<u>Original Paper</u>

Popular Evidence-Based Commercial Mental Health Apps: Analysis of Engagement, Functionality, Aesthetics, and Information Quality

Nancy Lau[1,2,3], PhD; Alison O'Daffer[1], BA; Joyce P Yi-Frazier[1], PhD; Abby R Rosenberg[1,3,4], MD, MSc, MA

[1]Palliative Care and Resilience Lab, Center for Clinical and Translational Research, Seattle Children's Research Institute, Seattle, WA, United States

[2]Department of Psychiatry and Behavioral Sciences, University of Washington School of Medicine, Seattle, WA, United States

[3]Cambia Palliative Care Center of Excellence, University of Washington, Seattle, WA, United States

[4]Department of Pediatrics, University of Washington School of Medicine, Seattle, WA, United States

Corresponding Author:
Nancy Lau, PhD
Palliative Care and Resilience Lab
Center for Clinical and Translational Research
Seattle Children's Research Institute
1920 Terry Ave
Seattle, WA, 98101
United States
Phone: 1 2068840569
Email: nancy.lau@seattlechildrens.org

Abstract

Background: There is a robust market for mobile health (mHealth) apps focused on self-guided interventions to address a high prevalence of mental health disorders and behavioral health needs in the general population. Disseminating mental health interventions via mHealth technologies may help overcome barriers in access to care and has broad consumer appeal. However, development and testing of mental health apps in formal research settings are limited and far outpaced by everyday consumer use. In addition to prioritizing efficacy and effectiveness testing, researchers should examine and test app design elements that impact the user experience, increase engagement, and lead to sustained use over time.

Objective: The aim of this study was to evaluate the objective and subjective quality of apps that are successful across both research and consumer sectors, and the relationships between objective app quality, subjective user ratings, and evidence-based behavior change techniques. This will help inform user-centered design considerations for mHealth researchers to maximize design elements and features associated with consumer appeal, engagement, and sustainability.

Methods: We conducted a user-centered design analysis of popular consumer apps with scientific backing utilizing the well-validated Mobile Application Rating Scale (MARS). Popular consumer apps with research support were identified via a systematic search of the App Store iOS (Apple Inc) and Google Play (Google LLC) and literature review. We evaluated the quality metrics of 19 mental health apps along 4 MARS subscales, namely, Engagement, Functionality, Aesthetics, and Information Quality. MARS total and subscale scores range from 1 to 5, with higher scores representing better quality. We then extracted user ratings from app download platforms and coded apps for evidence-based treatment components. We calculated Pearson correlation coefficients to identify associations between MARS scores, App Store iOS/Google Play consumer ratings, and number of evidence-based treatment components.

Results: The mean MARS score was 3.52 (SD 0.71), consumer rating was 4.22 (SD 0.54), and number of evidence-based treatment components was 2.32 (SD 1.42). Consumer ratings were significantly correlated with the MARS Functionality subscale (r=0.74, *P*<.001), Aesthetics subscale (r=0.70, *P*<.01), and total score (r=0.58, *P*=.01). Number of evidence-based intervention components was not associated with MARS scores (r=0.085, *P*=.73) or consumer ratings (r=–0.329, *P*=.16).

Conclusions: In our analysis of popular research-supported consumer apps, objective app quality and subjective consumer ratings were generally high. App functionality and aesthetics were highly consistent with consumer appeal, whereas evidence-based components were not. In addition to designing treatments that work, we recommend that researchers prioritize aspects of app design that impact the user experience for engagement and sustainability (eg, ease of use, navigation, visual appeal). This will

help translate evidence-based interventions to the competitive consumer app market, thus bridging the gap between research development and real-world implementation.

KEYWORDS

Introduction

In the Digital Age, smartphones have permeated all aspects of personal and professional life. There is a robust market for mobile health (mHealth) apps focused on self-help for mental health and behavioral health needs [1,2]. Despite the widespread appeal of mHealth for mental health, we raise 2 important considerations for its adoption. First, development and testing of mental health apps in formal research settings are limited and far outpaced by everyday consumer use [2-5]. Second, app design elements such as engagement and functionality impact whether users continue to use a mobile app for sustained behavior change over time beyond initial download [6].

The aim of this study was to conduct a user-centered design analysis of the usability, engagement, and quality of popular evidence-based apps for mental health self-management utilizing the well-established Mobile Application Rating Scale (MARS) [7]. Previous publications have utilized the MARS to evaluate apps on an eclectic array of health-related topics including blood pressure, mindfulness, nutrition, diet and physical activity, deafness and hard-of-hearing, and drug–drug interactions [8-13]. This study evaluated the quality of mental health apps that are successful across both research and commercial sectors. We evaluated the relationships between objective app quality, subjective user ratings, and evidence-based behavior change techniques. This will help inform design considerations for mHealth researchers to maximize consumer appeal, engagement, and sustainability.

Methods

Overview

In a recent review of consumer apps, we identified 21 mental health self-management apps that were publicly available and research supported [5]. Two have since been removed by the developers, consistent with previous findings that consumer apps are retired at a rapid rate [5]. For this pool of 19 apps, we conducted the following data collection and analyses (February to April 2021) to address the current study objectives: MARS evaluations by 2 independent coders, extraction of consumer ratings from the App Store iOS (Apple Inc) and Google Play (Google LLC), coding of evidence-based treatment components, and correlation analyses.

We utilized the MARS, a validated objective measure for assessing the quality of mHealth apps [7]. The 23-item MARS provides a total score and Engagement, Functionality, Aesthetics, and Information Quality subscale scores. MARS total score and subscale scores range from 1 to 5, with higher scores representing better quality. Independent raters (NL and

AO) downloaded and evaluated each app; individual item scores were averaged between raters according to accepted standards [11]. App consumer ratings were extracted and averaged across App Store iOS and Google Play. We coded app content as evidence-based behavior change techniques (ie, based in behavior change theory or psychological interventions shown to be efficacious or effective) or not evidence based (ie, digital content/modules such as daily inspirational quotes that are not a component of traditional evidence-based mental health interventions).

Statistical Analysis

Interrater reliability was assessed using the intraclass correlation coefficient (ICC) according to established guidelines; we ran a 2-way mixed effects, average measures model with a consistency of agreement definition [14,15]. We calculated Cronbach α to assess the internal consistency of the MARS [16,17]. Descriptive statistics were used to summarize MARS scores, consumer ratings, and number of evidence-based treatment components. Pearson correlation coefficients were calculated to compare (1) the MARS overall score with consumer ratings, (2) the MARS subscale scores with consumer ratings, and (3) number of evidence-based treatment components with MARS and consumer ratings. All analyses were conducted in IBM SPSS Statistics version 27.

Results

The MARS demonstrated high interrater reliability (ICC 0.97, 95% CI 0.97-0.98), and the total score had high internal consistency (Cronbach α=.99). Table 1 shows the MARS total score and subscale scores, mean consumer ratings obtained from app download platforms, and number of evidence-based treatment components for each of the 19 apps. The MARS total mean score for all apps was 3.52 (SD 0.71; range 2.22-4.32). The MARS subscale mean scores were as follows: Engagement, 3.98 (SD 0.82; range 2.30-4.80); Functionality, 3.42 (SD 0.80; range 2.00-4.63); Aesthetics, 3.23 (SD 0.90; range 1.67-4.67); and Information Quality, 3.47 (0.69; range 2.00-4.29).

Average consumer ratings across the App Store iOS and Google Play were 4.22 (SD 0.54; range 3.24-4.86). Average number of evidence-based treatment components was 2.32 (SD 1.42; range 1-5). Notably, 8/19 (42%) of apps consisted of a singular approach to treatment (ie, had only 1 evidence-based treatment component). Headspace had both the highest MARS total score and highest consumer rating, and is a unimodal intervention that teaches mindfulness meditation for emotion regulation and health behavior change.

Each of the MARS subscales was significantly correlated with each other (r=0.63-0.88, *P*<.01) and the MARS total score

(r=0.84-0.91, *P*<.001). Consumer ratings were correlated with the MARS Functionality subscale (r=0.74, *P*<.001; Figure 1), Aesthetics subscale (r=0.70, *P*<.01; Figure 2), and total score (r=0.58, *P*=.01; Figure 3). Number of evidence-based treatment components was not significantly associated with MARS scores (r=0.09, *P*=.73) or average consumer ratings (r=−.33, *P*=.17).

Table 1. MARS scores, consumer ratings, and number of evidence-based treatment components.

App name	MARS scores					Consumer ratings		Total number of evidence-based components
	Engagement	Functionality	Aesthetics	Information quality	Total score	Average user ratings	Total number of ratings	
10% Happier	3.80	4.50	4.33	3.43	4.02	4.80	96,707	3
AEON Mindfulness	2.80	3.00	2.17	2.71	2.67	4.00	45	1
Calm	4.60	4.13	4.67	3.43	4.21	4.66	1,431,242	1
DeStressify	4.10	3.00	2.33	3.36	3.20	4.15	17	3
Habitica	4.20	3.00	3.00	2.57	3.19	4.28	19,354	1
Happify	4.50	3.75	3.67	3.57	3.87	4.22	5615	4
Headspace	4.40	4.25	4.33	4.29	4.32	4.86	861,451	1
MindSurf	3.00	2.00	1.67	2.21	2.22	3.24	18	4
MoodMission	4.50	2.25	2.67	3.86	3.32	3.31	203	5
One Moment Meditation	2.30	2.38	2.33	2.00	2.25	4.81	1281	1
Pacifica/Sanvello	4.80	4.00	4.17	4.00	4.24	4.62	32,749	5
Provider Resilience	3.40	3.25	3.00	4.00	3.41	3.43	43	2
PTSD Coach	4.60	3.75	4.00	4.07	4.11	4.66	1781	3
Smiling Mind	3.60	3.50	3.50	4.07	3.67	3.86	3600	2
Stop, Breathe and Think/MyLife Meditation	4.60	4.13	3.83	3.79	4.09	4.68	39,329	1
SuperBetter	4.60	4.00	3.67	4.07	4.00	4.50	12,737	1
T2 Mood Tracker	2.40	2.25	2.00	2.71	2.34	3.60	1875	1
Virtual Hope Box	4.80	3.25	2.50	4.00	3.64	3.86	1133	3
Woebot	4.60	4.63	3.50	3.71	4.11	4.73	13,041	2

Figure 1. MARS Functionality subscale score × consumer rating scatterplot of 19 apps with variations in bubble size proportionate to app's total number of consumer ratings. MARS: Mobile Application Rating Scale.
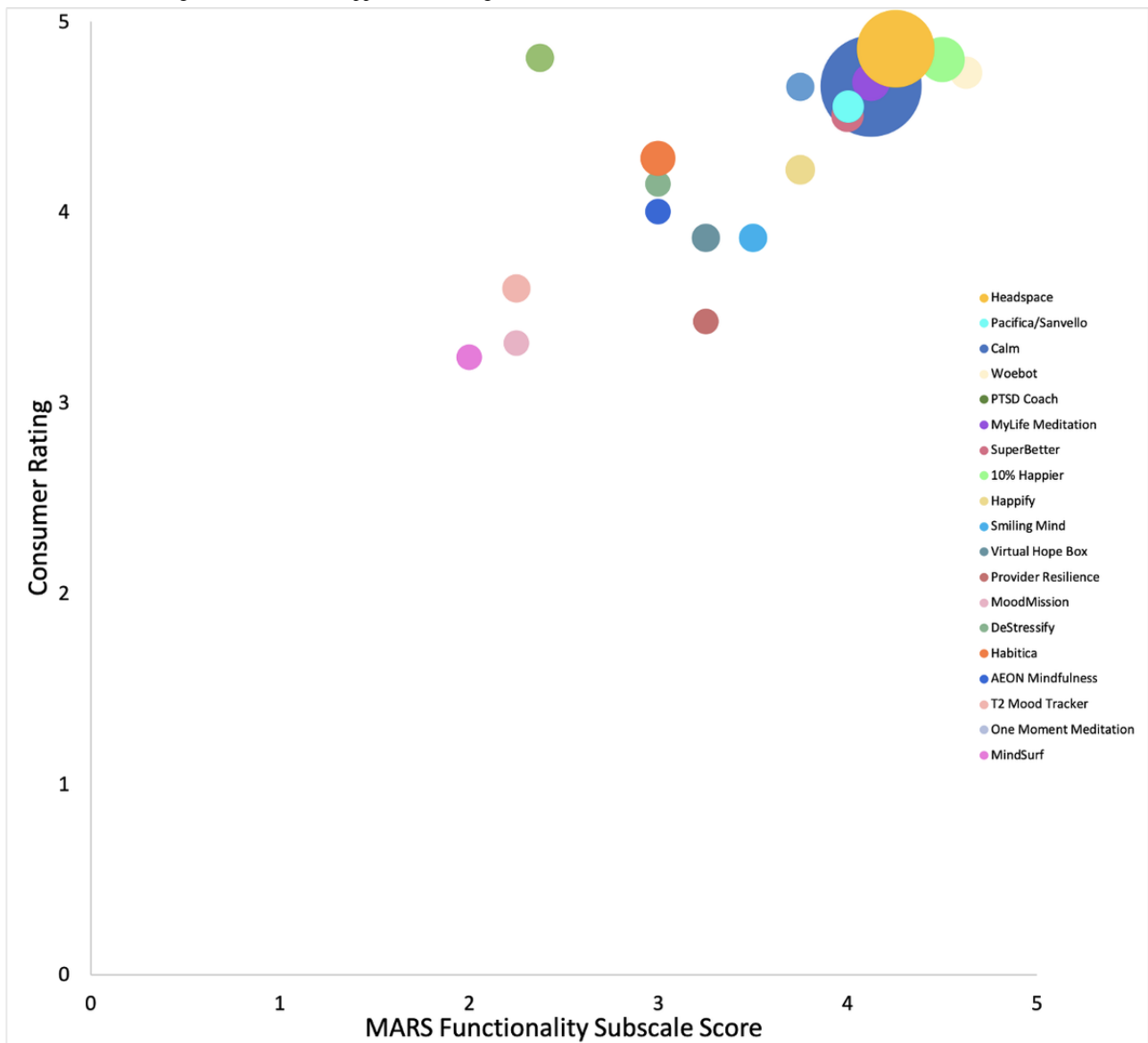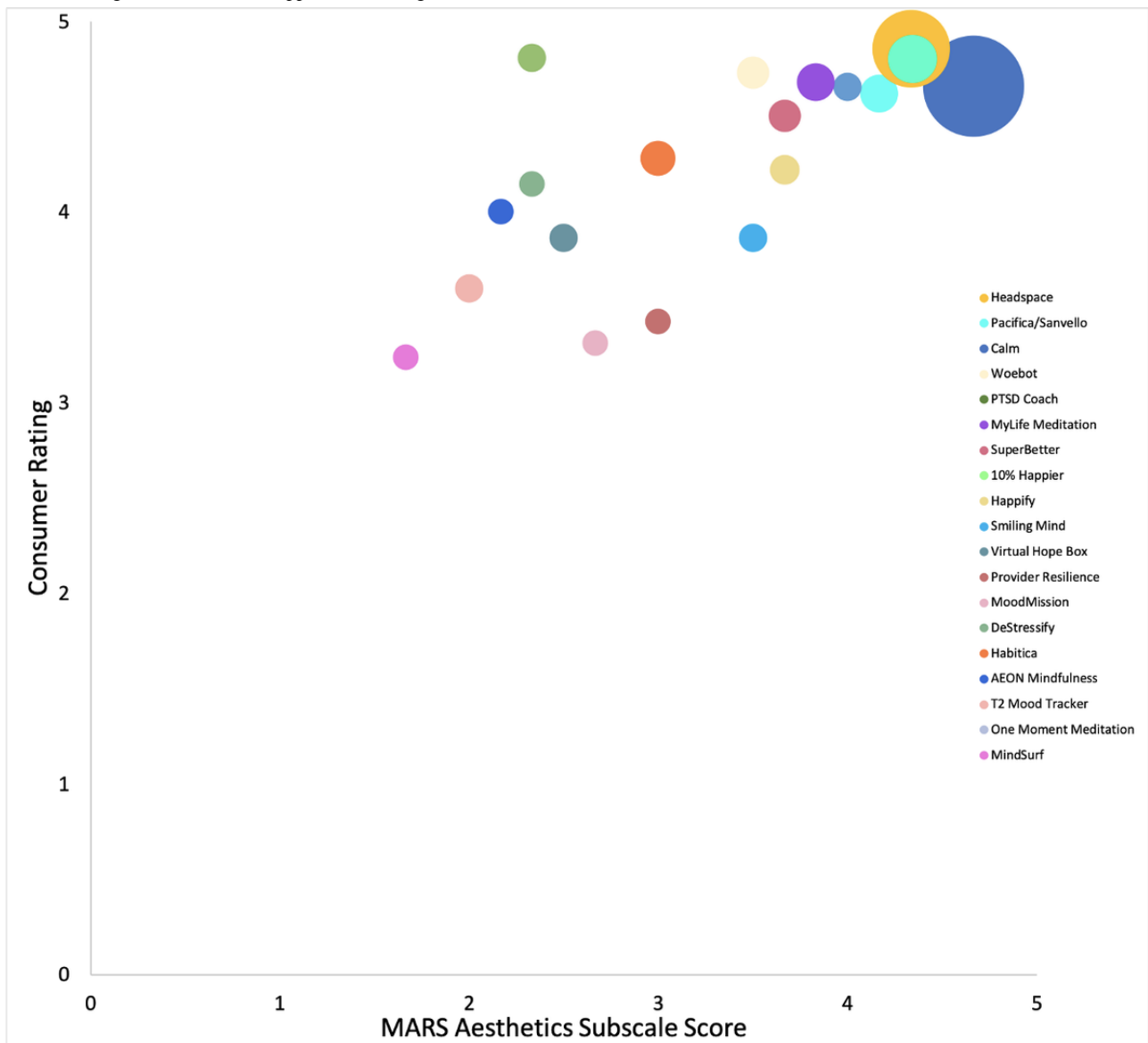
Figure 2. MARS Aesthetics subscale score × consumer rating scatterplot of 19 apps with variations in bubble size proportionate to app's total number of consumer ratings. MARS: Mobile Application Rating Scale.

Figure 3. MARS total score × consumer rating scatterplot of 19 apps with variations in bubble size proportionate to app's total number of consumer ratings. MARS: Mobile Application Rating Scale.
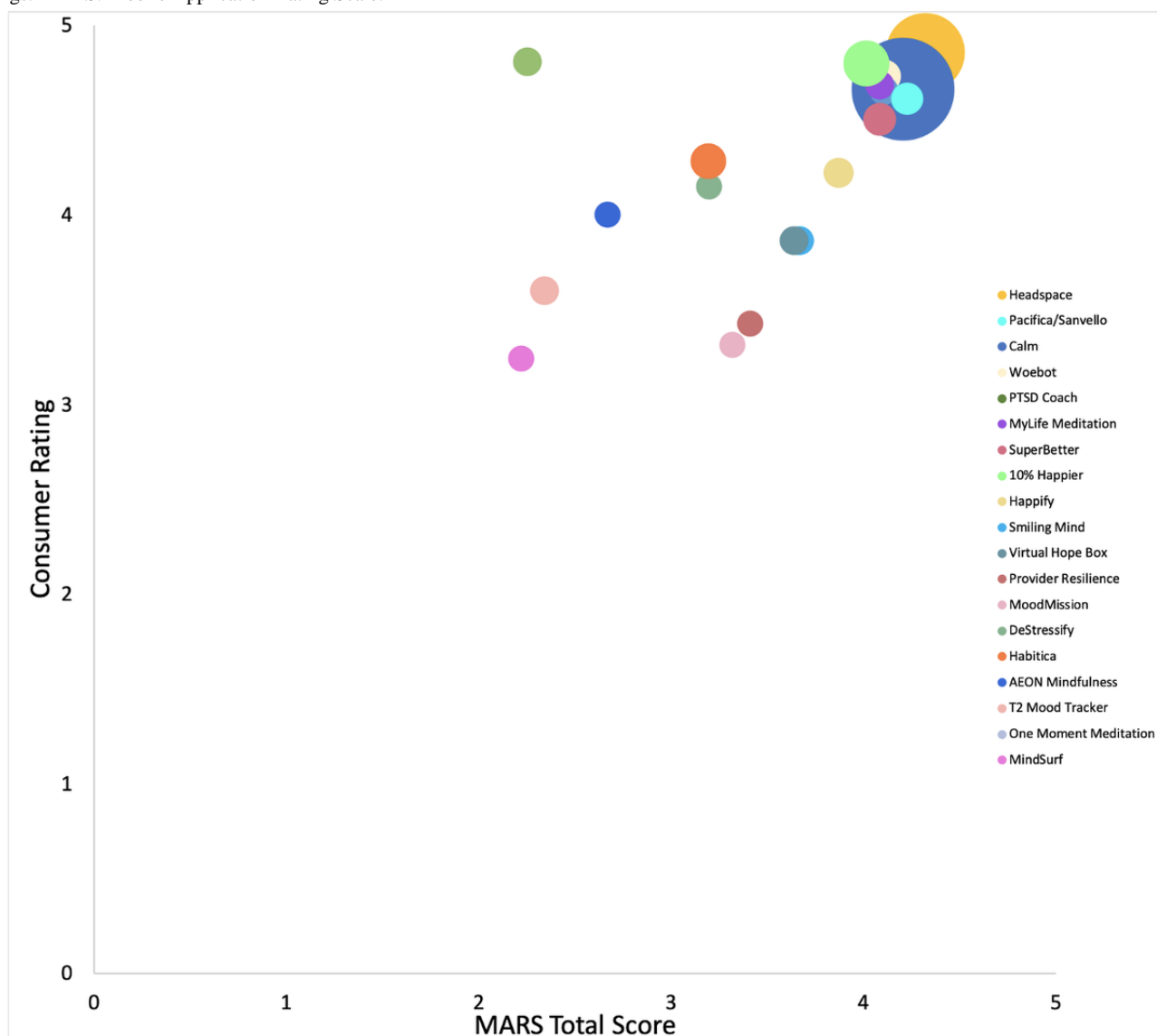
Discussion

Principal Findings

In this paper, we described an analysis of the usability and quality of popular research-supported consumer mental health apps using the MARS which provides a bite-sized synthesis of app usability and quality that are easily accessible to consumers and researchers alike [1,18]. The mental health apps we evaluated were of good quality overall. We observed overall MARS scores that were comparable to other reviews of self-management apps [12,19,20]. Previous research has compared MARS scores with consumer ratings for various health-related apps; results varied in whether MARS scores were correlated with consumer ratings [8,9,12].

With regard to popular research-supported mental health apps, we draw the following conclusions: First, we found that consumer ratings were related to objective quality of the app overall. This suggests alignment between subjective assessment of quality by app users and objective assessment of quality by researchers. Second, consumer ratings were related to

functionality and aesthetics. This suggests that design elements such as ease of use, navigation, graphics, and visual appeal may be more likely to impact the positivity of the user experience. Third, evidence-based treatment components were not associated with app quality or consumer ratings, and almost half of the apps had a singular skill focus. This suggests that quantity of evidence-based behavior change techniques designed and tested in traditional face-to-face mental health interventions is not what appeals to app consumers. Perhaps unimodal rather than multimodal intervention approaches lend themselves better to a self-guided format and decreases user burden.

Limitations

This study did not evaluate all mental health apps available for public download, nor would it be feasible to do so. Rather, our analysis focused on a small targeted subset of apps identified in a prior publication that we utilized to address a new research question focused on objective and subjective app quality and user-centered design considerations for engagement and sustainability. Thus, our analysis may not represent the whole of all available resources. Apps were limited to English language

that were available in the United States, downloadable on major app platforms, and with peer-reviewed publications.

Conclusions and Implications

We found that objective app quality—and functionality and aesthetics in particular—was highly consistent with consumer appeal. Quantity of evidence-based—and presumably effective—behavior change techniques was not associated with app quality or consumer appeal. To translate evidence-based interventions to the competitive consumer app space, researchers should prioritize aspects of design that impact the user experience such as ease of use, navigation, graphics, and visual appeal. This work may be informed by user-centered design approaches in which iterative development of apps prioritize end user's needs in the contexts in which the intervention will be implemented [21]. In addition, the complexity of evidence-based multimodal interventions may hinder chances of mHealth adoption. Adapting and optimizing design features to the individuals and settings that are unique to the digital space will help engage and retain users over time.

Conflicts of Interest

None declared.

References

1.   Proudfoot J, Parker G, Hadzi PD, Manicavasagar V, Adler E, Whitton A. Community attitudes to the appropriation of mobile phones for monitoring and managing depression, anxiety, and stress. J Med Internet Res 2010;12(5):e64 [FREE Full text] [doi: 10.2196/jmir.1475] [Medline: 21169174]

2.   Krebs P, Duncan DT. Health App Use Among US Mobile Phone Owners: A National Survey. JMIR Mhealth Uhealth 2015;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]

3.   Chib A, Lin SH. Theoretical Advancements in mHealth: A Systematic Review of Mobile Apps. J Health Commun 2018;23(10-11):909-955 [FREE Full text] [doi: 10.1080/10810730.2018.1544676] [Medline: 30449261]

4.   Bose J, Hedden S, Lipari R, Park-Lee E. Key substance use and mental health indicators in the United States: Results from the 2017 National Survey on Drug Use and Health (HHS Publication No. SMA 18-5068, NSDUH Series H-53). Rockville, MD: Center for Behavioral Health Statistics and Quality/Substance Abuse and Mental Health Services Administration; 2017. URL: https://www.samhsa.gov/data/report/2017-nsduh-annual-national-report [accessed 2021-06-27]

5.   Lau N, O'Daffer A, Colt S, Yi-Frazier JP, Palermo TM, McCauley E, et al. Android and iPhone Mobile Apps for Psychosocial Wellness and Stress Management: Systematic Search in App Stores and Literature Review. JMIR Mhealth Uhealth 2020 May 22;8(5):e17798 [FREE Full text] [doi: 10.2196/17798] [Medline: 32357125]

6.   Chang TR, Kaasinen E, Kaipainen K. What influences users' decisions to take apps into use? A framework for evaluating persuasive and engaging design in mobile apps for well-being. New York, NY: ACM; 2012 Presented at: Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia; December 2012; Ulm, Germany. [doi: 10.1145/2406367.2406370]

7.   Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]

8.   Jamaladin H, van de Belt TH, Luijpers LC, de Graaff FR, Bredie SJ, Roeleveld N, et al. Mobile Apps for Blood Pressure Monitoring: Systematic Search in App Stores and Content Analysis. JMIR Mhealth Uhealth 2018 Nov 14;6(11):e187 [FREE Full text] [doi: 10.2196/mhealth.9888] [Medline: 30429116]

9.   Kim BY, Sharafoddini A, Tran N, Wen EY, Lee J. Consumer Mobile Apps for Potential Drug-Drug Interaction Check: Systematic Review and Content Analysis Using the Mobile App Rating Scale (MARS). JMIR Mhealth Uhealth 2018 Mar 28;6(3):e74 [FREE Full text] [doi: 10.2196/mhealth.8613] [Medline: 29592848]

10. Romero RL, Kates F, Hart M, Ojeda A, Meirom I, Hardy S. Quality of Deaf and Hard-of-Hearing Mobile Apps: Evaluation Using the Mobile App Rating Scale (MARS) With Additional Criteria From a Content Expert. JMIR Mhealth Uhealth 2019 Oct 30;7(10):e14198 [FREE Full text] [doi: 10.2196/14198] [Medline: 31670695]

11. Schoeppe S, Alley S, Rebar AL, Hayman M, Bray NA, Van LW, et al. Apps to improve diet, physical activity and sedentary behaviour in children and adolescents: a review of quality, features and behaviour change techniques. Int J Behav Nutr Phys Act 2017 Jun 24;14(1):83 [FREE Full text] [doi: 10.1186/s12966-017-0538-3] [Medline: 28646889]

12. Flaherty S, McCarthy M, Collins A, McAuliffe F. Can existing mobile apps support healthier food purchasing behaviour? Content analysis of nutrition content, behaviour change theory and user quality integration. Public Health Nutr 2018 Feb;21(2):288-298. [doi: 10.1017/S1368980017002889] [Medline: 29081322]

13. Mani M, Kavanagh DJ, Hides L, Stoyanov SR. Review and Evaluation of Mindfulness-Based iPhone Apps. JMIR Mhealth Uhealth 2015 Aug 19;3(3):e82 [FREE Full text] [doi: 10.2196/mhealth.4328] [Medline: 26290327]

XSL•FO

RenderX

14. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979 Mar;86(2):420-428. [Medline: 18839484]

15. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychological Methods 1996;1(1):30-46. [doi: 10.1037/1082-989X.1.1.30]

16. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951 Sep;16(3):297-334. [doi: 10.1007/BF02310555]

17. Tavakol M, Dennick R. Making sense of Cronbach's alpha. Int J Med Educ 2011 Jun 27;2:53-55 [FREE Full text] [doi: 10.5116/ijme.4dfb.8dfd] [Medline: 28029643]

18. Schueller SM, Washburn JJ, Price M. Exploring Mental Health Providers' Interest in Using Web and Mobile-Based Tools in their Practices. Internet Interv 2016 May;4(2):145-151. [doi: 10.1016/j.invent.2016.06.004] [Medline: 28090438]

19. Weekly T, Walker N, Beck J, Akers S, Weaver M. A Review of Apps for Calming, Relaxation, and Mindfulness Interventions for Pediatric Palliative Care Patients. Children (Basel) 2018 Jan 26;5(2):16 [FREE Full text] [doi: 10.3390/children5020016] [Medline: 29373515]

20. Bardus M, van BSB, Smith JR, Abraham C. A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. Int J Behav Nutr Phys Act 2016;13(1):35 [FREE Full text] [doi: 10.1186/s12966-016-0359-9] [Medline: 26964880]

21. Lyon AR, Koerner K. User-Centered Design for Psychosocial Intervention Development and Implementation. Clin Psychol (New York) 2016 Jun;23(2):180-200 [FREE Full text] [doi: 10.1111/cpsp.12154] [Medline: 29456295]

Abbreviations

ICC: intraclass correlation coefficient
MARS: Mobile Application Rating Scale