<u>Original Paper</u>

Digital Phenotyping Data to Predict Symptom Improvement and Mental Health App Personalization in College Students: Prospective Validation of a Predictive Model

Danielle Currey[1,2], BS; John Torous[1], MD

[1]Department of Psychiatry, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, United States
[2]Case Western Reserve University School of Medicine, Cleveland, OH, United States

Corresponding Author:
John Torous, MD
Department of Psychiatry
Beth Israel Deaconess Medical Center
Harvard Medical School
330 Brookline Ave
Boston, MA, 02446
United States
Phone: 1 6176676700
Email: jtorous@gmail.com

Abstract

Background: Mental health apps offer a transformative means to increase access to scalable evidence-based care for college students. Yet low rates of engagement currently preclude the effectiveness of these apps. One promising solution is to make these apps more responsive and personalized through digital phenotyping methods able to predict symptoms and offer tailored interventions.

Objective: Following our protocol and using the exact model shared in that paper, our primary aim in this study is to assess the prospective validity of mental health symptom prediction using the mindLAMP app through a replication study. We also explored secondary aims around app intervention personalization and correlations of engagement with the Technology Acceptance Model (TAM) and Digital Working Alliance Inventory scale in the context of automating the study.

Methods: The study was 28 days in duration and followed the published protocol, with participants collecting digital phenotyping data and being offered optional scheduled and algorithm-recommended app interventions. Study compensation was tied to the completion of weekly surveys and was not otherwise tied to engagement or use of the app.

Results: The data from 67 participants were used in this analysis. The area under the curve values for the symptom prediction model ranged from 0.58 for the UCLA Loneliness Scale to 0.71 for the Patient Health Questionnaire-9. Engagement with the scheduled app interventions was high, with a study mean of 73%, but few participants engaged with the optional recommended interventions. The perceived utility of the app in the TAM was higher (P=.01) among those completing at least one recommended intervention.

Conclusions: Our results suggest how digital phenotyping methods can be used to create generalizable models that may help create more personalized and engaging mental health apps. Automating studies is feasible, and our results suggest targets to increase engagement in future studies.

International Registered Report Identifier (IRRID): RR2-10.2196/37954

XSL•FO
RenderX

Introduction

Digital mental health solutions, especially smartphone apps, are recognized as a scalable means to increase access to care. With the rising crisis around youth mental health [1], compounded by the pandemic and the chronic lack of adequate services on college campuses [2], college mental health is a prime application for such apps. Yet, the impact of these apps to date has been minimal, not because apps are ineffective but rather because engagement is often low [3,4].

This study explores methods for improving engagement, with a focus on personalization. While many apps can deliver evidence-based content, there is evidence that students want apps to be more tailored to their personalized needs [5]. This personalization requires first predicting individual symptom changes and second responding with appropriate interventions. Focusing on the first step, digital phenotyping can advance prediction by using smartphone sensors to derive behavioral features (eg, sleep, steps) that can be incorporated into models. While prior research on digital phenotyping has proposed models for college student mental health [6], these models have never been prospectively validated. Following our published protocol [7], in this study, we prospectively validated a digital phenotyping symptom prediction model for college students. To our knowledge, this is the first digital phenotyping algorithm to be prospectively validated.

Our study app, mindLAMP [8,9], provides a useful platform for this research, as it offers both digital phenotyping as well as a suite of cognitive therapy–based exercises and skills that participants can access on demand or as scheduled. This enables mindLAMP to be a responsive app and use digital phenotyping–derived symptom prediction to help recommend individual app activities for a student. While mindLAMP did not offer just-in-time adaptive interventions in this study, we piloted the feasibility of responsive interventions using passive data features as a secondary aim. Such an exploration of feasibility is important, as few other apps have used digital phenotyping data in predictive models. A recent review of just-in-time adaptive intervention apps found that 71% of these apps use only self-report (ie, not digital phenotyping data), and considering all apps, only 3.6% of all measurements involved sensor and device analytics [10].

This same report that noted how little is known about the role of sensor data in just-in-time adaptive interventions also discussed how there is a lack of research on the theoretical basis and mechanism of action for how these apps drive engagement or outcomes. Theoretical models like the Technology Acceptance Model (TAM) [11] can help elucidate factors associated with app engagement but, to date, have been rarely used in this research. Specifically, when using methods like digital phenotyping it remains unknown if these efforts toward increased personalization increase positive attitudes about the app, make the app more useful, or both. Other factors beyond those explored in the TAM, such as alliance and connection to the app (measured with the Digital Working Alliance Inventory [DWAI] [12]) are also important to explore.

While in prior research studies we have shown that students can engage with therapeutic activities within mindLAMP [13], our use of the app activities in this study was to assess the recommender model. Building off our prior work with the mindLAMP app [8,9], in this study, we sought to validate our digital phenotyping symptom prediction model as a primary goal while secondarily assessing the feasibility of an app activity recommender model and the correlations with elements of the TAM and DWAI.

Methods

Ethics Approval

This study was approved by the Beth Israel Deaconess Medical Center institutional review board (protocol 2020P000310).

Study

The protocol for this paper was published in *JMIR Research Protocols* [7]. This study used the open source mindLAMP app developed by the Digital Psychiatry lab at Beth Israel Deaconess Medical Center to collect survey and sensor data [8,9]. Briefly, the study recruited undergraduate participants via Reddit to complete a screener survey that was required to participate. The following inclusion criteria were used: students must be 18 years or older, score 14 or higher on the Perceived Stress Scale (PSS) [14], be enrolled as an undergraduate for the duration of the study, own a smartphone able to run mindLAMP, be able to sign informed consent, and pass a run-in period (outlined below). A link to the REDCap screener survey was posted on 73 different college Reddit web pages. Participants that were screened completed an informed consent quiz, which then automatically generated a mindLAMP account log-in.

Participants entered a 3-day run-in period, and if they completed the requirements (daily surveys and GPS passive data coverage checks), they were moved into the 28-day enrollment period. Passive data coverage was estimated as the percent of 10-minute intervals that had at least one GPS data point collected. If coverage was greater than 20%, participants were able to continue into the enrollment period. It is important to note that some level of missingness is expected in digital phenotyping work. We did not perform any imputation of passive data in our analysis.

During the enrollment phase of the study, participants had a feed with activities scheduled each day (eg, mindfulness or gratitude journaling) and completed longer weekly surveys (which included the Patient Health Questionnaire-9 [PHQ-9] [15], Generalized Anxiety Disorder-7 [GAD-7] [16,17], PSS [14], UCLA Loneliness Survey [18], Pittsburgh Sleep Quality Index [PSQI] [19], DWAI [12], and TAM-related questions [11]) each week. Participants were compensated for completing weekly surveys at weeks 1, 3, and 4 (US $50 in total) but not for engaging with any interventions of the coverage of digital phenotyping captured through their smartphones. However, participants were warned via email if they did not complete any activities for 3 days and were discontinued from the study if they still had not completed any activities after 5 consecutive days. In addition to these activities, one-third of participants received emails from a "research assistant" (emails were

automated but signed with a researcher's name) encouraging them to complete an additional activity based on their data. Another one-third of participants were sent an activity suggestion from the study bot "Marvin." The final one-third of the participants did not receive any additional activity suggestions. As explained in the protocol [7], participants were sequentially assigned to each of the three groups.

This study is important because it was a successful iteration of a fully automated study. All aspects of the study, including enrollment, data quality checks, payment, and activity suggestions and scheduling, were automated by Python study workers. For further details on the study, please refer to the protocol [7]. We would like to note that there were no deviations from the protocol. The code for the automated workers is open source and available on GitHub [20].

Enrollment

A total of 67 participants were used for this analysis. The number of participants that completed each phase is shown in Figure 1. A total of 636 participants filled out the initial screening survey, 481 of whom met the study requirements outlined above; 170 of these participants completed and passed a required informed consent quiz designed to ensure participants understood the expectations of the study. Of these, 154 created mindLAMP accounts and entered the run-in period. A total of 46 of these participants did not make it through the run-in because they either did not complete all the daily surveys (bad active data) or did not meet the passive data coverage requirement (bad passive data). Some participants had both "bad" active and passive data and were counted under "bad

active." Of the 108 participants that entered the enrollment period of the study, 34 were discontinued after not completing any activities in the app for 5 consecutive days. A total of 74 participants completed the study.

Some participants were excluded from the analysis. First, the initial screener survey had an error where 3 participants who reported being unable to meet if needed were allowed to complete the informed consent. These participants are shown in Figure 1 under "unable to meet if needed" and were excluded from the analysis (the numbers do not sum to the total for screen-fail as some people were excluded for multiple reasons). Second, after retrospectively reviewing the REDCap screener survey entries, it seems that some participants filled out the survey multiple times, changing their responses to be included in the study. One participant changed their status from "graduate" to "undergraduate," and 4 other participants changed their responses to the PSS survey to increase their scores by double or more. While stress does fluctuate over time, such a large change over only about 10 minutes is unlikely. These 5 participants are not shown in Figure 1 and were excluded from the analysis.

Finally, 7 of the 74 participants that completed the study previously participated in the College Study (V1 or V2), so they were excluded from the analysis to ensure the sample used for model testing was completely distinct from the training and testing sets. The demographics of participants used in the analysis are outlined in Table 1. Of the 67 participants used for analysis, 52 used iOS phones and 15 used Android phones. Participants had a mean age of 20 (SD 2) years.

Figure 1. Flowchart showing counts of participants at different stages in the study. IFC: informed consent; PSS: Perceived Stress Scale.
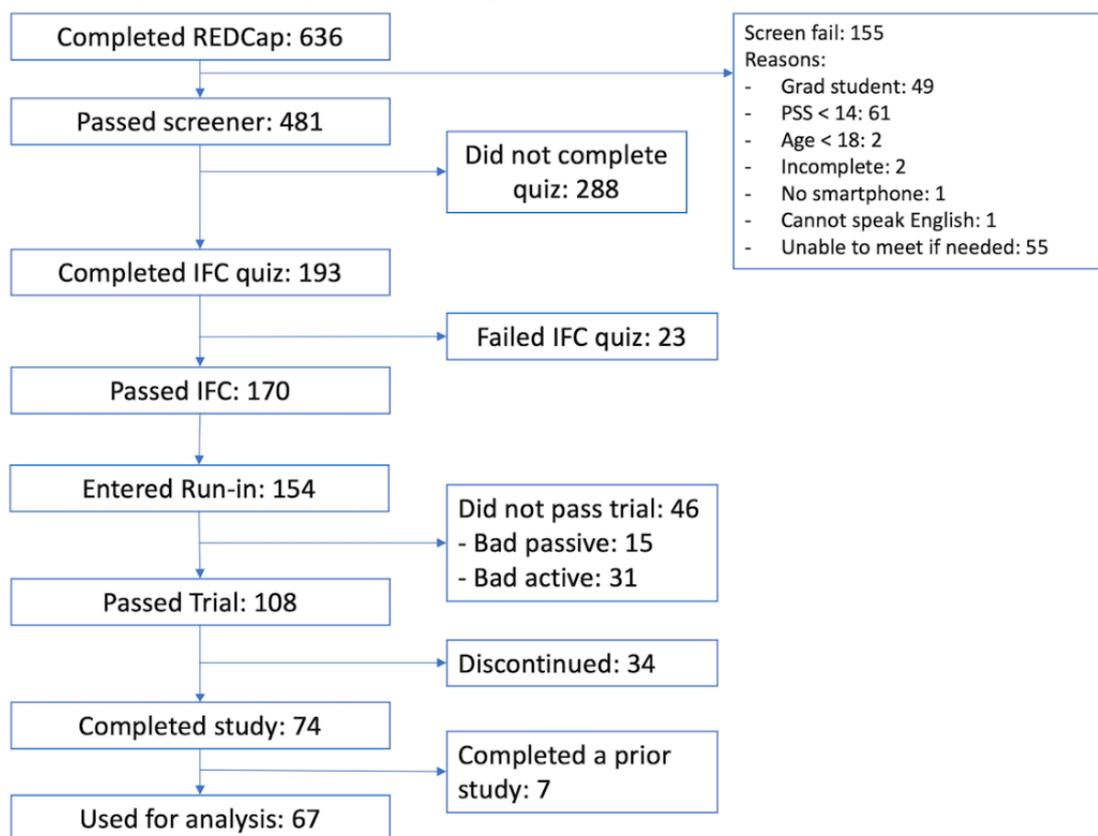
Table 1. Demographic information.

Race	Female, n	Male, n	Nonbinary, n	Total, n
Asian	11	6	1	19[a]
African American	2	0	0	2
Latinx	7	1	0	8
White	24	8	6	38
Total	44	15	7	67

[a]One Asian student marked their gender as "prefer not to say."

Prospective Model

The model reported in the protocol [7] was prospectively tested on this data set. The 67 participants used for testing have not previously participated in a College Study and were completely distinct from the testing and validation training sets.

Groupwise Analysis

As a secondary outcome, the Python scipy.stats [21] was used to perform ANOVA tests on each feature to determine if differences existed between the groups receiving suggestions from a digital navigator or a bot, or receiving no suggestions. We also used the scipy.stats module to perform *t* tests comparing the group that completed the suggested activities to the group that did not. Here, we examined the TAM [11] and DWAI [12,22] questions from the weekly survey to investigate the relationships between attitudes toward the app and behavior.

We did not aim to validate the TAM but rather to explore questions related to engagement. *P* values were corrected using the Hochberg method via the statsmodels.stats.multitest.multipletests module in Python [23].

Results

The primary goal of this study was to prospectively validate a model to predict whether a participant would improve over the course of the study given the average of each of their passive and active data features. The change in score by participant (scaled by the number of points in the survey) is shown in Figure 2. Except for the GAD-7, which showed a slight average increase in the score (0.0025), participants' scores on average decreased. The mean and SD for the weekly survey scores across students are listed in Table 2.

Figure 2. Violin plots showing the change in scores for each survey scaled by the total possible score (PHQ-9: 27, GAD-7: 21, PSS: 40, UCLA: 60, PSQI: 15). Score differences are plotted as starting score – ending score, so that positive values indicate improvement. GAD-7: Generalized Anxiety Disorder-7; PHQ-9: Patient Health Questionnaire-9; PSQI: Pittsburgh Sleep Quality Index; PSS: Perceived Stress Scale; UCLA: UCLA Loneliness Survey.
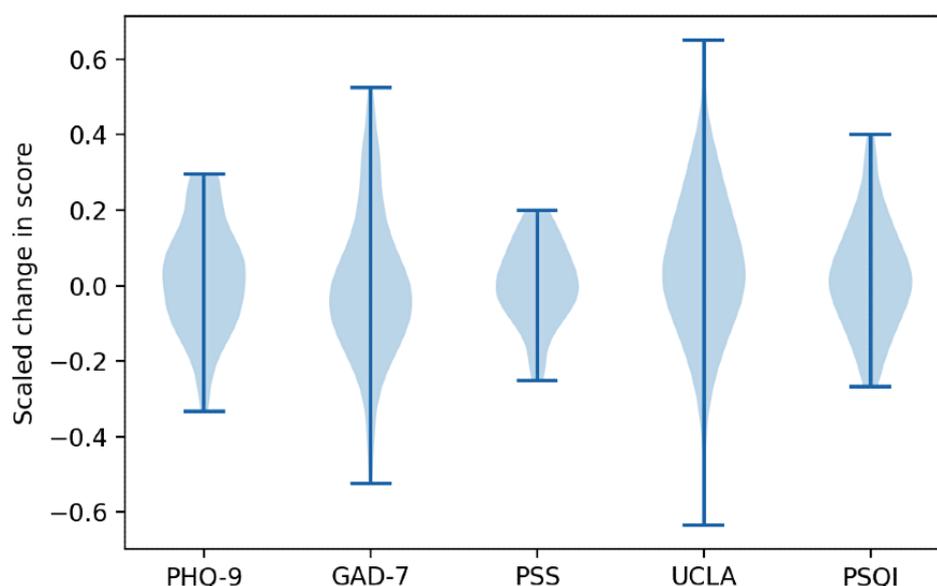
Table 2. Mean and SD of weekly survey scores across students.

Survey	Scores, mean (SD)
Patient Health Questionnaire-9	7.55 (4.45)
Generalized Anxiety Disorder-7	6.49 (3.90)
Perceived Stress Scale	18.04 (6.33)
UCLA Loneliness Survey	15.56 (15.05)
Pittsburgh Sleep Quality Index	5.30 (2.77)

Area under the curve (AUC) values range from 0.58 (UCLA Loneliness) to 0.71 (PHQ-9). Specifically, the AUC values were 0.71 for PHQ-9, 0.60 for GAD-7, 0.68 for PSS, 0.58 for UCLA Loneliness, and 0.60 for PSQI.

There were no significant differences in the survey features ($P>.05$), the number of activities completed ($P>.99$), or the percentage of module activities completed between the three groups ($P>.99$). Moreover, the changes in the PHQ-9 and GAD-7 scores were not significantly different between the three groups ($P=.42$ and $P=.72$, respectively). Despite the relatively high completion of scheduled module activities (73% on average), few participants completed the optional activities suggested by the recommendation algorithm. Over half (24/41, 59%) of the participants who received suggestions completed at least one activity. Two-sided t tests were performed to compare the DWAI and TAM question scores between the group that completed at least one activity and the group that did not complete any of the optional activities. These P values can be found in Table 3.

The differences between the groups were in the questions about the perceived usefulness of the app. There were no significant correlations between the magnitude of the PHQ-9 or GAD-7 scores' improvement and the DWAI score, the number of activities completed, or the percentage of assigned activities completed. Finally, there were no significant correlations between the average PHQ-9 and GAD-7 scores and DWAI or TAM scores.

Table 3. P values for the t test comparison between participants that did and did not complete at least one suggested activity.

Question	Component of TAM[a] model	P value
DWAI[b]		
Total score	N/A[c]	.77
I agree that the tasks within the app are important for my goals.	Attitude toward using	.07
I believe the app tasks will help me to address my problems.	Attitude toward using	.07
I trust the app to guide me toward my personal goals.	Attitude toward using	.24
The app encourages me to accomplish tasks and make progress.	Attitude toward using	.31
The app is easy to use and operate.	Perceived ease of use	.48
The app supports me to overcome challenges.	Perceived usefulness	.39
TAM		
Total score	N/A	.07
I want to use the app daily.	Behavioral intention to use	.27
I would want to use it after the study ends.	Behavioral intention to use	.18
The app allows me to easily manage my mental health.	Perceived usefulness	.31
The app makes me better informed of my mental health.	Perceived usefulness	*.01*[d]
The app provides me with valuable information or skills.	Perceived usefulness	*.01*

[a]TAM: Technology Acceptance Model.

[b]DWAI: Digital Working Alliance Inventory.

[c]N/A: not applicable.

[d]Italicized P values indicate values less than .05.

Discussion

Principal Findings

The primary outcome, validation of the symptom prediction model, demonstrated overall success with an AUC of 0.71 for change in depression symptoms measured by the PHQ-9, which is similar to the results from prior studies referenced in the protocol [7]. This prospective validation indicates that such models may be able to generalize across samples and thus be applicable to a broad range of college students.

Our results also explored engagement with apps as secondary outcomes. Overall engagement with assigned tasks in the app was 73%. Participants who completed at least one of the optional recommended activities scored differently on certain TAM questions. In particular, the usefulness questions around the belief that the app provides some helpful/valuable information differed, indicating that these types of attitudes toward the utility of the app may be necessary for participants to engage. However, our data set is small and not powered for these outcomes, so further work is needed to explore questions around which participants are best suited to benefit from using digital mental health apps.

Deploying the app recommendation algorithm demonstrated feasibility but did not in itself change engagement. This was likely due to our study not being designed or powered to change engagement but rather to replicate the prediction algorithm and demonstrate the feasibility of using it to automate recommendations.

Strengths and Limitations

Our study was limited by the sample size and the fact that all participants were college students. First, a larger sample size would provide better training and testing data, and would allow for improvements in the symptom prediction model. Moreover, the small number of participants in each engagement subgroup (about 20 per group) makes it difficult to compare them. In the future, larger sample sizes should be recruited to further investigate the difference between interacting with a digital navigator and not.

Second, although there is a high level of need in the college student population for mental health resources, using college students can also be considered a limitation of this study, as college mental health may not be representative of the broader experience for the general population. Moreover, the way that college-aged students interact with their phones may be different from the rest of the population, making phenotyping methods difficult to transfer to other groups. Future work should explore generalization across age and other demographic groups and seek a more gender-balanced sample than ours. Related work suggests that bias across races and different ethnic groups may be low [24], but this needs to be assessed in future work.

Qualitative results around app engagement with mindLAMP, as we have done in the past with college students [25], will be important to explore with new populations. More work is also necessary to validate symptom prediction models in different populations, especially those with lower digital health literacy or students with different backgrounds than those featured in our sample. Still, given the high degree of mental health needs in this population, our results can support future efforts to personalize apps toward delivering more tailored care.

Future Directions

While our activity tailoring algorithm did not drive engagement, overall engagement was high in our study. Weekly therapeutic module completion was high despite these activities not being required or compensated. We did not assess the reasons for this higher engagement, but perhaps by scheduling activities in the feed, participants felt that they should complete these daily activities, while the additional activities recommended by the activity recommendation algorithm were explicitly provided as suggestions. Adding the recommendation algorithm activity suggestions into the feed is a simple next step to assess in future studies. Moreover, since our work suggests that attitudes around app usefulness contribute to engagement, future work should also explore whether participant attitudes can be changed. If the belief that the app is helpful is the key to engagement, then focusing on changing this attitude may be the key to reaching technology-resistant participants. We also note that our results around engagement were secondary outcomes, and our analysis involved a first-pass overview with assumptions such as the underlying data being normally distributed. As the field seeks to better operationalize measures of engagement [26,27], digital phenotyping metrics like those featured in this paper may play a future role [28,29].

Our results around overall study recruitment and retention are also important for planning and powering such future studies [30]. We were able to obtain a high-quality data set, but this required recruitment of almost three times our goal due to the loss of participants through the run-in period and the requirement for a baseline level of participation. In addition, the fact that we had at least five participants providing false information to enter the study underscores the challenge of web-based recruitment the field is now growing aware of [17]. We expect this likely impacts all web-based studies and hope that, by calling attention here, others will also carefully consider who is enrolled in their digital health research.

Conclusion

Overall, this study presents evidence that a digital phenotyping symptom prediction model can prospectively generalize to a new population of college students. The success of the automated study protocol holds promise for being able to efficiently run even larger studies in the future, and the results around activity tailoring suggest areas for future improvement.

XSL•FO

RenderX

Data Availability

The code for the automated workers is open source and is available on our GitHub. Data from this study is not available given the personally identifiable nature of the information.

Conflicts of Interest

None declared.

References

1. Office of the Surgeon General. Protecting Youth Mental Health: The U.S. Surgeon General's Advisory. U.S. Department of Health & Human Services. 2021 Nov 10. URL: https://www.hhs.gov/sites/default/files/surgeon-general-youth-mental-health-advisory.pdf [accessed 2023-02-02]

2. Pedrelli P, Nyer M, Yeung A, Zulauf C, Wilens T. College students: mental health problems and treatment considerations. Acad Psychiatry 2015 Oct;39(5):503-511 [FREE Full text] [doi: 10.1007/s40596-014-0205-9] [Medline: 25142250]

3. Ng MM, Firth J, Minen M, Torous J. User engagement in mental health apps: a review of measurement, reporting, and validity. Psychiatr Serv 2019 Jul 01;70(7):538-544 [FREE Full text] [doi: 10.1176/appi.ps.201800519] [Medline: 30914003]

4. Torous J, Michalak EE, O'Brien HL. Digital health and engagement-looking behind the measures and methods. JAMA Netw Open 2020 Jul 01;3(7):e2010918 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.10918] [Medline: 32678446]

5. Lattie EG, Cohen KA, Hersch E, Williams KDA, Kruzan KP, MacIver C, et al. Uptake and effectiveness of a self-guided mobile app platform for college student mental health. Internet Interv 2022 Mar;27:100493 [FREE Full text] [doi: 10.1016/j.invent.2021.100493] [Medline: 35036331]

6. Melcher J, Hays R, Torous J. Digital phenotyping for mental health of college students: a clinical review. Evid Based Ment Health 2020 Nov;23(4):161-166. [doi: 10.1136/ebmental-2020-300180] [Medline: 32998937]

7. Currey D, Torous J. Digital phenotyping data to predict symptom improvement and app personalization: protocol for a prospective study. JMIR Res Protoc 2022 Nov 29;11(11):e37954 [FREE Full text] [doi: 10.2196/37954] [Medline: 36445745]

8. Vaidyam A, Halamka J, Torous J. Enabling research and clinical use of patient-generated health data (the mindLAMP Platform): digital phenotyping study. JMIR Mhealth Uhealth 2022 Jan 07;10(1):e30557 [FREE Full text] [doi: 10.2196/30557] [Medline: 34994710]

9. Torous J, Wisniewski H, Bird B, Carpenter E, David G, Elejalde E, et al. Creating a digital health smartphone app and digital phenotyping platform for mental health and diverse healthcare needs: an interdisciplinary and collaborative approach. J Technol Behav Sci 2019 Apr 27;4(2):73-85. [doi: 10.1007/s41347-019-00095-w]

10. Teepe GW, Da Fonseca A, Kleim B, Jacobson NC, Salamanca Sanabria A, Tudor Car L, et al. Just-in-time adaptive mechanisms of popular mobile apps for individuals with depression: systematic app search and literature review. J Med Internet Res 2021 Sep 28;23(9):e29412 [FREE Full text] [doi: 10.2196/29412] [Medline: 34309569]

11. Davis FD. A technology acceptance model for empirically testing new end-user information systems: theory and results [dissertation]. Massachusetts Institute of Technology. 1986. URL: https://dspace.mit.edu/handle/1721.1/15192 [accessed 2023-01-26]

12. Goldberg SB, Baldwin SA, Riordan KM, Torous J, Dahl CJ, Davidson RJ, et al. Alliance with an unguided smartphone app: validation of the Digital Working Alliance Inventory. Assessment 2022 Sep;29(6):1331-1345. [doi: 10.1177/10731911211015310] [Medline: 34000843]

13. Melcher J, Patel S, Scheuer L, Hays R, Torous J. Assessing engagement features in an observational study of mental health apps in college students. Psychiatry Res 2022 Apr;310:114470. [doi: 10.1016/j.psychres.2022.114470] [Medline: 35227991]

14. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]

15. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]

16. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]

17. Löwe B, Decker O, Müller S, Brähler E, Schellberg D, Herzog W, et al. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. Med Care 2008 Mar;46(3):266-274. [doi: 10.1097/MLR.0b013e318160d093] [Medline: 18388841]

18. Russell DW. UCLA Loneliness Scale (Version 3): reliability, validity, and factor structure. J Pers Assess 1996 Feb;66(1):20-40. [doi: 10.1207/s15327752jpa6601_2] [Medline: 8576833]

19. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]

20. Currey D, Hays R, Vaidyam A, Torous J. mindLAMP College Study V3 [computer software]. GitHub. 2022. URL: https://github.com/BIDMCDigitalPsychiatry/LAMP-college-study/ [accessed 2023-02-02]

XSL•FO

RenderX

21. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, SciPy 1.0 Contributors. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 2020 Mar;17(3):261-272 [FREE Full text] [doi: 10.1038/s41592-019-0686-2] [Medline: 32015543]

22. Henson P, Wisniewski H, Hollis C, Keshavan M, Torous J. Digital mental health apps and the therapeutic alliance: initial review. BJPsych Open 2019 Jan;5(1):e15 [FREE Full text] [doi: 10.1192/bjo.2018.86] [Medline: 30762511]

23. Seabold S, Perktold J. Statsmodels: econometric and statistical modeling with Python. In: Proceedings of the Python in Science Conference. 2010 Jun 1 Presented at: The Python in Science Conference; 2010; Austin, Texas. [doi: 10.25080/majora-92bf1922-011]

24. Currey D, Torous J. Increasing the value of digital phenotyping through reducing missingness: a retrospective analysis. medRxiv Preprint posted online on May 17, 2022. [doi: 10.1101/2022.05.17.22275182]

25. Melcher J, Torous J. Smartphone apps for college mental health: a concern for privacy and quality of current offerings. Psychiatr Serv 2020 Nov 01;71(11):1114-1119. [doi: 10.1176/appi.ps.202000098] [Medline: 32664822]

26. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the concept of adherence to eHealth technology: systematic review on when usage becomes adherence. J Med Internet Res 2017 Dec 06;19(12):e402 [FREE Full text] [doi: 10.2196/jmir.8578] [Medline: 29212630]

27. Kelders SM, van Zyl LE, Ludden GDS. The concept and components of engagement in different domains applied to eHealth: a systematic scoping review. Front Psychol 2020;11:926 [FREE Full text] [doi: 10.3389/fpsyg.2020.00926] [Medline: 32536888]

28. Nickels S, Edwards MD, Poole SF, Winter D, Gronsbell J, Rozenkrants B, et al. Toward a mobile platform for real-world digital measurement of depression: user-centered design, data quality, and behavioral and clinical modeling. JMIR Ment Health 2021 Aug 10;8(8):e27589 [FREE Full text] [doi: 10.2196/27589] [Medline: 34383685]

29. Hsu M, Ahern DK, Suzuki J. Digital phenotyping to enhance substance use treatment during the COVID-19 pandemic. JMIR Ment Health 2020 Oct 26;7(10):e21814 [FREE Full text] [doi: 10.2196/21814] [Medline: 33031044]

30. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]

Abbreviations

AUC: area under the curve
DWAI: Digital Working Alliance Inventory
GAD-7: Generalized Anxiety Disorder-7
PHQ-9: Patient Health Questionnaire-9
PSQI: Pittsburgh Sleep Quality Index
PSS: Perceived Stress Scale
TAM: Technology Acceptance Model

XSL•FO
RenderX