

# Mental Health App Evaluation: Updating the American Psychiatric Association's Framework Through a Stakeholder-Engaged Workshop

Sarah Lagan, B.S., Margaret R. Emerson, D.N.P., A.P.R.N., Darlene King, M.D., Sonia Matwin, Ph.D., Steven R. Chan, M.D., M.B.A., Stephon Proctor, Ph.D., Julia Tartaglia, M.D., Karen L. Fortuna, Ph.D., L.C.S.W., Patrick Aquino, M.D., Robert Walker, M.S., C.O.A.P.S., Michelle Dirst, M.P.P., Nicole Benson, M.D., Keris Jan Myrick, M.B.A., M.S., Nathan Tatro, M.A., David Gratzner, M.D., F.R.C.P.(C.), John Torous, M.D.

The app evaluation framework of the American Psychiatric Association (APA) has emerged as an adaptable tool for those seeking to navigate the ever-growing space of mental health apps. The authors describe a meeting convened in December 2019 to refine the APA framework. The expert panel comprised 16 individuals across health care fields, with representation from psychiatry, psychology, social work, nursing, clinical informatics, peer support

specialists, and individuals with lived mental health experience. This meeting resulted in an update to the APA framework with the addition of clearer question descriptions and the release of an eight-question screener that may be useful in clinical settings.

*Psychiatric Services 2021; 72:1095–1098; doi: 10.1176/appi.ps.202000663*

With the large number of health-related mobile apps available on various marketplaces, it is not surprising that, although individuals may want to use apps to support mental health treatment, finding a safe and effective app is challenging (1). Although many prior app evaluation efforts have noted the need for multistakeholder engagement to advance app evaluation efforts, to date, most efforts have been created by or consisted of a single stakeholder group (2). To support and advance the decision-making process around mobile app technology, the American Psychiatric Association (APA) put out a call for the development of an app expert panel that would further expand the organization's ongoing app evaluation efforts (3). In this column, we outline the composition of the panel and explore how a 2-day meeting of this diverse group led to tangible changes in the APA app evaluation model.

## Description of the Panel and Meeting

The expert panel convened in December 2019 in Washington, D.C. Panelists were chosen from a pool of 71 applicants and were reviewed by both the APA Committee on Health Information Technology and a team of two predetermined panel leaders. Panelists were selected on the basis of demonstrated interest and experience in digital mental health

and were compensated fully for travel and related expenses. The interprofessional panel consisted of 16 members who traveled from across the United States and Canada, with representation from psychiatry (N=5), nursing (N=1), peer support specialists (N=2), psychology (N=2), social work (N=1), clinical informatics (N=1), medical students and trainees (N=2), as well as APA staff (N=2). The panel included service users and individuals with lived mental health experience

## HIGHLIGHTS

- In December 2019, a diverse panel (including representation from psychiatry, nursing, peer support specialists, psychology, social work, clinical informatics, service users and individuals with lived mental health experience, those training in health care, and staff from the American Psychiatric Association [APA]) convened to update the APA app evaluation framework.
- Over the course of the 2-day meeting, the APA app evaluation framework was updated by using a modified Delphi process to be more applicable to diverse audiences, and a new brief screener was created.
- The updated framework, new screener, and tutorial are publicly available online.

(for brief details on the panelists, see the online supplement to this column). The diversity of backgrounds and the array of experiences that were represented among the panelists highlighted the unique nature of what it means to be an expert in this field and accounted for the wide variety of contexts in which the framework may be implemented.

*APA app evaluation framework.* The levels in the original APA app evaluation framework included background information, risk-privacy and security, evidence, ease of use, and interoperability. A series of questions enabled an individual to consider various components related to each level. The framework was structured in a hierarchical fashion so that the evaluator could choose to stop the inquiry if he or she noted concerns about an app early in the evaluation process (3). For example, if an app lacked a privacy policy and did not specify data usage, limiting its use in clinical settings, there would be no need to continue the evaluation and consider facets of evidence and usability for a clinician or consumer concerned with maintaining a high standard of privacy and security. Before meeting in person, each of the panelists was asked to watch a three-part series available online through the APA, which provided the foundation for the framework and app evaluation process in its original format. This framework was created by harmonizing 961 questions from 45 different app evaluation frameworks (3, 4).

On the first day of the in-person meeting, the group convened and reviewed the APA app evaluation framework as it existed in its previous form. The group reviewed the most recent evidence on mobile apps for mental health and noted that limited evidence of benefit in high-quality studies and mounting privacy concerns warranted a cautious approach. The full framework was scrutinized through structured discussion among the panelists to ensure that the evaluation tool was effective in providing a thorough inquiry into the factors that clinicians and patients should consider when using a mobile app. Although the conversation typically focused on the use of apps within the psychiatric clinical practice setting, the group felt it was important to create a tool that could be used broadly to evaluate apps intended for use by an individual or clinician regardless of specialty.

To make decisions regarding the app evaluation framework, the panel used a modified Delphi procedure, with iterative discussion through each level until a consensus was reached. Three defined rounds of discussion took place. Although voting scores were not recorded, a consensus was reached after each round before proceeding to the next. First, the panel discussed whether to retain the initial 2018 framework with edits or to restructure it entirely. Although the panel decided to keep the hierarchical pyramid structure of the framework, in the second round of discussion, the five levels were slightly altered to better reflect their constituent questions. The first level became “accessibility and background” instead of “background information,” because

the panel emphasized a renewed focus on access. If an individual is unable to access an app because of lack of online connectivity or incompatibility with smartphone accessibility features, the app is not usable, and consideration of the framework’s upper levels is unnecessary. Access is thus a foundational component of evaluation, as reflected by the new questions in the first level.

The second level of the pyramid, which addresses privacy and security, remained largely the same, whereas the third level became “clinical foundation” instead of “evidence” to emphasize not only the app’s specific evidence but also the body of clinical research underpinning its various use cases. The group debated what level of evidence is necessary for an app and realized that this question is complex, especially because the lack of published research is not necessarily indicative of low quality of newer tools that may not have undergone more longitudinal testing. The fourth level remained largely the same, although the panel chose to use the label “usability” instead of “ease of use” to better incorporate the various features that users may be drawn to in an app. In the third round of discussion, efforts were made to offer more concrete questions to guide assessment of this very personalized and unique facet of app evaluation. Finally, the fifth level maintained its focus on data integration but was termed, more descriptively, “data integration toward therapeutic goal” instead of “interoperability.” In the third round of discussion, the original list of 27 questions was clarified and expanded to 36 questions (see online supplement).

*From questions to screener.* Another core objective for the first day of the meeting was to determine specific questions that providers and patients could use to complete a quick and simple screening process for an app intended for their use. All members agreed that although a more thorough assessment is always beneficial, it may be possible to offer users value with a brief screening tool that can be quickly applied. Thus, each level within the APA app evaluation framework was discussed and evaluated to determine which questions were key to providing a brief but thorough screening of mobile apps. The 36 questions from the full framework were reduced to eight key questions across the five broader levels (see online supplement). This “screener” can be completed more quickly than the full framework but retains the spirit and priorities of the full version. The screener version was not designed to replace the full version but rather to offer a tool more applicable to a busy clinical setting. Although it accomplishes the goal of providing a condensed, easily deployable tool, the screener does not prompt users to consider as many dimensions of the app as the full framework.

*Putting the framework in action.* The second day of the in-person meeting tasked the panel with putting the APA app evaluation framework into practice. The day commenced

with a group evaluation of a chatbot-based app that provided online cognitive-behavioral therapy. Together, the panel went through the process of evaluating this chatbot by using the updated framework. On the basis of this group exercise, minor adjustments were made to the wording of questions. After this group evaluation, each panelist evaluated an app of their choosing and presented the findings to the group. These evaluations were conducted in pairs to identify any confusing questions that may have needed further refinement. The framework proved to be a feasible tool for a wide variety of apps, ranging from child-focused apps, to apps with corresponding wearable devices, to popular artificial intelligence therapy apps. Each of these diverse app types was evaluated by two panelists, with the framework evaluations demonstrating strong reliability between pair members.

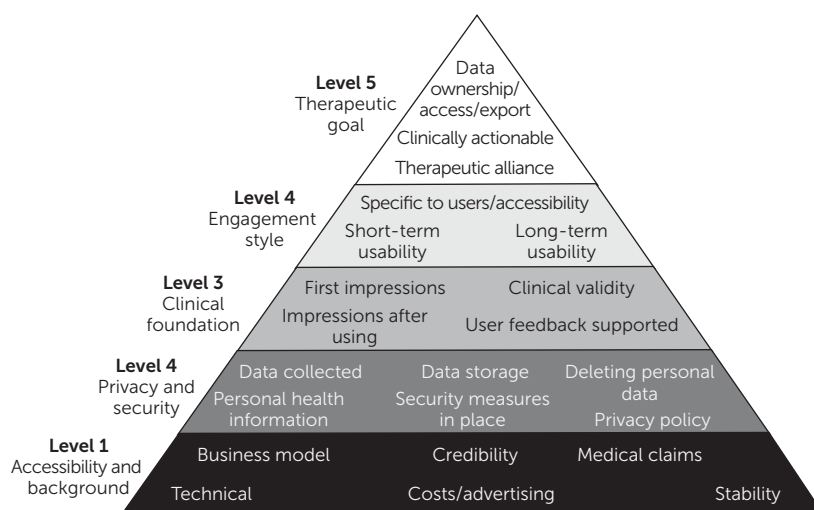
The evaluation begins at the foundational level, with a consideration of background and accessibility information. At this level, we addressed questions pertaining to various facets of access, including cost, offline functionality, and stability (reflected by an update within the past 180 days). The evaluation then proceeded through privacy and security, clinical foundation, engagement style and usability, and data integration toward therapeutic goal (Figure 1). Tutorials for evaluating 11 apps with this framework are now publicly available on the APA website (<https://www.psychiatry.org/psychiatrists/practice/mental-health-apps>), and 11 tutorials apply the new framework to diverse mental health-related apps.

## Future Directions

The December 2019 APA meeting was the first to convene a panel of experts with diverse backgrounds, ranging from peer support, to psychiatry, to nursing, to social work, for a weekend of structured discussion and to use a modified Delphi procedure to refine the APA app evaluation framework. A major contribution of the panel is the publication of a framework with two versions: a comprehensive 36-question framework and an abbreviated screener that poses eight critical questions. This screener facilitates easier adoption of the framework by busy clinicians and patients with limited time to spare.

What distinguishes this framework from other emerging app evaluation efforts is its comprehensiveness, flexibility, and potential for adoption in diverse contexts. Since the panel met in December 2019, the framework has been adopted by both a large public mental health system (New York City Department of Mental Health and Hygiene; <https://nycwell.cityofnewyork.us/en/app-library>) and a community mental health system (Vinfen; <https://vinfen.org/resources/vinfen-app-library>) to produce tangible outcomes: app libraries and detailed guides for consumers of mental health apps, tailored to specific patient populations. It has also been used to inform a database of mental health apps (<https://apps.digitalpsych.org>) (5) and other app-rating frameworks (6). These use cases underscore the feasibility of the APA framework in diverse settings.

**FIGURE 1. Updated American Psychiatric Association mental health app evaluation framework, by hierarchical level**



org/resources/vinfen-app-library) to produce tangible outcomes: app libraries and detailed guides for consumers of mental health apps, tailored to specific patient populations. It has also been used to inform a database of mental health apps (<https://apps.digitalpsych.org>) (5) and other app-rating frameworks (6). These use cases underscore the feasibility of the APA framework in diverse settings.

As the number of app-rating tools increases alongside the number of health apps, the APA app evaluation framework is unique in its focus on factors from accessibility, to security, to clinical foundation. The framework addresses all the domains identified as key standards for app evaluation by mobile health leaders in industry and academia (7). In reviews of app evaluation models by third parties, the APA framework has stood out for its comprehensive analysis of privacy (8). It was also selected by the Federal Trade Commission as an example of patient protection at the virtual PrivacyCon 2020.

The panel recognizes the limitations of this work. The implementation of app evaluation in clinical settings often requires sensitivity to local needs, which may require tools beyond the guiding questions of the APA app evaluation framework. Implementation guides, such as the Technology Evaluation and Assessment Criteria for Health apps (9), offer a useful means of tailoring the framework to specific contexts.

The landscape of mental health apps is rapidly changing. Although this framework is intended to be durable and adaptable, the APA panel recognizes the need to make updates when necessary. The panel will continue to be involved in app evaluation efforts, with each panelist aiding in the publication of instructional videos on the APA website.

## AUTHOR AND ARTICLE INFORMATION

Division of Digital Psychiatry, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston (Lagan, Torous); College of Nursing,

University of Nebraska Medical Center, Omaha (Emerson); Department of Psychiatry, Utah Southwestern Medical Center, Dallas (King); Department of Psychiatry and Behavioral Sciences, Vanderbilt University Medical Center, Nashville, Tennessee (Matwin); U.S. Department of Veterans Affairs Palo Alto Health Care System, and Department of Psychiatry, Stanford Medical Center, Palo Alto, California (Chan); Behavioral Health Informatics, Children's Hospital of Philadelphia, Philadelphia (Proctor); Department of Psychiatry, Northwell Health, New York City (Tartaglia); Department of Psychiatry, Geisel School of Medicine, Dartmouth College, Hanover, New Hampshire (Fortuna); Department of Psychiatry, Lahey Hospital and Medical Center, Boston (Aquino); Department of Mental Health, Office of Recovery and Empowerment, Boston (Walker); Department of Practice Management and Delivery Systems Policy (Dirst) and Department of Digital Health (Tatro), American Psychiatric Association, Washington, D.C.; Department of Psychiatry, McLean Hospital, Belmont, Massachusetts (Benson); Los Angeles County Department of Mental Health, Los Angeles (Myrick); Scarborough Hospital Toronto, and Department of Psychiatry, University of Toronto, Toronto (Gratzer). Dror Ben-Zeev, Ph.D., is editor of this column. Send correspondence to Dr. Torous (jtorous@bidmc.harvard.edu).

This work was supported by a grant from Jeremy Wertheimer. Dr. Fortuna was funded by a K01 award from the National Institute of Mental Health (K01-MH-117496). Dr. Benson receives support from the National Library of Medicine (NLM-T15-LM007092).

Dr. Chan performs academic research for the University of California, Davis, under contract and was formerly compensated as a clinical fellow at the University of California, San Francisco. Dr. Chan formerly saw patients as a contracted physician of HealthLinkNow and Traditions Behavioral Health and taught at and was financially compensated by Guidewell Innovation and by the North American Center for Continuing Medical Education. He was financially compensated by Advanced Clinical for consulting; Scholastic Expeditions and the Arizona Psychiatric Society for teaching; the University of California, Davis, for contract consulting; and the University of Wisconsin–Madison School of Public Health. He consulted with Orbit Telepsychiatry for potential future stock options yet to be granted. Dr. Fortuna provides consulting services through Social Wellness. Dr. Myrick has had travel expenses

covered by the American Psychiatric Association. Dr. Torous reports unrelated research support from Otsuka.

The other authors report no financial relationships with commercial interests.

Received September 5, 2020; revisions received October 7 and November 6, 2020; accepted November 12, 2020; published online April 22, 2021.

## REFERENCES

1. Torous J, Roberts LW: Needed innovation in digital health and smartphone applications for mental health: transparency and trust. *JAMA Psychiatry* 2017; 74:437–438
2. Rodriguez-Villa E, Torous J: Regulating digital health technologies with transparency: the case for dynamic and multi-stakeholder evaluation. *BMC Med* 2019; 17:226
3. Torous JB, Chan SR, Gipson SYT, et al: A hierarchical framework for evaluation and informed decision making regarding smartphone apps for clinical care. *Psychiatr Serv* 2018; 69:498–500
4. Henson P, David G, Albright K, et al: Deriving a practical framework for the evaluation of health apps. *Lancet Digit Health* 2019; 1:e52–e54
5. Lagan S, Aquino P, Emerson MR, et al: Actionable health app evaluation: translating expert frameworks into objective metrics. *NPJ Digit Med* 2020; 3:100
6. Levine DM, Co Z, Newmark LP, et al: Design and testing of a mobile health application rating tool. *NPJ Digit Med* 2020; 3:74
7. Torous J, Andersson G, Bertagnoli A, et al: Towards a consensus around standards for smartphone apps and digital mental health. *World Psychiatry* 2019; 18:97–98
8. Nurgalieva L, O'Callaghan D, Doherty G: Security and privacy of mHealth applications: a scoping review. *IEEE Access* 2020; 8: 104247–104268
9. Camacho E, Hoffman L, Lagan S, et al: Technology Evaluation and Assessment Criteria for Health apps (TEACH-apps): pilot study. *J Med Internet Res* 2020; 22:e18346