, via subjective and objective metrics, may impact clinical outcomes.

This study seeks to improve mental health app engagement through piloting digital phenotyping based recommendations. We hypothesize that this recommendation approach will lead to greater app engagement as compared to a control condition of participant self selection of apps. As a secondary outcome, we explore different metrics of engagement and how different measures of engagement may inform different clinical outcomes related to app use. We hypothesize that subjective measures like self reported engagement and D-WAI will better correlate to clinical outcomes as compared to objective measures of app use measured from screen-time logs.

METHODS

Study Design

In this five week study, the first week was observational and used to gather digital phenotyping data on all participants. After this first week of data collection, all participants were randomized to receive app recommendations based on their digital phenotyping data or to select an app without any assistance or data. Over the next four weeks, participants utilized their designated app and completed pre/post questionnaires via online study visits.

Participants

All subjects were recruited from ResearchMatch. Inclusion criteria included being above 18 years of age, proficient in English, able to sign an informed consent form through an online process, having a PCP or psychiatrist, owning an Apple or Android smartphone, and scoring higher than 5 on the General Anxiety Questionnaire (GAD-7) at the initial visit. No study compensation or payment was offered for pataking.

Materials

All participants utilized two mobile health apps throughout the study. The first app that every participant used was mindLAMP, an app developed by the Division at Digital Psychiatry at BIDMC [24]. In this study, mindLAMP served solely as a digital phenotyping data collection tool. The second app varied between participants and served as an intervention tool. Participants downloaded one of four intervention apps: UCLA Mindful, How We Feel, Insomnia Coach, or Nike Training Club.

mindLAMP

mindLAMP is a digital phenotyping app developed at BIDMC [24, 25]. mindLAMP has a customizable interface with 5 main sections: Feed, Learn, Assess, Manage, and Portal. While it can be customized to offer both interventional and data capacities, this study only utilized its data collection capacity including custom surveys and sensors (GPS, accelerometer, screen use metrics).

Figure 1: *Assess Tab on mindLAMP*

Interventional Mental Health Apps

All participants were designated to engage with one of four mobile health apps for four weeks: UCLA Mindful, How We Feel, Insomnia Coach, Nike Training Club. UCLA Mindful offers guided meditations for users [26]. How We Feel is a mood tracking app that offers a range of emotions for users to choose while also tracking aspects of their physical health such as sleep and exercise [27].

Insomnia Coach guides users with their sleep through CBT and offers weekly training with sleep coach, tips, log, and diary [28]. Nike Training Club Fitness offers home workouts to healthy recipes [29]. All apps were found through the Mobile Health Index and Navigation Database on MINDapps.org developed by the Division of Digital Psychiatry [30]. For this study, we created a new filter on MINDapps to display the app(s) when participants were recommended or selected an app (see Figure 2 below). The selection of apps to include was based on feedback from patients in our clinic, advisory board, volunteer MINDapps.org app raters, and our prior research on app engagement [21]

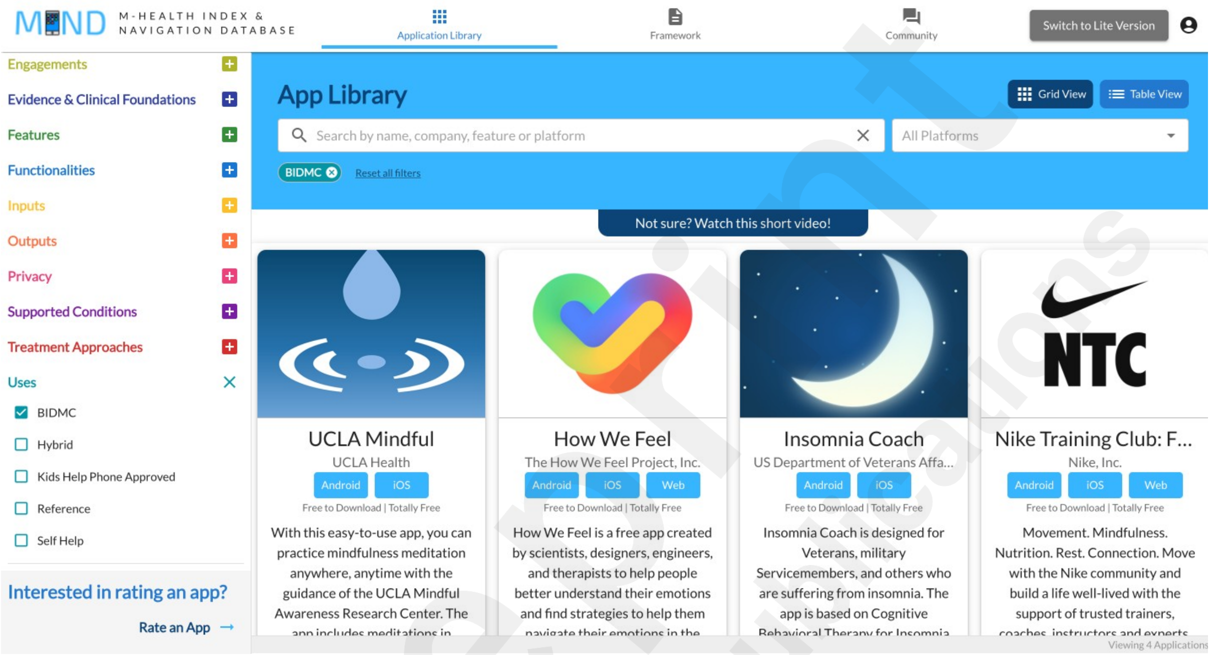


Figure 2: *Mental health apps used in study from mindapps.org database with BIDMC study filter that identified apps selected for this study.*

Data Collection

Both active and passive data features were collected as a part of this study. Passive data collection was continuously collected through the mindLAMP app for the duration of the study. Passive data specifically refers to: GPS, accelerometer, phone use (eg. screen on/off, phone on/off), and step count data. Active data is categorized as survey responses and was collected at different time points. We also collected objective data on the use of each app which is reported accessible via the participant taking a screenshot of their screen time page in their settings app and not possible to automatically capture with digital phenotyping across Apple and Android devices (See Figure 3 below). Participants were asked to take screenshots of their screen time page at the final study visit as part of the digital data collection.

Figure 3: *Screentime page in settings app. On the left shows overall screen time and on the right shows app specific screen time (How We Feel).*

Active Data (Surveys)

This study included a total of 11 surveys completed at different time points through redcap and mindLAMP. On study visit days (3 times total), participants completed a battery of standardized neuropsychiatric tests on symptoms and cognition to establish baseline, interim, and evaluation scores. The psychiatric scales consisted of: The General Anxiety Questionnaire 7 (GAD-7), Insomnia Index Scale (ISI) + Question about time of sleep onset/offset, Social Interaction Anxiety Scale, UCLA Loneliness Scale, The Flourishing Scale, and the Perceived Stress Scale (PSS). During the intake appointment, researchers completed the Clinical Global Impression Scale to evaluate participant's illness severity. Four additional surveys were administered throughout the study: The Daily Survey, System Usability Scale (SUS), Digital Working Alliance Scale (D-WAI), and the Final Engagement Survey. The daily survey was developed by the Digital Psychiatry research team and consists of 6 questions to briefly assess daily activity mental health status (See Appendix A for the full survey). Participants took the daily survey twice per day between visit 1 and 2. Between visits 2 and 3, participants reduced daily survey completion to three times per week. The System Usability Scale and The Digital Working Alliance Inventory (D-WAI) were completed during visit 1, 2, and 3. They are standardized scales used to assess app usability/perceived satisfaction and therapeutic alliance in smartphone-based interventions, respectively [23, 31]. The final engagement survey was also developed by the Digital Psychiatry team (See Appendix A for the full survey) and completed on the final day of the study to understand participants' perception of their engagement with the interventional mental health app downloaded during visit 2.

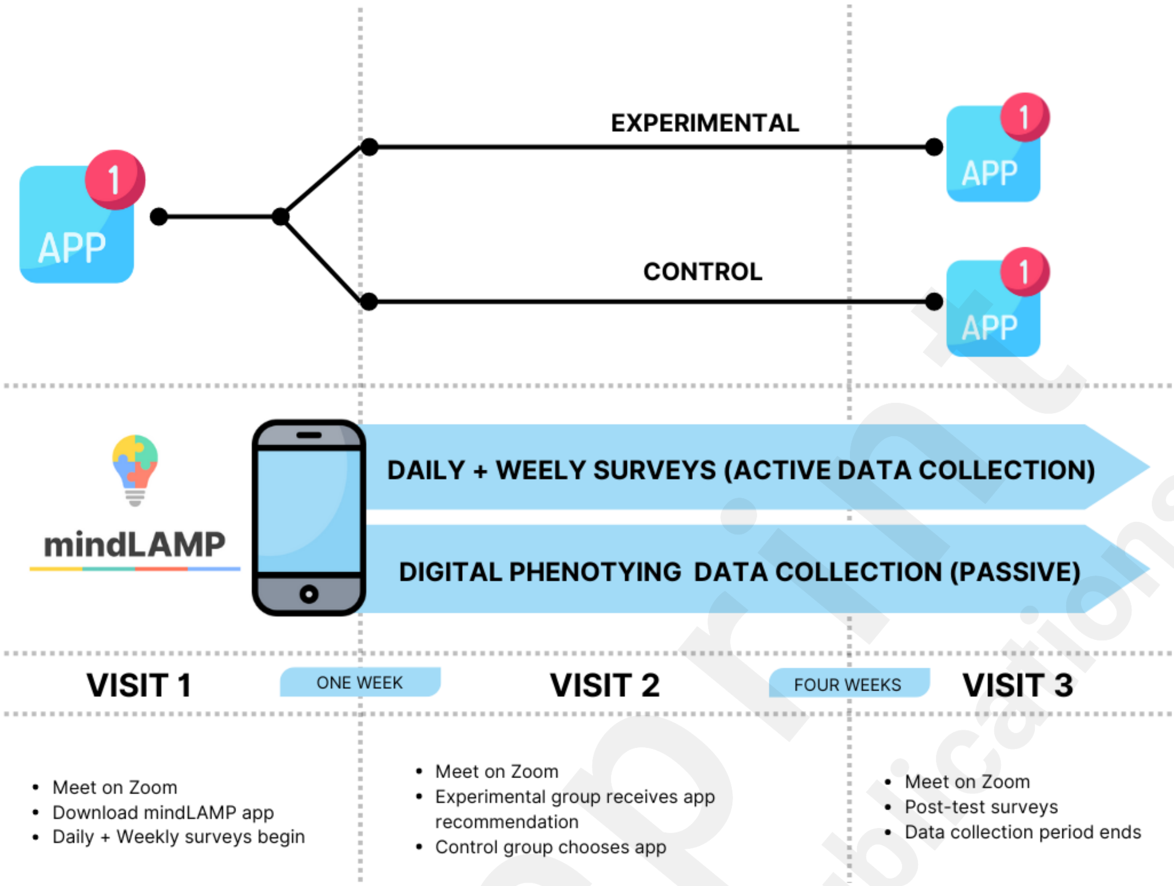


Figure 4: *Study Flow and Data Collection*

Group 1: Precision App Recommendation (Experimental)

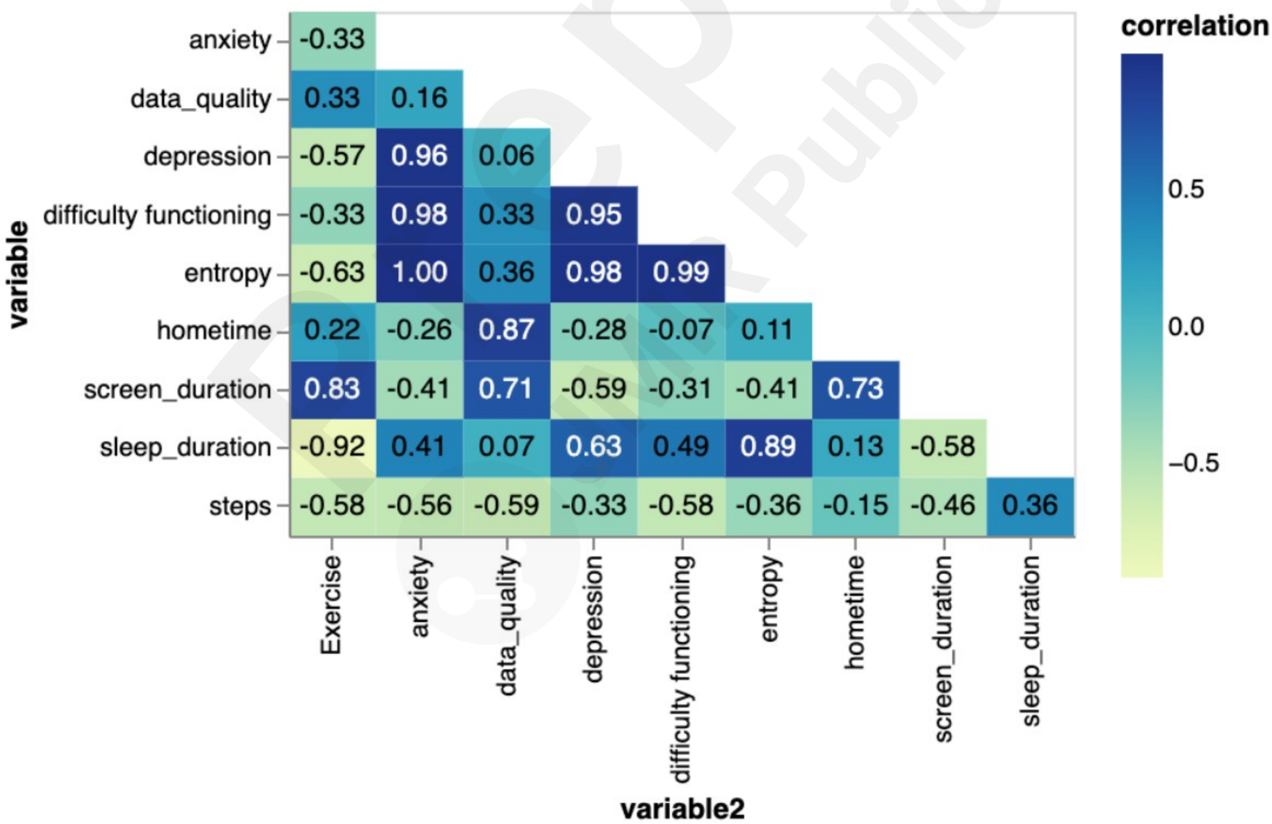
After a week of capturing digital phenotyping data about the participant, the digital navigator reviewed visualizations of that data and shared it back with the participant. The data visualizations are shown in Figure 5 and consist of a radar plot (see figure 5.1) and correlation matrix (5.2) of their data from the previous week. To recommend an app, the digital navigator assessed how mental health correlated with functioning. They selected mental health targets that featured elevated correlations for impaired functioning and persistence of this relationship over the week, and ultimately recommended an app that targeted the identified symptoms.

Figure 5.1
Radar Plot

Note: The radar plot shows the participant's score on The General Anxiety Questionnaire 7 (GAD-7), Insomnia Index Scale (ISI) + Question about time of sleep onset/offset, Social Interaction Anxiety Scale, UCLA Loneliness Scale, The Flourishing Scale, and the Perceived Stress Scale (PSS).

Figure 5.2

Correlation Matrix



Note: The correlation matrix displays factors from participant's active and passive data collected via mindLAMP; daily questionnaires related to anxiety, depression, sleep, exercise, and difficulty functioning are correlated with passive factors such as screen duration, entropy, and home time.

Group 2: Unguided App Selection (Control)

Participants in the control group also collected digital phenotyping for one week, but this data was not shared back with them until after the study. They were asked at one week to select one of the same four mental health apps (UCLA Mindful, How We Feel, Nike Training Club, or Insomnia Coach) without guidance from their digital phenotyping data or the digital navigator. A standardized description of each app was given to the participants in the control group prior to their choosing.

Interventional App Use (Both Groups)

After downloading one of the four mental health apps, both groups were asked to utilize that app for the remaining four weeks of the study. With the goal of capturing naturalistic engagement, participants were not given specific instruction regarding how or when to interact with the app. Researchers instructed participants to: “Engage with the app in your daily life as you see fit.” After four weeks, participants had their third and final meeting where their objective (screentime) and self-report (Final Engagement Survey) engagement data was collected.

Data Analysis Techniques

To assess if a personalized app recommendation would increase engagement effectively, we assessed correlations between self-reported engagement and objective engagement. To measure this relationship, an Ordinary Least Squares (OLS) linear regression model was implemented using the statsmodels library in Python. To assess that between participant’s engagement with their respective app and the participant’s anxiety symptoms, a series of OLS linear regressions were performed using the statsmodels library in Python.

A regression model was created for every measure of engagement, which was either objective (mean app screen time) or subjective (participant’s self reported engagement gathered from a survey). From these measures of engagement, we compared them to the change in structured surveys they took during the study. These surveys include; Generalized Anxiety Disorder-7 (GAD-7), Flourishing Scale, UCLA Loneliness Scale (UCLAQ), Social Interaction Anxiety Scale (SIAS), Insomnia Severity Index (ISI), and Perceived Stress Scale (PSS-10). The specific change in structured survey scores was calculated from appointment 2, when app use occurred, to appointment 3, the final appointment.

Qualitative Analysis:

Following the Braun and Clarke framework, a group of five research assistants initially reviewed the raw responses of the open ended questions in the Final Engagement survey [32]. They identified themes associated with use and engagement of the app and added them to a table: Notifications, Memory, Ease of Use, Content of app, etc. Each individuals’ final engagement survey was printed out and rated by at least two research assistants to ensure inter-rater reliability. They marked where they saw the theme and indicated whether it seemed positive or negative (ie: “The notifications were annoying” → Notifications → Negative). In the case of dispute, an additional research assistant contributed a rating until a consensus was identified. A spreadsheet was developed to indicate themes and positive, negative, or both associations.

This study was approved by the BIDMC IRB as Protocol 2022P001143.

RESULTS

Demographics and Groups

N = 82 adults were recruited and enrolled in the study. There were no significant differences in gender for the control and experimental groups. There were no significant differences of baseline anxiety or depressive symptoms in each group.

There was a total dropout of N = 35. A total of 22 participants dropped out of the study after the first meeting divided evenly between the treatment (N = 11) and control groups (N = 11). Of those 22, 9 left for unknown reasons, 6 lost interest, 3 left due to the time commitment, 3 for data quality reasons, and 1 participant left for a family emergency. After meeting two, 13 participants dropped out (N = 7 in control) (N = 6 in treatment). Two left for data quality reasons while the rest were unknown.

Table 1 below shows the full breakdown of the demographics for all 47 participants.

Sample Characteristics	(n=47)
Age	
Mean in years (SD)	n = 43.0 (15.8)
Sex	
Female	n = 37 (78.7%)
Male	n = 7 (14.9%)
Other	n = 3 (6.4%)
Race	
White/Caucasian	n = 42 (89.4%)
Black/African American	n = 2 (4.2%)
Multiracial or other	n = 3 (6.4%)
Education	
High School Graduate/GED	n = 1 (2.1%)
Some college	n = 10 (21.3%)
4-year college graduate	n = 20 (42.6%)
Masters degree or higher	n = 14 (29.8%)
Missing	n = 2 (4.2%)

Table 1: *Age, Sex, Race, and Education demographics of participants*

Of note, only How We Feel and UCLA Mindful had enough participants complete the study while using the app to produce meaningful results (see Figure 6).

Engagement

To determine if a personalized app recommendation would increase engagement at a population level, we plotted the mean objective engagement (mean screentime) of the control and experimental groups broken down by app.

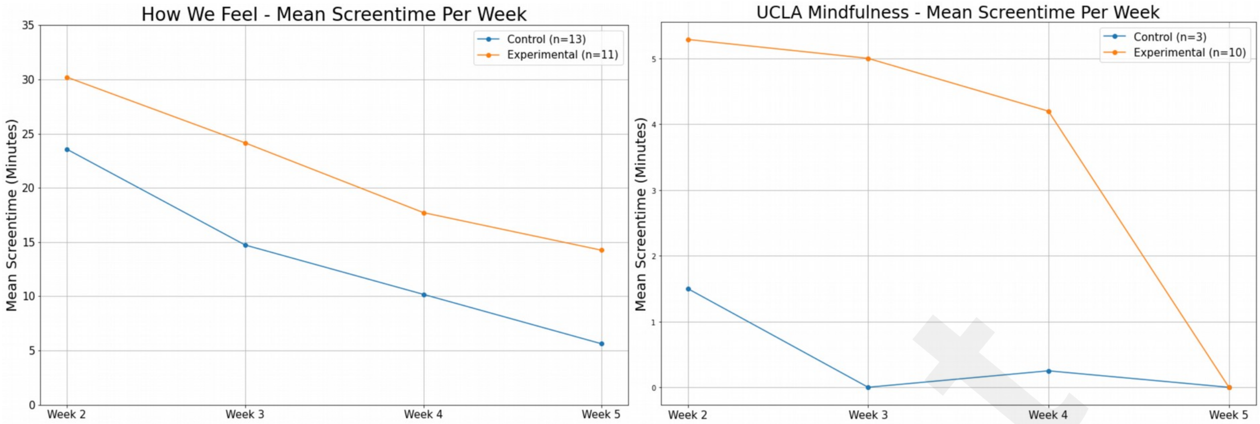


Figure 6: *Mean screentime (in Minutes) for How We Feel and UCLA Mindful apps in Control vs. Experimental groups across weeks*

How We Feel App

In both the control and experimental groups for the How We Feel (HWF) app, mean screen time was the highest during the first week of use and steadily declined throughout the four weeks (See Figure 6a). While statistically insignificant, the experimental group for HWF showed higher screen time over all.

UCLA Mindful App

In the UCLA Mindful app group, participants in the control group barely used the app after week 2 of the study while the experimental group tended to use the app more in the beginning, with a steep drop at the end of the study (See Figure 6b).

Self-Reported Engagement vs. Objective Engagement

The mean screentime values across apps could not be directly compared to self reported engagement. In order to compare all apps against each other, we utilized the MinMaxScaler function from the sklearn python library to map all mean screentime values to a 0-1 range for each app separately before combining the data for analysis.

Through OLS regression we found no significant correlation (R-squared: 0.0188, p-value: 0.3924) between self-reported engagement and scaled screentime in our pilot results. When participants were asked to rate their engagement on a scale of 1-10 (10 being the highest engagement), the control group's mean rating was 6.42 as compared to the experimental group's mean rating of 6.30.

Engagement vs. D-WAI

In addition to comparing self-reported engagement to mean screentime, we also compared both self-reported engagement and mean screentime to participant's mean score on the Digital Working Alliance Inventory (D-WAI) scale (Figure 7 below) utilizing the MinMaxScaler function noted above.

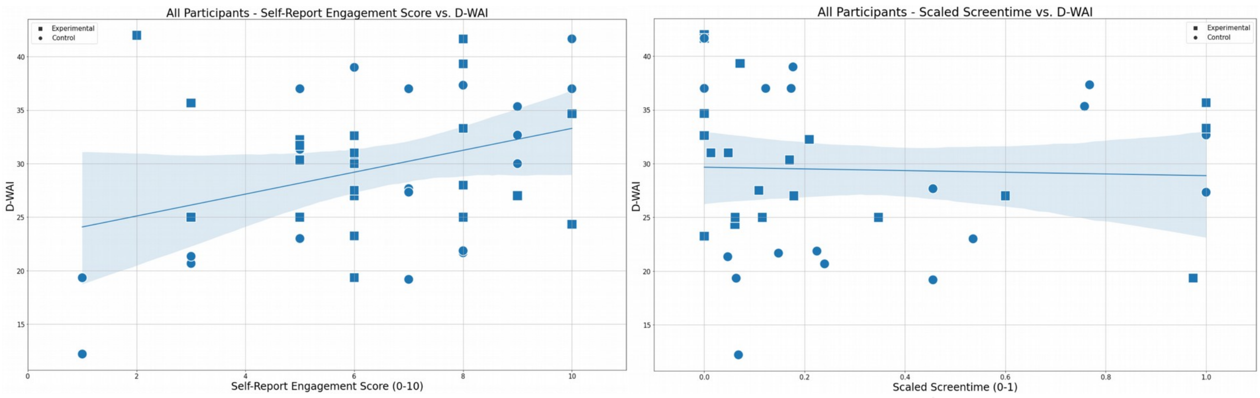


Figure 7: *Self-Reported Engagement (1-10) and Scaled Screentime (0-1) vs. Mean DWAI (0-50) for all participants and apps with regression line and 95% confidence interval.*

There was a significant positive correlation between self-reported engagement and DWAI scores (R-squared: 0.1199, p-value: 0.02132, coefficient: 1.0238), but no equivalent correlation between scaled screen time and DWAI values (R-squared: 0.0013, p-value: 0.8306, Coefficient: -0.7849). In both cases these findings were driven more by the control group than the experimental group.

Engagement vs Change in Structured Survey Scores

In addition to comparing different types of engagement, a preliminary analysis compared engagement scores to changes in various clinical surveys. We explored how engagement metrics correlated with clinical symptom score changes after one month of app use in all participants. Overall, correlations between app engagement (via any metric) and clinical changes (via any survey) were small and most results were not statistically significant. The small sample size precludes us from making any significant claims about the findings. A full table of results from our regression analyses for the How We Feel and UCLA Mindful apps can be found in Appendix B.

Qualitative Results

Following the Braun and Clarke framework, a team of 5 blinded codes reviewed Final Engagement Surveys for thematic analysis [32]. Through qualitative analysis, the team of coders initially identified 9 independent subjects linked to engagement with the interventional mental health apps. The subjects were: Perceived Usefulness, Perceived Clinical Value, Share Data Visuals, Ease of Use, Cost/Privacy, External Factors/Other Apps, Memory/Reminders, Habits/Motivation, Use as Needed/ Mental Health Status. Next, the team further categorized the 9 topics into 3 overarching themes, including: Perceived Therapeutic Utility, Access and Usability, and Behavioral Usage. See Figure 8 below.

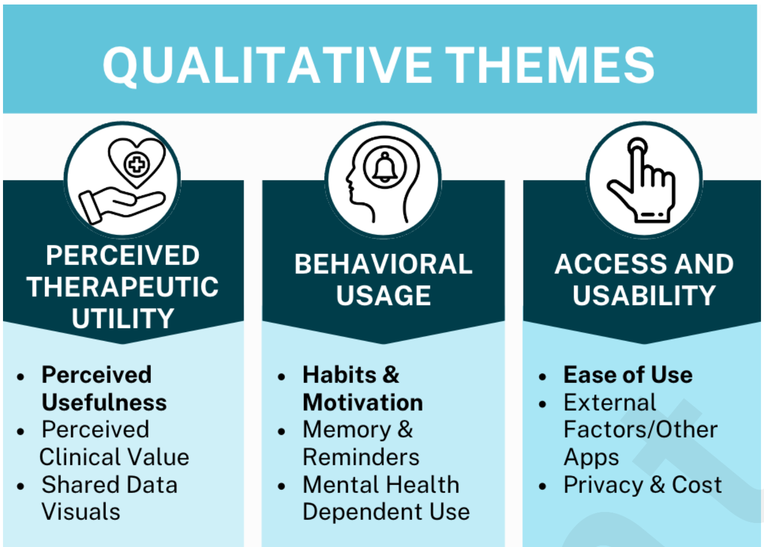


Figure 8: *Qualitative Themes*

Perceived Therapeutic Utility

Referenced 51 times in total (see figure 9), perceived therapeutic utility emerged as the most prevalent theme. Subjects frequently commented on the subjective usefulness of the application, the clinical benefit they perceived it to have, or the personal value of reviewing their data. These topics appeared to have both positive and negative influences on engagement. In some circumstances, perceived usefulness drove engagement with multiple people reporting that subjects found the mental health app "useful in moments of stress." However, in some cases it also had the opposite effect if perceived usefulness was low: "I'd have engaged more if it had more information I needed."

Behavioral Usage

Behavioral Usage was the second leading theme. Cited 44 times throughout 47 self reports (see figure 9), this theme highlighted the role of memory, habit formation, and individual motivation in *sustaining* app use. Without motivation and a habit of use, engagement was not guaranteed in the long-term: "It seemed to have useful tools…however, I just was not self-driven enough," and "Limited usage… mainly due to issues establishing a habit of entering data." Participants also had conflicting opinions on the role of reminders to use the app, with some suggesting a role in increasing engagement: "I forgot to use…If I had reminders, it would have been more useful," and others suggested against them: "App push notifications were a bit disruptive and too frequent."

Access and Usability

Access and usability were the third most prevalent themes. Referenced 33 times in total (see figure 9), subjects commented on the fundamental factors associated with access and use of the mental health app. Referenced in more than half of the reports (n=25), ease of use was the most predominant sub-category under access and usability. Without being directly prompted to report on usability, the exact phrases "easy to use" or "easy to navigate" were used in 12 independent reports. However, while access and usability emerged as importantly associated with use, it did not ensure engagement. One subject reported, "The app was intuitive to use and had a pleasing user experience, but I didn't feel particularly engaged by it." Another subject reported: "It was easy to use but meditation is not something that works well for me."

Qualitative Coding of Final Engagement Survey			
Theme	**Perceived Therapeutic Utility**	**Behavioral Usage**	**Access and Usability**
Total Amount Discussed	**51**	**44**	**33**
Postive Association	29	21	25
Negative Association	15	22	7
Neutral	7	1	1

Figure 9: *Qualitative Analysis Results (Currently making this a bar graph)*

DISCUSSION

Using digital phenotyping to guide mental health app recommendation is feasible and results in higher levels of engagement over four weeks, although assessing the clinical impact of that higher engagement remains complex due to challenges in assessing meaningful engagement. Through qualitative analysis, we were able to better understand the participants' perception of engagement driven less by perceived ease of use and perceived utility, and instead, more by habit formation. These results have implications about, first the potential of clinical app recommendation to drive engagement, second the importance of collecting both subjective and objective engagement metrics, and third the role of habit formation for sustained app use.

In our study, objective engagement was highest in both groups during the first week of interventional app use and continuously decreased throughout the study period with the largest drop from week one to week two. The high engagement across both groups at the beginning suggests that common factors may drive initial engagement, but the then differing course of engagement (see figure 6), suggests distinct factors affect sustained use. As noted in the results (see figure 6), mean weekly screen time nearly every week, the digital phenotyping app recommendation group sustained higher mean screen times, suggesting the potential of this approach.

Our study results also highlight differences between objective and subjective measures of engagement. Despite having lower objective engagement scores as measured by screen time, the control group self-reported their engagement slightly higher than the digital phenotyping group. This lack of consistency between self reported and actual use is well known [33] but rarely explored as a methodological consideration in mental health app use studies. Inconsistencies in self reported app use raise concerns about research methods relying solely on this as a measure of engagement. However, the value of either subjective or objective metrics of engagement is hard to determine, as there were few statistically significant correlations with any clinical changes after using these apps for four weeks. This negative finding could be because the use of self help apps is often not associated with large clinical changes and our sample size was underpowered to detect any small effects.

Our results on the Digital Working Alliance Inventory scores (see figure 7) showed a correlation with subjective, but not objective, screentime, measures of engagement. This result is notable as prior studies have shown that this alliance metric may be a predictor of successful clinical outcomes with self-guided mental health apps but, have not explored how it may impact engagement [20]. If alliance is related to subjective engagement as our results suggest, this raises mechanistic questions about how apps function and the need for further research exploring the dual role of subjective and objective factors in driving outcomes.

Additionally, our qualitative results suggest facets of a more nuanced picture of engagement beyond metrics like screen time or alliance. Participants agreed that perceived ease of use and perceived utility were important factors in engaging with an app, which aligns with prior research findings grounded in the Technology Acceptance Model [22]. But while these two core factors were necessary for initial engagement, results suggest that sustained engagement requires the addition of habit formation. The higher rates of engagement that we saw for the digital phenotyping recommendation group may have been driven by the feedback and digital phenotyping information that could have helped participants create routines and patterns around app use.

Limitations

Only two apps were picked for final analysis in order to produce meaningful results because the other two apps (Insomnia Coach and Nike Training Club) did not produce an adequate sample size. The lack of uptake of those two apps was related to our clinical algorithm consistently recommending participants target depression and anxiety symptoms leading to disproportionate recommendations of their respective apps. Participants in the control group most frequently chose to download the other apps as well. While we picked only 4 apps for participants to select from in this study, future studies could pick a larger number or different apps given the generalizability of this approach.

The study was not designed to assess clinical impact and instead engagement. Future studies, powered appropriately, can explore if high engagement (both subjective and objective) is actually associated with improvements in depression or anxiety. While this study offered no compensation in any form, paying participants to attend study visits would likely have increased engagement but also confounded results.

Conclusion

Digital phenotyping app recommendation is feasible and may increase rates of engagement. However, such models need to be carefully assessed before use in larger-scale studies as they may bias recommendations toward a subset of apps. Assessing the mechanism of how this approach increases engagement, whether through Digital Working Alliance or habit formation, can help advance the use of digital phenotyping for app recommendation.

References

1. Nwosu A, Boardman S, Husain MM, Doraiswamy PM. Digital therapeutics for mental health: is attrition the Achilles heel?. Frontiers in Psychiatry. 2022 Aug 5;13:900615.
2. Woolley MG, Klimczak KS, Davis CH, Levin ME. Predictors of adherence to a publicly available self-guided digital mental health intervention. Cognitive Behaviour Therapy. 2024

Apr 14:1-5.

3. Balaskas A, Schueller SM, Cox AL, Doherty G. Understanding users' perspectives on mobile apps for anxiety management. Frontiers in Digital Health. 2022:178.
4. Forbes A, Keleher MR, Venditto M, DiBiasi F. Assessing Patient Adherence to and Engagement With Digital Interventions for Depression in Clinical Trials: Systematic Literature Review. Journal of Medical Internet Research. 2023 Aug 11;25:e43727.
5. Baumel, A., Muench, F., Edan, S., & Kane, J. M. (2019). Objective user engagement with mental health apps: systematic search and panel-based usage analysis. *Journal of medical Internet research*, *21*(9), e14567.
6. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. BMJ Ment Health. 2018 Aug 1;21(3):116-9
7. Torous J, Staples P, Slaters L, Adams J, Sandoval L, Onnela JP, Keshavan M. Characterizing smartphone engagement for schizophrenia: results of a naturalist mobile health study. Clinical schizophrenia & related psychoses. 2017 Aug 4.
8. Jaworski BK, Ramsey KM, Taylor K, Heinz AJ, Senti S, Mackintosh MA, Rosen CS, Owen JE, Pietrzak RH. Mental health apps and US military veterans: Perceived importance and utilization of the National Center for Posttraumatic Stress Disorder app portfolio. Psychological Services. 2023 Nov 2.
9. Torous J, Roberts LW. Needed innovation in digital health and smartphone applications for mental health: transparency and trust. JAMA Psychiatry. 2017:74(5):437-438.
10. Chen, K., Lane, E., Burns, J., Macrynikola, N., Chang, S., & Torous, J. (2024). The Digital Navigator: Standardizing Human Technology Support in App-Integrated Clinical Care. *Telemedicine and e-Health*.
11. Posselt J, Baumann E, Dierks ML. A qualitative interview study of patients' attitudes towards and intention to use digital interventions for depressive disorders on prescription. Frontiers in Digital Health. 2024 Feb 5;6:1275569.
12. Jongeneel A, Delespaul P, Tromp N, Scheffers D, van der Vleugel B, de Bont P, Kikkert M, Croes CF, Staring AB, Riper H, van der Gaag M. Effects on voice hearing distress and social functioning of unguided application of a smartphone app—A randomized controlled trial. Internet Interventions. 2024 Jan 26:100717.
13. Khan W, Jebanesan B, Ahmed S, Trimmer C, Agic B, Safa F, Ashraf A, Tuck A, Kavic K, Wadhawan S, Abbott M. Stakeholders' views and opinions on existing guidelines on "How to Choose Mental Health Apps". Frontiers in Public Health. 2023;11.
14. Rawnsley C, Stasiak K. Unlocking the Digital Toolbox—A Mixed Methods Survey of New Zealand Mental Health Clinicians' Knowledge, Use and Attitudes Towards Digital Mental Health Interventions. Journal of Technology in Behavioral Science. 2024 Mar 13:1-1.
15. McGee-Vincent P, Mackintosh M, Jamison AL, Juhasz K, Becket-Davenport C, Bosch J, Avery TJ, Glamb L, Hampole S. Training Staff Across the Veterans Affairs Health Care System to Use Mobile Mental Health Apps: A National Quality Improvement Project. JMIR Ment Health 2023;10:e41773
16. Suggs BG, Sanderfer Stull M, Baker SR, Erwin KT, Savinsky DM. A Tide of Technical Trends: Technology Competence Among Licensed Counselors. Journal of Technology in

Counselor Education and Supervision. 2022;2(1):2.

17. Mordecai D, Histon T, Neuwirth E, Heisler WS, Kraft A, Bang Y, Franchino K, Taillac C, Nixon JP. How Kaiser Permanente created a mental health and wellness digital ecosystem. NEJM Catalyst Innovations in Care Delivery. 2021 Jan 1;2(1).

18. Hoffman L, Benedetto E, Huang H, Grossman E, Kaluma D, Mann Z, Torous J. Augmenting mental health in primary care: a 1-year study of deploying smartphone apps in a multi-site primary care/behavioral health integration program. Frontiers in psychiatry. 2019 Feb 28;10:418996.

19. Mohr DC, Schueller SM, Tomasino KN, Kaiser SM, Alam N, Karr C, Vergara JL, Gray EL, Kwasny MJ, Lattie EG. Comparison of the Effects of Coaching and Receipt of App Recommendations on Depression, Anxiety, and Engagement in the IntelliCare Platform: Factorial Randomized Controlled Trial. J Med Internet Res 2019;21(8):e13609

20. Lattie EG, Cohen KA, Hersch E, Williams KD, Kruzan KP, MacIver C, Hermes J, Maddi K, Kwasny M, Mohr DC. Uptake and effectiveness of a self-guided mobile app platform for college student mental health. Internet Interventions. 2022 Mar 1;27:100493.

21. Kopka M, Camacho E, Kwon S, Torous J. Exploring how informed mental health app selection may impact user engagement and satisfaction. PLOS Digital Health. 2023 Mar 29;2(3):e0000219.

22. Currey D, Torous J. Digital phenotyping data to predict symptom improvement and mental health app personalization in college students: prospective validation of a predictive model. Journal of Medical Internet Research. 2023 Feb 9;25:e39258.

23. Goldberg, S. B., Baldwin, S. A., Riordan, K. M., Torous, J., Dahl, C. J., Davidson, R. J., & Hirshberg, M. J. (2022). Alliance with an unguided smartphone app: validation of the digital working alliance inventory. *Assessment*, *29*(6), 1331-1345.

24. Torous, J., Wisniewski, H., Bird, B., Carpenter, E., David, G., Elejalde, E., ... & Keshavan, M. (2019). Creating a digital health smartphone app and digital phenotyping platform for mental health and diverse healthcare needs: an interdisciplinary and collaborative approach. *Journal of Technology in Behavioral Science*, *4*, 73-85

25. Vaidyam, A., Halamka, J., & Torous, J. (2022). Enabling research and clinical use of patient-generated health data (the mindLAMP Platform): digital phenotyping study. *JMIR mHealth and uHealth*, *10*(1), e30557.

26. UCLA Health. (2019, May 23). *UCLA mindful*. App Store. https://apps.apple.com/us/app/ucla-mindful/id1459128935)

27. The How We Feel Project, Inc. (2022, April 18). *How we feel*. App Store. https://apps.apple.com/us/app/how-we-feel/id1562706384

28. US Department of Veterans Affairs (VA). (2018, February 6). *Insomnia coach*. App Store. https://apps.apple.com/ca/app/insomnia-coach/id1341944736

29. Nike, I. (2009, January 15). *Nike Training Club: Wellness*. App Store. https://apps.apple.com/us/app/nike-training-club-wellness/id301521403.

30. Lagan, S., Aquino, P., Emerson, M. R., Fortuna, K., Walker, R., & Torous, J. (2020). Actionable health app evaluation: translating expert frameworks into objective metrics. npj Digital Medicine, 3, 1–8

31. Lewis, J. R. (2018). The system usability scale: past, present, and future. *International Journal of Human–Computer Interaction*, *34*(7), 577-590.
32. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol (2006) 3:77–101. doi: 10.1191/1478088706qp063oa.
33. Griffioen, N., van Rooij, M., Lichtwarck-Aschoff, A., & Granic, I. (2020). Toward improved methods in social media research.