CO-CARES

Colorado Alliance for Resilient & Equitable Systems
Serving the Health Care & Public Health Workforce

CO-CARES Mobile
Health Evaluation Toolkit

Behavioral Health &
Wellness Program

CU
Department of Psychiatry
SCHOOL OF MEDICINE
UNIVERSITY OF COLORADO ANSCHUTZ MEDICAL CAMPUS

This CO-CARES Mobile Health Evaluation Toolkit was completed by the
University of Colorado Anschutz Medical Campus,
School of Medicine, Department of Psychiatry
Behavioral Health and Wellness Program
November 2023

For further information about this report, please contact:

Behavioral Health and Wellness Program
University of Colorado Anschutz Medical Campus
School of Medicine
Department of Psychiatry
1890 N Revere Street
Mail Stop F478
Aurora, Colorado 80045

Phone: 303.724.3713
Email: bh.wellness@ucdenver.edu
Website: www.bhwellness.org

Citation for this Toolkit: Mescher, T., Hacker, R.L., Morris, C.D., & Garver-Apgar, C.E., (2023). CO-CARES Mobile Health Evaluation Toolkit.

About this Toolkit

Who is this toolkit for?
This toolkit is designed for health care workers that would benefit from implementing mobile health into their organizations or who want to make mobile health recommendations to patients. These materials are intended for healthcare and behavioral healthcare providers, system administrators, and health officials.

How does one use this toolkit?
This toolkit contains evidence-informed information about mobile health, why and how to evaluate it, and free searchable mobile health databases. The information herein may be used to develop an evaluation plan for mobile health in an organization and to evaluate mobile health applications prior to implementing and/or recommending to patients.

Why Focus on Mobile Health?

In 2008, both Apple and Google launched their app stores; Apple launched 500 initial apps and Google just 50. Fifteen years later, however, Apple has nearly 2 million apps, Google has more than 3 million, and more are added daily (Apple, 2023; Google, 2023). With between 100,000 and 350,000 mobile health (mHealth) apps currently available, many people turn to mHealth apps for added support in accomplishing their health and well-being goals (Ceci, 2022a-b; Levine et al., 2020). mHealth apps are making their way into clinical practice via patients asking providers about them and providers recommending specific apps to patients. Identifying which mHealth apps should be recommended by healthcare providers and used by patients can be daunting due to the volume, lack of oversight, and minimal review of mHealth apps. By October 2020, more than 79 frameworks for evaluating mHealth apps existed in the literature, and this number has likely grown as frameworks are added and revised with advances in the field (Lagan et al., 2021a).

What is in this toolkit?
This toolkit overviews why app evaluation is important and describes both cautions and benefits of using mHealth. Frameworks and scales are reviewed for use in health care settings. While other frameworks and scales could have been included, those selected are comprehensive, widely used, accessible to providers, have validation data or were peer reviewed, and are believed to be most beneficial in helping providers make recommendations in treatment settings. This toolkit also offers a summary of what should be included in an mHealth evaluation framework or scale, describes what the implementation process could look like, and introduces two free searchable databases developed using two of the frameworks/scales reviewed herein that may facilitate app selection. The goal is to help healthcare providers competently and confidently select mHealth options to recommend to patients, implement into treatment workflows, and improve patient outcomes.

What is mHealth?

Digital Health encompasses all technologies used to revolutionize health care (e.g., artificial intelligence applications; eHealth; e-tools to monitor, change, and evaluate personal health domains; electronic medical records; mHealth; telehealth and telemedicine platforms; wearable devices; etc.) and exists across several domains (Table 1). Digital Health has "the vast potential to improve our ability to accurately diagnose and treat disease and to enhance the delivery of health care to all people" (U.S. Food & Drug Administration, 2020). mHealth is a subset of Digital Health that specifically uses applications (apps) on mobile devices and smartphones to accomplish these goals (Chan, 2021). Coinciding with increased functionality and use of smartphones, tablets, and other personal electronic devices, the use of mHealth has increased exponentially in the past two decades.

Table 1. Digital Health Domains

Domain	Examples
Education	Pregnancy education, language learning
Fitness	GPS activity, guided classes or exercises
Medical Record	Provider- or clinic-provided medical information app
Nutrition	Calorie and macronutrient counting, diet-specific recipes
Psychological Interventions	Online therapy, meditation, self-care, cognitive-behavioral interventions
Reminders	Medication or treatment reminders
Symptom Tracking	Migraine symptoms, bipolar disorder events, menstrual cycle

mHealth can facilitate positive change in clinical practice and consumer health. For instance, mHealth can help patients monitor symptoms and medication adherence between appointments and easily share real-time data directly with providers. Collecting and reviewing this information outside of appointment times could allow for more accurate data collection compared to retroactive recall, allow providers to use the brief face-to-face time they have with patients in an overburdened healthcare system more efficiently, and ultimately improve patient and provider relationships and health outcomes (American Psychiatric Association, 2023a). mHealth provides opportunities for health care teams and patients to record data, to set and observe progress toward goals, and to increase access to interventions that improve health outcomes. However, comprehensive evaluation is needed to address concerns about widespread implementation of mHealth apps in healthcare settings.

Why is evaluating mHealth important?

Evaluating mHealth prior to implementation into clinic workflows or recommending apps to patients is crucial. Finding the most appropriate app that will meet clinic and patient needs is difficult since mHealth apps are on an exponential rise. Additionally, existing apps get updated, some disappear, and new apps become available regularly (American Psychiatric Association, 2023a). It is also essential that providers are confident that mHealth apps are reliable, user friendly, and appropriate for their needs before implementing and recommending them. Healthcare providers often hesitate to recommend mHealth apps to patients because they do not have adequate knowledge of

available apps (Wangler and Jansky, 2021). Wangler and Jansky (2021) surveyed 2,138 German primary care physicians and found that only 18% of respondents recommended apps frequently or occasionally; over half never recommended apps. Teaching providers how to evaluate mHealth could increase the use of mHealth apps in clinical settings. Providers would gain the knowledge necessary to recommend appropriate apps to their patients while also determining which apps may potentially improve patient health outcomes.

While many mHealth options are available and may promote positive change for patients and clinics, little is known about their efficacy. Apps are often not evaluated beyond the app store's 5-star rating system, which does not correlate with whether an app is effective at doing what it purports to do (Levine et al., 2020). mHealth clinical trials are rarely conducted, and deployment in clinical settings is often unstudied. Thus, it is important to assess whether the app has an evidence base.

The notable interest in mHealth and the growing market have enabled many companies, often led by investors and tech specialists, not health or medical experts, to focus on building and selling mHealth without adequately testing for feasibility and effectiveness, nor sufficiently considering data security. Apps not developed by health content experts in the field for which they are to be implemented may deliver inaccurate information that could lead to adverse outcomes. Therefore, patient health and safety may be at risk if an app has not been properly evaluated (Roberts et al., 2021). Apps should be assessed on whether they were developed via scientific methods and by a company with a health focus, not a company simply trying to break into a lucrative, rapidly expanding industry.

In addition, mHealth is not under the purview of the US Food and Drug Administration (FDA) like other medical devices. While the FDA does recognize this gap and deployed a pilot program for regulating mHealth between 2017-2021, the Pre-Cert Program is not finalized nor comprehensive (U.S. Food and Drug Administration, 2022; Lagan et al., 2021a). Thus, very little oversight and regulation leaves end users to determine whether an mHealth product is safe, effective, and capable of doing what it advertises. End user ratings, as reflected in app store star-based ratings, do not correlate with comprehensive evaluation framework ratings (Levine et al., 2020). mHealth products, therefore, are often not developed nor properly tested with efficacy and reliability in mind. This often leads to apps failing, not being updated, and ultimately being removed from app stores. Some critics fear, and evidence supports, that mHealth may cause harm to consumers, violating the healthcare principle of beneficence. With little oversight, mHealth products may provide incorrect or misleading information to consumers, be ineffective, lack privacy and security measures, or sell users' personal data (American Psychiatric Association, 2023a). In fact, a variety of well-known mHealth apps have faced legal consequences in response to their sharing of sensitive information with external marketing and analytics firms and being ineffective and misleading in what they claimed to do (Wicklund, 2017). The Federal Trade Commission (FTC) has also acted against deception and false advertising by mHealth apps (Wagner, 2020). Therefore, mHealth apps should be assessed for privacy and security since they

are being used in healthcare settings where protecting patients' personal health information is of the utmost importance.

Overall, it is important to evaluate mHealth apps for effectiveness, quality, and safety. Apps should have an evidence base, as noted above, however, it is very common for companies to develop and release apps without testing their effectiveness, much less assess whether the app is delivering high-quality information and/or data that could harm patients. Since the FDA does not currently oversee mHealth, it is the end user's responsibility to evaluate any app they may use, recommend to patients, or implement in a clinic.

What should one look for in an evaluation framework or scale?
The most recent and thorough review of mHealth apps was completed in 2021 and spanned the literature through October 2020 (Lagan et al., 2021a). Lagan et al (2021a) reviewed 70 different apps, yet the number continues to grow. In addition, existing apps and tools continually update and important aspects of them may change over time. Since the mHealth industry is constantly evolving, it is important to know how to evaluate a tool or app rather than relying on the app reviews in the scientific literature. An early step to evaluating mHealth apps and tools is to identify an appropriate framework. Answering the following questions will assist in choosing a framework for a specific app or tool.

WHAT METRICS SHOULD THE FRAMEWORK OR SCALE ASSESS?
mHealth evaluation frameworks and scales evaluate a variety of criteria, most often assessing for usability and accessibility (e.g., learnability, efficiency, memorability, errors, and satisfaction; Aljaber et al., 2015). Additional criteria include engagement, privacy and security, subjective and content quality, therapeutic goal/impact, and evidence-base/research. Reliability is also commonly assessed via error prevention and consistency, and standards and productivity may be measured using metrics for efficiency, effectiveness, attractiveness, responses, and time (Aljaber et al., 2015).

DOES THE FRAMEWORK OR SCALE ASSESS THE TYPE OF TECHNOLOGY OF INTEREST?
When assessing mHealth, it is important that the selected framework or scale have been developed and tested for use with mHealth specifically. The use of apps in health care settings creates unique uses and needs, including clinical setting use nuances and privacy and security considerations, which need to be considered by the selected framework or scale.

IS THE FRAMEWORK OR SCALE EVIDENCE-INFORMED AND TESTED?
It is important to choose an evaluation tool that was based on scientific literature and has been peer reviewed. In addition, a scale should be well-researched and statistically validated. Having a peer-reviewed evidence base increases the credibility of a framework or scale.

mHealth Evaluation Frameworks and Scales

How were evaluation frameworks and scales chosen for this toolkit?

Evaluation frameworks and scales were required to meet an established list of criteria to be included in this toolkit. First, included frameworks and scales had to be accessible online, available in English, and focused on mHealth specifically, rather than general health or broader digital health. Frameworks and scales had to have an evidence base, and scientific data needed to be used during the development process. It was also important that they were published in a peer reviewed journal and that scales were validated for internal consistency and interrater reliability, at minimum. Finally, as this toolkit focuses on recommending mHealth to patients or integrating mHealth into healthcare clinics, assessing privacy and security was strongly preferred. If a framework or scale was created for a specific population (e.g., older adults), to measure a specific type of outcome (e.g., potential for behavior change), or did not assess for privacy and security but were otherwise well-designed and researched, they were included with the suggestion that they be used in conjunction with a more comprehensive framework or scale.

There are several frameworks and scales available online that were not peer reviewed in scientific literature or lacked validation data and were excluded from this toolkit for those reasons. The authors believe those selected for this toolkit are comparable and comprehensive. Below are analyses of five frameworks and scales that meet the above criteria: 1) the Adapted Mobile App Rating Scale, 2) the American Psychiatric Association App Evaluation Model, 3) THESIS, 4) mHealth for Older Users, and 5) the App Behavior Change Scale.

Adapted Mobile App Rating Scale

The Mobile App Rating Scale (MARS) was developed in 2015 to evaluate user experience and was the first app evaluation tool. It was updated to the Adapted MARS (A-MARS) by a team at the University of Sydney's Brain and Mind Centre to expand its use for e-tool (e.g., websites, online courses) evaluation, in addition to apps (Roberts et al., 2021). While A-MARS is comprehensive, widely used, and available in multiple languages, it does not evaluate privacy or security, crucial considerations for health care settings.

The 28-item scale measures app and e-tool quality across eight domains (Engagement, Functionality, Aesthetics, Information, Subjective Quality, and Health-Related Quality), and items are scored on a five-point Likert scale where higher scores are better. Each domain receives an average score, and each app is assigned an overall score (see Appendix A for the scale). Preliminary internal testing of the A-MARS showed high internal consistency (α = 0.94) and interrater reliability. In addition, A-MARS maintains this consistency when looking at the first four domains independent from the two quality-related subscales (α = 0.91; Roberts et al., 2021).

A-MARS was created specifically for use with mHealth and can evaluate both mHealth apps and e-tools, whereas most scales evaluate apps only (Roberts et al., 2021). Implementing A-MARS requires adequate training and is time-consuming. The creators

recommend piloting the scale by evaluating 3-5 apps or e-tools and having multiple people intentionally rate each app or e-tool under consideration to ensure interrater reliability. Therefore, providers may not have the time nor skill to effectively utilize A-MARS. The developers encourage organizations to consider hiring a digital navigator to focus on identifying and evaluating mHealth specific to their patient population (Roberts et al., 2021). Importantly for healthcare settings, A-MARS does not assess data privacy or security. This is a significant limitation, so A-MARS should be used in conjunction with a more comprehensive framework or scale. Alternatively, security and privacy-specific items from other frameworks may be integrated to assess these domains separately and decrease risk to patient data.

SUMMARY: A-MARS is an effective and reliable scale that can be used to evaluate apps and e-tools throughout the development process. It can also be used by adequately trained health care providers to identify appropriate mHealth apps and e-tools for their practices, with the understanding that this tool does not assess data privacy and security.

American Psychiatric Association App Evaluation Model

The American Psychiatric Association (APA) App Evaluation Model was originally developed to help psychiatric providers, other health care providers, and their patients identify apps to support their mental health-related treatment goals. The APA recognized app selection needs to be an individualized decision and should consider many factors unique to the patient, the clinical context, and version of the app (Torous et al., 2018). Patients regularly use apps and ask for provider input, yet the APA acknowledged that psychiatrists and other health care providers do not receive adequate training to identify appropriate apps or make sufficiently informed recommendations through their traditional education (American Psychiatric Association, 2023b). Additionally, previous frameworks did not reliably consider an app's safety and usefulness (Torous et al., 2018). In 2018, the APA first offered their framework for mental health apps. It was expanded by an international multistakeholder app expert panel, including people with lived experience, in 2021 to include use across all mHealth apps, increased focus on accessibility, supporting clinical research for use case, and the use of more accurate terminology throughout (Lagan et al., 2021b). This same committee also developed a screener consisting of eight key questions that could be more easily applied in busy clinical settings (Lagan et al., 2021b).

The APA Framework presents an "adaptable scaffold for informed decision making" when selecting an mHealth app (Torous et al., 2018). This structure uses hierarchical stages, presented as a pyramid, which allow the evaluator to stop an app evaluation if concerns are noted early in the process. The five levels, starting with the most foundational, are Accessibility, Privacy and Security, Clinical Foundation, Engagement Style, and Therapeutic Goal (Fig. 1; see Appendix B for all items). While this framework does not give a formal score for questions, levels, or the app overall, it ensures healthcare providers and their patients have adequate information to make an informed decision based on their unique circumstances (American Psychiatric Association, 2023b).

The APA Framework is straightforward, comprehensive, flexible, and relevant to diverse contexts. It prioritizes accessibility, privacy, and security, all of which are important considerations in healthcare settings. This framework encompasses all the domains identified as key standards for app evaluation by mHealth leaders in both industry and academia and it can be used across all mHealth settings. Furthermore, the screener version makes the framework more accessible to busy clinicians and patients with limited time. The APA website provides a comprehensive page overviewing their model; it includes information about the screener, sample evaluations, and video tutorials on how to use the model to rate apps (American Psychiatric Association, 2023a).
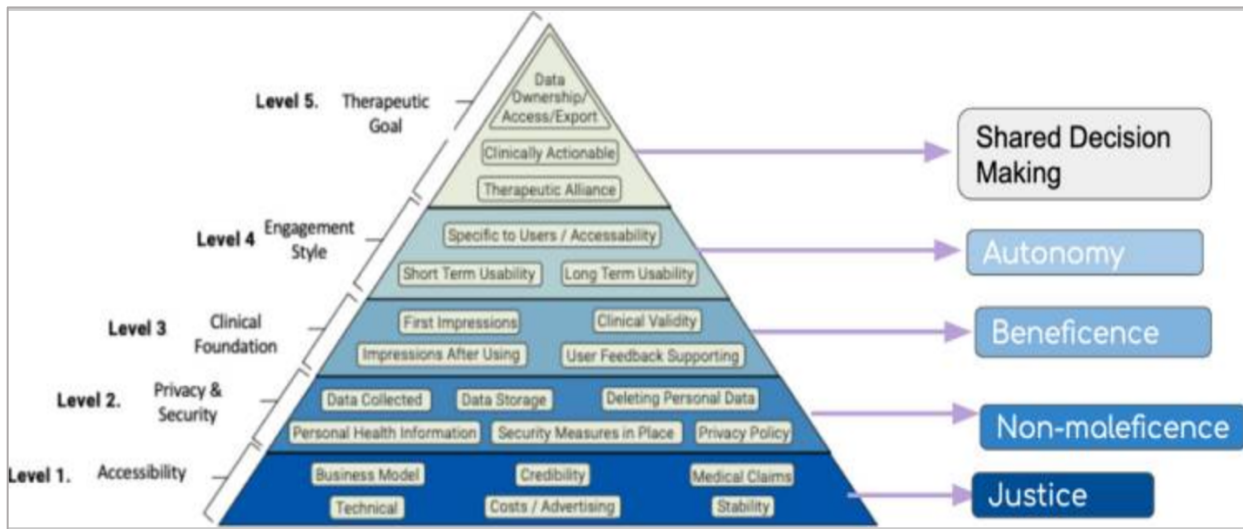

Figure 1. American Psychiatric Association mental health app evaluation framework from Lagan et al, 2020.

SUMMARY: The APA Framework is a comprehensive, adaptable, and hierarchical model, developed by experts and people with lived experience, that can be used by health care providers and patients to make informed decisions on which mHealth apps can best support individual health and behavioral health goals.

THESIS

The THESIS rating tool was developed to bridge gaps in other evaluation tools (e.g., concerns of access, privacy, security, and interoperability) and to specifically evaluate mHealth apps for chronic disease, including mental health conditions. The developers recognized that most apps are developed for "relatively healthy patients and few are developed specifically for high-cost, high-need patients, or patients with chronic disease" (Levine et al., 2020). It is well known that longitudinal care benefits this population, yet few apps are developed and intended for long-term use. A panel of experts and patient representatives was convened in 2017 to rate and review criteria identified through review of other evaluation tools; criteria about bandwidth and device memory requirements not present in other tools were added, considering the population of interest. Bandwidth and use of device memory significantly impact whether a patient

can download and use an mHealth app, particularly important considerations for people with chronic conditions within lower socioeconomic demographics.

THESIS is an acronym for the six domains evaluated by this tool (Transparency, Health Content, Excellent Technical Content, Security/Privacy, Usability, and Subjective). A score of 1-5 is provided for the app overall and within each domain allowing for comparison across apps (see Appendix C for the scale).

THESIS developers acknowledge that their rating scale does not cover every aspect of mHealth apps that may be of interest; instead, their goal was to create a scale that was quick to use. THESIS can be completed in about 12 minutes per mHealth app by raters with a college-level education or tech background. Inter-rater reliability was moderate (K = 0.3-0.6) and internal consistency is high (α = 0.85). THESIS could benefit from further validation using a larger cohort of raters from varying backgrounds and areas of expertise, however (Levine et al., 2020).

SUMMARY: THESIS is a rating tool and acronym that can be used to evaluate mHealth apps for chronic health conditions, including mental health conditions. It is one of the only frameworks or scales that includes consideration of app size and the impacts of this on access; however, it may be less accessible to raters without a college degree or a tech background.

mHealth for Older Users

It is common for mHealth evaluation tools and frameworks to assess usability; however, they fail to address important considerations for the aging population who "interact differently with information technology compared to younger people" (Wildenbos et al., 2018). To remedy this, Wildenbos, et al. (2018) created the mHealth for Older Users (MOLD-US) framework to assess mHealth usability for the growing older adult population. Based on the scientific literature, they identified four barriers that influence mHealth usability: 1) cognition, 2) motivation, 3) physical ability, and 4) perception (Fig. 2, bright blue). From there, they identified medical conditions that contribute to these barriers (Fig. 2, light blue). Finally, they identified five elements of mHealth user experience that can be affected by these barriers and medical conditions (Fig. 2, orange). For example, cognitive barriers may include decline in working
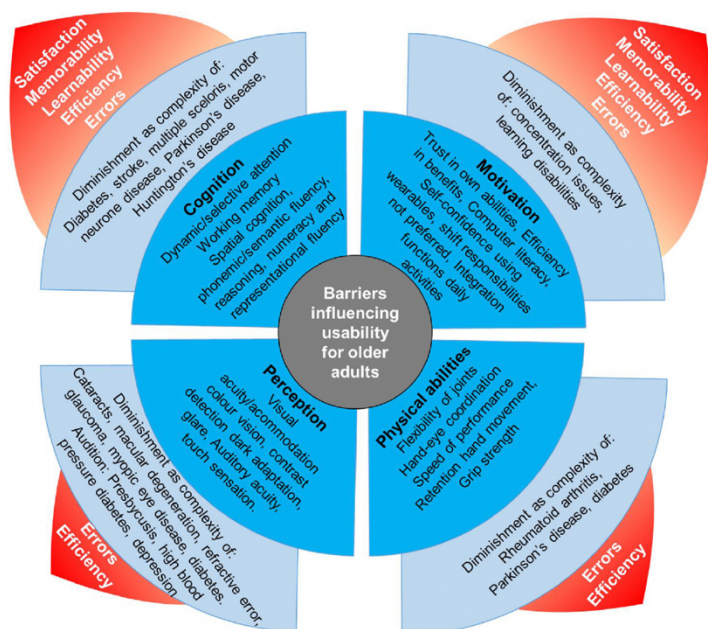

Figure 2. mHealth for Older Users (MOLD-US) evaluation framework from Wildenbos, et al., 2018.

memory due to stroke, which may lead to satisfaction, memorability, learnability, efficiency, and error issues in the user experience (Fig. 2, top left quartile).

MOLD-US is unique in identifying barriers to use for older adults while also providing examples of health conditions that may lead to specific usability issues. It provides a roadmap for developers and end users to comprehensively evaluate mHealth for usability through a population-specific lens, thus filling a large gap in the implementation of and access to mHealth in aging populations. Importantly, these results can "be classified and interpreted based on impediments intrinsic to encountered issues" (Wildenbos et al., 2018). MOLD-US, however, lacks a validated scale that can be used to assess these unique usability barriers in mHealth, making adoption of MOLD-US in health care settings more difficult. It is only recommended for evaluation of apps for older adults and apps for medical conditions most often diagnosed in older adults, but MOLD-US could be used to inform other innovative population-specific usability frameworks in the future. In addition, MOLD-US should be utilized in conjunction with a more comprehensive framework or scale, such as APA or THESIS since it does not assess other important aspects of mHealth, such as data privacy and security.

SUMMARY: MOLD-US is a framework that identifies mHealth barriers specific to older adults. It describes how these barriers are related to common medical conditions in this population and what aspects of usability may be affected by both the barriers and medical conditions. While filling an important gap in mHealth development and implementation, it does not have a validated scale for rating app usability for older adults, thus making adoption in health care settings difficult. MOLD-US should be used in conjunction with another more comprehensive framework or scale.

App Behavior Change Scale
Evaluating mHealth for its design, usability, data privacy and security, and other features is crucial, but if an app claims to change behavior, it is also important to assess its potential to do so. The App Behavior Change Scale (ABACUS) is a validated 21-item scale that assesses an app's potential behavior change across four domains: 1) Knowledge and Information, 2) Goals and Planning, 3) Feedback and Monitoring, and 4) Actions (McKay et al., 2019; Alslaity et al., 2022). See Appendix D for the scale. Based on health behavior change interventions, the scale underwent rigorous testing prior to final validation. Researchers validated the scale by rating 20 apps using the scale; results indicated high internal consistency ($\alpha = 0.93$) and interrater reliability (McKay et al., 2019).

In addition to high consistency and reliability, ABACUS is widely used in the literature. Its innovation in assessing behavior change is a strength that is missing from the majority of existing mHealth frameworks and tools. While measuring *potential* for behavior change may not translate to actual behavior change, it may act as a proxy for effectiveness when deciding whether to recommend an app to a patient. Since ABACUS specifically measures potential behavior change, it does not measure other important aspects of mHealth, such as usability, data privacy and security, subjective (e.g., look and feel) quality, content (e.g., health information) quality, evidence-base/research, and

engagement. Therefore, ABACUS should be used in conjunction with another more comprehensive framework or scale.

SUMMARY: ABACUS is a validated and widely used scale to measure mHealth apps' potential behavior change. It may serve as a proxy to measure potential effectiveness of mHealth apps and should only be used in conjunction with another scale or framework that assesses other critical evaluation criteria, including usability, data privacy and security, subjective and content quality, evidence-base and research, and engagement.

The Implementation Process

A model for app evaluation and implementation

Proper mHealth evaluation includes more than simply assessing mHealth for various criteria. Camacho, et al. (2023) describes a comprehensive model for evaluating and implementing mHealth apps (Fig. 3). The Technology Evaluation and Assessment Criteria for Health Apps (TEACH-Apps) model has four stages: Pre-Conditions, Pre-Implementation, Implementation, and Maintenance and Evolution. The model begins with identifying the needs of the local end users and connecting with stakeholders to gather names of apps for consideration. Considering the findings from the pre-
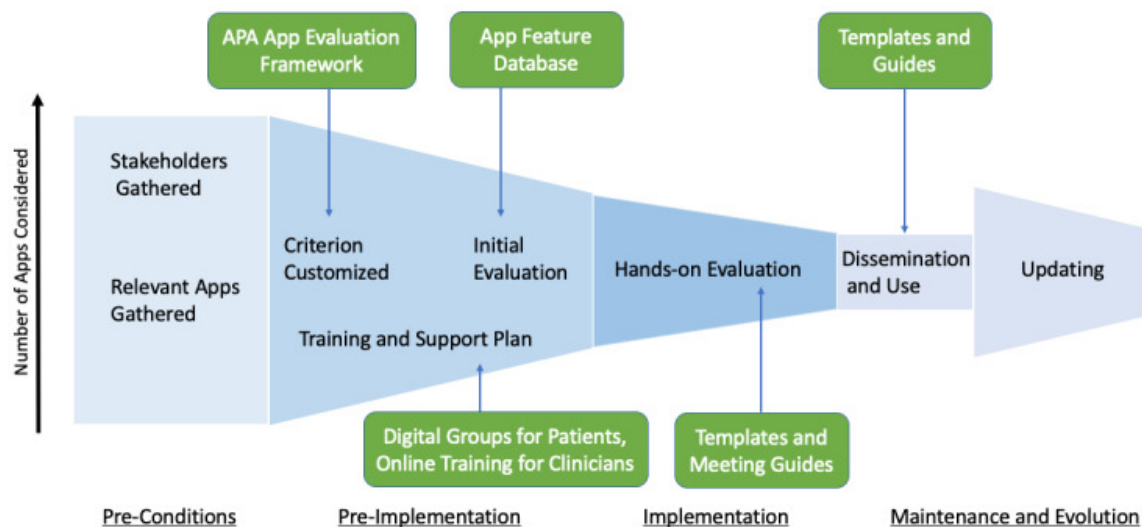
Figure 3. Technology Evaluation and Assessment Criteria for Health Apps (TEACH-Apps) Model from Camacho et al., 2023.

conditions phase, pre-implementation involves identifying criteria that reflect the priorities and needs of the local end users. A committee then evaluates each app using frameworks or scales, such as those described herein, to identify which apps best align with the defined criteria (Camacho et al., 2023). Once inclusion or exclusion decisions are made for all apps, the organization moves onto implementation. During this phase, committee members test the apps and provide feedback to determine which apps are the best fit for the setting. Once apps are selected, an educational handout (e.g., flyer, webpage) discussing each app's pros and cons should be created; the apps and handout are then offered to patients. Since mHealth apps constantly change, the

maintenance and evolution phase is ongoing. It is recommended that evaluations and handouts be updated quarterly, but at least twice a year, (Camacho et al., 2023).

mHealth Evaluation Databases

Providers should independently evaluate apps prior to integrating mHealth into a clinic or recommending them to patients. This helps ensure the selected app(s) will meet the patients' and/or clinics' needs. A few free searchable databases do the work of evaluating mHealth apps based on frameworks reviewed herein, so their ratings and reviews may be used to assist in identifying the best apps for the given circumstances. However, these searchable databases often focus on evaluating apps targeting mental health. Below, we review two searchable databases for mHealth apps that most often target mental health.

mHealth Index and Navigation Database

The mHealth Index and Navigation Database (MIND) is a searchable database of more than 600 health apps based on the APA App Evaluation Model. While the majority are mental health-focused, other mHealth apps are also included (e.g., fitness apps, food diary apps, and physical health trackers) In addition to the five APA levels noted above, MIND includes review of how data gets into the app and how it outputs information (Fig. 4, Inputs and Outputs). MIND was developed by the Digital Psychiatry Lab, a research team at the Beth Israel Deaconess Medical Center, in collaboration with the APA. It uses 105 objective yes/no questions to make the APA Framework "functionable and actionable for public use" (Division of Digital Psychiatry, 2023). See Appendix E for the
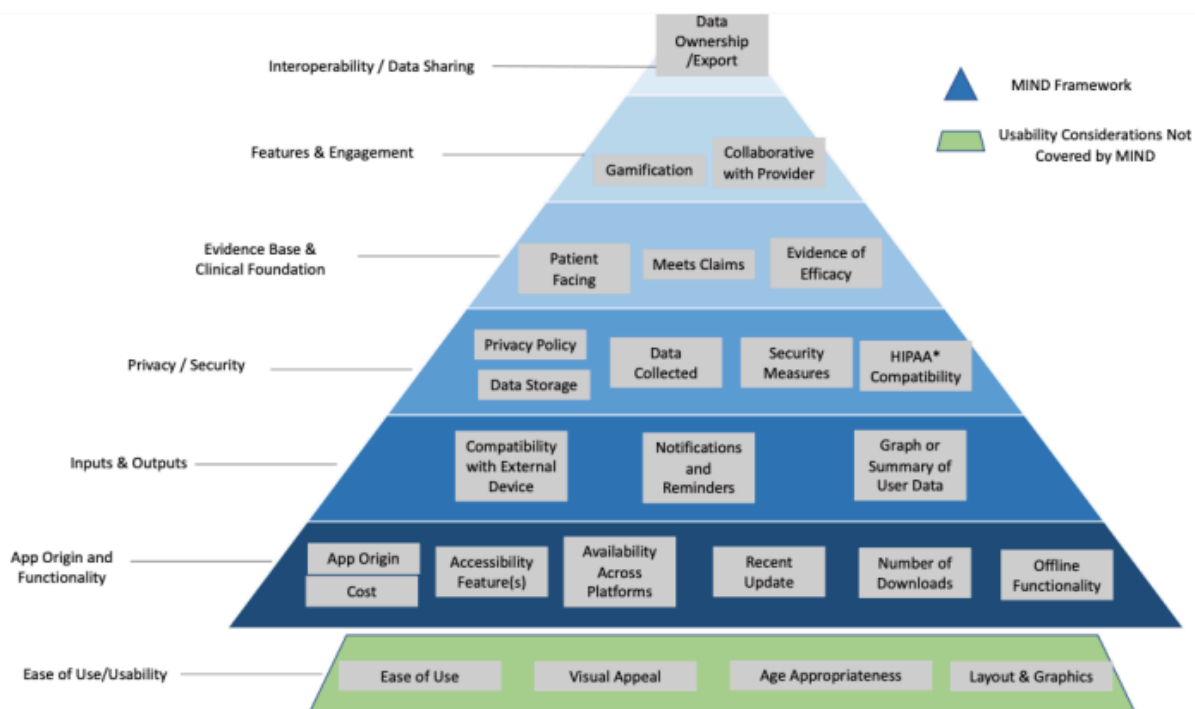

Figure 4. mHealth Index and Navigation Database's (MIND) most frequently addressed questions and unaddressed usability considerations from Lagan et al., 2021a.

full set of questions. Subjective questions, such as whether the app is easy to use, and objective questions that are not easily answered by available app data are not included (Lagan et al., 2021a). MIND does not provide a formal score; however, the database is searchable based on what is most important to the end user. MIND allows providers and patients to identify mHealth apps that may be most useful given the clinical context and patient's individual goals. App reviews are completed by trained volunteer app raters who completed a three-hour online training program and are reviewed by a member of the Digital Psychiatry Lab before the review is published in MIND. Apps are rated every six months to increase review accuracy as apps update.

One Mind PsyberGuide

One Mind PsyberGuide is another searchable database of over 230 mental health apps that is operated through a collaboration between Northwestern University and University of California, Irvine (One Mind PsyberGuide, 2023a & c). The database was created in 2013 and is a non-profit whose goal is to improve access to high quality mental health apps that can improve mental wellness without bias or endorsement (One Mind PsyberGuide, 2023a). A team of mental health and technology experts developed and maintains the database by assessing app credibility, user experience, and transparency. Credibility includes items focused on proposed goals, evidence-based content, research base and independence, software updates, development team and process. In addition to subjective quality and perceived impact scores, the original MARS is used by this team to evaluate user experience, including engagement, functionality, aesthetics, and information. Finally, transparency is scored as acceptable, questionable, or unacceptable based on strict guidelines. To be rated as having acceptable transparency, an app must provide accessible information and must conform to standard policies regarding data collection, storage, and exchange (One Mind PsyberGuide, 2023a). See Appendix F for the scoring criteria.

One Mind PsyberGuide is user-friendly, allows users to filter results by mental health concern, platform, audience, and cost, and provides clear professional reviews for some apps. The database prioritizes reviews of popular apps based on the number of reviews in the Apple and Google Play App Stores, but requests for an app review can be submitted by contacting PsyberGuide. Many apps listed, however, have incomplete ratings (e.g., credibility rating but no other information or reviews), ratings are not regularly updated (e.g., some as old as 7 years), and the database only evaluates mental health apps (One Mind PsyberGuide, 2023a-c). One Mind PsyberGuide's policy regarding updating ratings and re-evaluating apps as they update and evolve is unclear.

Conclusions and Best Practices

mHealth apps have the potential to revolutionize healthcare, to make treatment more accessible, make health data more accurate, and improve health outcomes. Yet, identifying, evaluating, and recommending mHealth apps to patients is complicated, particularly when added to the regular demands of the healthcare system in the United States and because app developers are often more focused on profit than testing their apps for efficacy, security, and reliability. Hundreds of thousands of mHealth apps

currently exist and more are being added, others updated, and some go obsolete. Thus, mHealth apps are constantly in flux, the number of proposed evaluation frameworks is overwhelming, and the time commitment can be a barrier to implementation.

This toolkit reviewed mHealth evaluation frameworks that show the greatest potential for easing app selection for healthcare providers. Two free searchable databases that review mental health apps were presented and an implementation process for healthcare settings with available resources was introduced. The goal of this document is to facilitate more streamlined implementation of mHealth into healthcare through an overview of available frameworks, scales, and other available resources.

A-MARS expands the first framework for mHealth evaluation; it focuses on user experience, is beneficial for app developers, and can be used to evaluate mHealth apps and e-tools by end users. A-MARS does not adequately consider data security, and thus should be used in conjunction with a more thorough framework that considers this.

The APA framework provides a more comprehensive framework that can be applied to all mHealth apps. It prioritizes accessibility and data safety and security considerations and encourages individualized app selection, not by giving an overall score, but rather providing sufficient information for end users to determine whether an app is appropriate for their needs.

THESIS builds on MARS by adding considerations specific to data safety and security and evaluation of mHealth apps for chronic disease, whereas MOLD-US provides a framework for considering unique factors when developing and selecting apps for aging adults and related to diseases more common to this demographic. The ABACUS framework offers a unique evaluation process for mHealth apps focused on facilitating behavior change. Each of these frameworks is more narrowly focused and should ideally be used in conjunction with a more comprehensive evaluation framework, such as A-MARS or APA. Table 2 summarizes criteria each framework, scale, and searchable database assesses.

Frameworks herein require training for implementation, vary in completion times, and may be seen as burdensome for busy healthcare professionals. Two searchable databases are currently available where trained evaluators complete the evaluation process. Databases may be useful for identifying mHealth apps, but they often focus reviews on mental health apps. One Mind PsyberGuide reviews popular apps based on app store ratings and focuses their evaluation on credibility, user experience, and transparency while utilizing the MARS scale. MIND reviews more apps, uses the more comprehensive APA framework, and includes additional items focused on app inputs and outputs. MIND invites people from all demographics to get involved in the evaluation process and includes formal training and professional review of evaluations before inclusion in the database. Additionally, MIND has a built-in process for ongoing app review so database users can be confident that the information is recent, whereas One Mind PsyberGuide's ongoing review process is unclear.

Table 2. Summary of Criteria Assessed in Frameworks, Scales, and Searchable Databases

	Accessibility/ Usability	Engagement	Privacy/ Security	Subjective or Content Quality	Efficacy Data/ Evidence-Base
Frameworks and Scales					
A-MARS	X	X		X	X
APA	X	X	X	X	X
THESIS	X		X	X	X
MOLD-US*	X				
ABACUS**					
Searchable Databases					
MIND	X	X	X	X	X
One Mind PsyberGuide	X	X	X	X	X

*MOLD-US is a usability-focused framework in the older adult population that does not assess other typical criteria assessed in other frameworks, scales, and searchable databases
**ABACUS is a non-traditional app evaluation scale that only measures aspects of potential behavior change, not the typical criteria assessed in other frameworks, scales, and searchable databases.

For healthcare settings with the resources, a four-stage implementation process has been shown to be effective for identifying mHealth apps that may best meet the needs of a healthcare setting's unique patient population. This process ideally involves the development of a committee, creation of resources that can be easily shared, and a plan for ongoing evaluation and updates.

In summary, selecting mHealth apps can generate significant burden in healthcare systems that are already overstretched; however, committing to a formal evaluation and implementation process can enhance services, alleviate system burden over time, improve health outcomes, and increase patient satisfaction.

References

Aljaber, T., Gordon, N., Kambhampati, C., & Brayshaw, M. (2015) An evaluation framework for mobile health education software. *2015 Science and Information Conference* 786-790. https://doi.org/10.1109/SAI.2015.7237233

Alslaity, A., Suruliraj, B., Oyebode, O., Fowles, J., Steeves, D., & Orji, R. (2022) Mobile applications for health and wellness: A systematic review. *Proc ACM Hum-Comput Interact 6*(171):1-29. https://doi.org/10.1145/3534525

American Psychiatric Association. (2023a) App Advisor: An American Psychiatric Association Initiative. *American Psychiatric Association.* https://www.psychiatry.org/psychiatrists/practice/mental-health-apps

American Psychiatric Association. (2023b) App Advisor: An American Psychiatric Association initiative: The App Evaluation Model. *American Psychiatric Association.* https://www.psychiatry.org/psychiatrists/practice/mental-health-apps/the-app-evaluation-model

Apple (2023) App Store. *Apple.* https://www.apple.com/app-store/

Camacho, E., Hoffman, L., Lagan, S., Rodriguez-Villa, E., Rauseo-Ricupero, N., Wisniewksi, H., Hensen, P., & Torous, J. (2020) Technology Evaluation and Assessment Criteria for Health Apps (TEACH-Apps): Pilot study. *J Med Internet Res 22*(8):e18346. https://doi.org/10.2196%2F18346

Ceci, L. (2022a) Apple App Store: Number of available medical apps as of Q3 2022. *Statista.* https://www.statista.com/statistics/779910/health-apps-available-ios-worldwide/

Ceci, L. (2022b) Google Play: Number of available medical apps as of Q3 2022. *Statista.* https://www.statista.com/statistics/779919/health-apps-available-google-play-worldwide/

Chan, J. (2021) Exploring digital health care: eHealth, mHealth, and librarian opportunities. *J Med Libr Assoc 109*(3):376-381. https://doi.org/10.5195%2Fjmla.2021.1180

Division of Digital Psychiatry (2023) M-Health Index & Navigation Database. *Beth Israel Deaconess Medical Center* https://mindapps.org

Google (2023) How Google Play Works. *Google.* https://play.google/howplayworks/

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 Digit Med* 3:100. https://doi.org/10.1038/s41746-020-00312-4

Lagan, S., Sandler, L., & Torous, J. (2021a) Evaluating evaluation frameworks: A scoping review of frameworks for assessing health apps. *BMJ Open 11*:e047001. https://doi.org/10.1136/bmjopen-2020-047001

Lagan, S., Emerson, M. R., King, D., Matwin, S., Chan, S. R., Proctor, S., Tartaglia, J., Fortuna, K. L., Aquino, P., Walker, R., Dirst, M., Benson, N., Myrick, K. J., Tatro, N., Gratzer, D., & Torous, J. (2021b) Mental health app evaluation: Updating the American Psychiatric Association's framework through a stakeholder-engaged workshop. *Psychiatric services 72*(9):1095–1098. https://doi.org/10.1176/appi.ps.202000663

Levine, D.M., Co, Z., Newmark, L.P., Groisser, A.R., Holmgren, A.J., Haas, J.S., & Bates, D.W. (2020) Design and testing of a mobile health application rating tool. *NPJ Digit Med 3*:74. https://doi.org/10.1038/s41746-020-0268-9

McKay, F.H., Slyderman, S., & Dunn, M. (2019) The App Behavior Change Scale: Creation of a scale to assess the potential of apps to promote behavior change. *JMIR Mhealth Uhealth 7*(1):e11130 https://doi.org/10.2196/11130

One Mind PsyberGuide (2023a) About One Mind PsyberGuide. *One Mind PsyberGuide.* https://onemindpsyberguide.org/about-psyberguide/

One Mind PsyberGuide (2023b) Beating the Blues: A professional review. *One Mind PsyberGuide.* https://onemindpsyberguide.org/expert-review/beating-the-blues-expert-review/

One Mind PsyberGuide (2023c) Help me find an app. *One Mind PsyberGuide.* https://onemindpsyberguide.org/apps/

Roberts, A.E., Davenport, T.A., Wong, T., Moon, H., Hickie, I.B., & LaMonica, H.M. (2021) Evaluating the quality and safety of health-related apps and e-toolsL Adapting the Mobile App Rating Scale and developing a quality assurance protocol. *Internet Interv 24*:100379. https://doi.org/10.1016/j.invent.2021.100379

Torous, J.B., Chan, S.R., Tan Gipson, S.Y., Kim, J.W., Nguyen, T., Luo, J., & Wang, P. (2018) A hierarchical framework for evaluation and informed decision making regarding smartphone apps for clinical care. *Psychiatr Serv 69*(5):498-500. https://doi.org/10.1176/appi.ps.201700423

U.S. Food and Drug Administration (September 22, 2020) What is Digital Health? *U.S. Food and Drug Administration.* https://www.fda.gov/medical-devices/digital-health-center-excellence/what-digital-health

U.S. Food and Drug Administration (September 2022) The Software Precertification (Pre-Cert) pilot program Tailored total product lifecycle approaches and key findings. *U.S. Food and Drug Administration.* https://www.fda.gov/media/161815/download?attachment

 Wagner, J.K. (2020) The Federal Trade Commission and consumer protection for mobile health apps. *J Law Med Ethics 48*(1 Suppl):103-114 https://doi.org/10.1177%2F1073110520917035

Wangler, J. & Jansky, M. (2021) The use of health happs in primary care – Results from a survey amongst general practitioners in Germany. *Wien Med Worchenschr 171*:148-156. https://doi.org/10.1007/s10354-021-00814-0

Wicklund, E., (March 24, 2017) mHealth apps for cardiac care sanctioned for misleading claims. *mHealth Intelligence* https://mhealthintelligence.com/news/mhealth-apps-for-cardiac-care-sanctioned-for-misleading-claims

Wildenbos, G.A., Puete, L., & Jaspers, M. (2018) Aging barriers influencing mobile health usability for older adults: A literature based framework (MOLD-US). *Int J Med Inform 114*:66-75. https://doi.org/10.1016/j.ijmedinf.2018.03.012

Adapted Mobile App Rating Scale (A-MARS)
APP / ETOOL NAME: _____

SECTION A
Engagement – fun, interesting, customizable, interactive, has prompts (e.g., sends alerts, messages, reminders, feedback, enables sharing)

1. Engagement: Is the app/e-tool engaging for the user? Do you feel engaged enough to complete the e-tool program or use the app on multiple occasions? Does it have components that make you want to use it more than similar apps/e-tools?
    1 - Dull, not engaging, not encouraged to start using app/e-tool
    2 - Mostly boring, would start program but never finish OR would download app and only use once or twice
    3 - OK, engaging enough to us app for a brief time (<5 min), would finish up to half an e-tool program
    4 - Moderately engaging, would use app/e-tool for some time; most likely complete the full program
    5 - Highly engaging, would stimulate repeat use of app/e-tool or would use for 5-10 min, OR would finish the full program, complete additional programs and/or return to use program again and engaged in tools/strategies given/learnt

2. Interest: Is the app/e-tool interesting to use? Does it present information in an interesting way compared to other apps/e-tools or offline/traditional tools?
    1 - Not interesting at all,
    2 - Mostly uninteresting
    3 - OK, neither interesting nor uninteresting; appears the same as offline/traditional tools/similar apps
    4 - Moderately interesting; would engage user for some time, somewhat more interesting than offline/traditional tools/similar apps
    5 - Very interesting, would engage user in repeat use, more engaging than traditional/offline tools/similar apps

3. Customization: Does the app/e-tool need to be customized to make it more user *friendly for you to use? Can you change settings such as sound, content, notifications, email/SMS reminders, display more to your liking?*
    1 - App/e-tool is not user friendly; has no customization options or requires setting to be input every time
    2 - App/e-tool allows little customization; App/e-tool could be improved with more customization options
    3 - Basic customization to function adequately and/or can use app/e-tool without customization
    4 - Allows numerous options for customization and/or easy to use app/e-tool, customization somewhat unnecessary

5 - Does not need any customization for me to use the app/e-tool effectively; Allows complete tailoring the user's characteristics/preferences, remembers all settings

4. Interactivity/Interoperability: Does it allow user input, provide feedback, contain prompts (reminders, sharing options, notifications, etc.)? Does the app/e-tool adapt based on user input? Does it allow exchange of data with other apps, e-tools, or wearable devices (if applicable)?

    1 - No interactive and features and/or no response to user input; does not adapt based off user information; has no function for exchange of data with other apps, e-tools, or wearables (if applicable)

    2 - Some, but not enough interactive and/or interoperability features which limits app/e-tool functions, some adaptability

    3 - Basic interactive features to function adequately; has some capacity for exchanging data with other apps, e-tools, or wearables (if applicable)

    4 - Offers a variety of interactive features, feedback and user input options, app/e-tool adapts somewhat to user input; with some effort can exchange data with multiple different apps, e-tools, and wearables (if applicable)

    5 - Very high level of responsiveness through interactive features, feedback, and user input options; app/e-tool adapts as user inputs; allows easy exchange of data with apps, e-tools, or wearable devices (if applicable)

5. Target group: Is the content (visuals, language, design) appropriate for the target audience?

    1 - Completely inappropriate, unclear, or confusing

    2 - Mostly inappropriate, unclear, or confusing

    3 - Acceptable but not specifically designed for the target audience. May be inappropriate/ unclear/confusing at times

    4 - Designed for the target audience, with minor issues

    5 - Designed specifically for the target audience, no issues found

A. **Engagement mean score = _____**

SECTION B

Functionality –app/e-tool functioning, easy to learn, navigation, flow logic, and intuitive design of app/e-tool

6. Performance: How accurately/fast does the app/e-tool run (functions) and do all components with the app/e-tool (buttons/menus) work? Are there any error messages, glitches, crashes?

    1 - App/e-tool is broken; no/insufficient/inaccurate response (e.g., crashes/bugs/broken features, etc.) 2 Some functions work, but lagging or contains major technical problems

    3 - App/e-tool works overall. Some technical problems need fixing, or is slow at times

    4 - Mostly functional with minor/negligible problems

5 - Perfect/timely response; no technical bugs found, or contains a 'loading time left' indicator (if relevant)

7. **Ease of use:** How easy is it to learn how to use the app/e-tool; how clear are the menu labels, icons, and instructions? Is the sign-up process quick and/or simple? Are there relevant help buttons/FAQs?

1 - No/limited instructions; menu labels, icons are confusing; complicated; sign up process is complicated with no help buttons/FAQ's

2 - Takes a lot of time or effort, sign up process is somewhat complicated and/or asks for too much information and/or offers little help

3 - Takes some time or effort

4 - Easy to learn (or has clear instructions); sign up process relatively simple; some help/FAQ's

5 - Able to use app/e-tool immediately; intuitive; simple (no instructions needed); relevant support is obvious and helpful

8. **Navigation:** Does moving between screens make sense; Is it easy to move from one section of the app/e-tool to another? Does the app/e-tool provide all necessary links between screens?

1 - No logical connection between screens at all/navigation is difficult

2 - Understandable after a lot of time/effort

3 - Understandable after some time/effort

4 - Easy to understand/navigate

5 - Perfectly logical, easy, clear, and intuitive screen flow throughout, and/or has shortcuts

9. Design: Are there intuitive popup boxes, videos, animations, audio clips, flash images etc. within the e-tool or are there consistent taps/swipes, pinches/scrolls within the app/e-tool? Are these relevant/accurate/make sense and in theme with the rest of the app/e-tool?

1 - Completely confusing/inconsistent, information lacks relevance or is inaccurate/unnecessary

2 - Often confusing/inconsistent, information of little relevance or contains some unnecessary/incorrect

3 - Okay, some confusing and/or unnecessary information or some inconsistencies

4 - Mostly intuitive, with negligible problems with majority of information is accurate/necessary

5 - Perfectly consistent and intuitive, information is accurate/necessary

B. Functionality mean score = _____

SECTION C
Aesthetics – graphic design, overall visual appeal, color scheme, and stylistic consistency

10. Layout: Is arrangement and size of buttons, icons, menus, and content on the screen appropriate?

    1 - Very bad design, cluttered, some options impossible to select, locate, see, or read

    2 - Bad design, random, unclear, some options difficult to select/locate/see/read

    3 - Satisfactory, few problems with selecting/locating/seeing/reading items

    4 - Mostly clear, able to select/locate/see/read items

    5 - Professional, simple, clear, orderly, logically organized

11. Graphics: How high is the quality/resolution of graphics used for buttons, icons, menus, and content?

    1 - Graphics appear amateur, very poor visual design - disproportionate, stylistically inconsistent

    2 - Low quality/low resolution graphics; low quality visual design – disproportionate

    3 - Moderate quality graphics and visual design (generally consistent in style)

    4 - High quality/resolution graphics and visual design – mostly proportionate, consistent in style

    5 - Very high quality/resolution graphics and visual design - proportionate, consistent in style throughout

12. **Visual appeal:** How good does the app/e-tool look?

    1 - Ugly, unpleasant to look at, poorly designed, clashing, mismatched colors

    2 - Bad – poorly designed, bad use of color, visually boring

    3 - OK – average, neither pleasant, nor unpleasant

    4 - Pleasant – seamless graphics – consistent and professionally designed

    5 - Beautiful – very attractive, memorable, stands out; use of color enhances app/e-tool features/menus

C. Aesthetics mean score = _____

SECTION D

Information – Contains high quality information (e.g., text, feedback, measures, references) from a credible source

13. Goals: Does app/e-tool have specific, measurable, and achievable goals (are these goals specified/obvious within the app/e-tool)?

    N/A - Description does not list goals, or app/e-tool goals are irrelevant to research goal (e.g., using a game for educational purposes)

    1 - App/e-tool has no chance of achieving its stated goals

    2 - Description lists some goals, but app/e-tool has very little chance of achieving them

    3 - OK. App/e-tool has clear goals, which may be achievable.

    4 - App/e-tool has clearly specified goals, which are measurable and achievable

    5 - App/e-tool has specific and measurable goals, which are highly likely to be achieved

14. Quality of information: Is the content within the app/e-tool correct (including description in app store – if an app)? Is app/e-tool up to date with current research, well written, and relevant to the goal/topic of the app/e-tool?

    N/A - There is no information within the app/e-tool

    1 - Irrelevant/inappropriate/incoherent/incorrect

    2 - Poor. Barely relevant/appropriate/coherent/may be incorrect

    3 - Moderately relevant/appropriate/coherent/and appears correct

    4 - Relevant/appropriate/coherent/correct

    5 - Highly relevant, appropriate, coherent, and correct

15. Quantity of information: Is the information within the app/e-tool comprehensive and/or relevant but concise?

    N/A - There is no information within the app/e-tool

    1 - Minimal or overwhelming

    2 - Insufficient or possibly overwhelming

    3 - OK but not comprehensive or concise

    4 - Offers a broad range of information, has some gaps or unnecessary detail; or has no links to more information and resources

    5 - Comprehensive and concise; contains links to more information and resources

16. Visual information: Is visual explanation of concepts – through charts/graphs/images/videos, etc. – clear, logical, correct?

    N/A - There is no visual information within the app/e-tool (e.g., it only contains audio, or text)

    1 - Completely unclear/confusing/wrong or necessary but missing

    2 - Mostly unclear/confusing/wrong

    3 - OK but often unclear/confusing/wrong

    4 - Mostly clear/logical/correct with negligible issues

    5 - Perfectly clear/logical/correct

17. Credibility of source: does the information within the app/e-tool seem to come from a credible source?

    1 - Source identified but legitimacy/trustworthiness of source is questionable (e.g., commercial business with vested interest)

    2 - Appears to come from a legitimate source, but it cannot be verified (e.g., has no webpage)

    3 - Developed by small NGO/institution (hospital/center, etc.) /specialized commercial business, funding body

    4 - Developed by government, university or as above but larger in scale

    5 - Developed using nationally competitive government or research funding (e.g., Australian Research Council, NHMRC)

18. Evidence base: Has the app/e-tool been trialed/tested; must be verified by evidence (in published scientific literature)?

    N/A - It has not been trialed/tested

    1 - The evidence suggests the app/e-tool does not work

2 - App/e-tool has been trialed (e.g., acceptability, usability, satisfaction ratings) and has partially positive outcomes in studies that are not randomized controlled trials (RCTs), or there is little or no contradictory evidence.
3 - App/e-tool has been trialed (e.g., acceptability, usability, satisfaction ratings) and has positive
outcomes in studies that are not RCTs, and there is no contradictory evidence.
4 - App/e-tool has been trialed and outcome tested in 1-2 RCTs indicating positive results
5 - App/e-tool has been trialed and outcome tested in > 3 high quality RCTs indicating positive results

D. Information mean score = _____*
* Exclude questions rated as "N/A" from the mean score calculation

SECTION E
App/e-tool subjective quality rating

19. Would you recommend this app/e-tool to people who might benefit from it?
    1 - Not at all - I would not recommend this app/e-tool to anyone
    2 - There are very few people I would recommend this app/e-tool to
    3 - Maybe - There are several people whom I would recommend it to
    4 - There are many people I would recommend this app/e-tool to
    5 - Definitely - I would recommend this app/e-tool to everyone

20. How many times do you think you would use this app/e-tool in the next 12 months if it was relevant to you?
    1 - None
    2 - 1-2
    3 - 3-10
    4 - 10-50
    5 - >50

21. Would you pay for this app/e-tool?
    1 - No
    2 -
    3 - Maybe
    4 -
    5 - Yes

22. What is your overall star rating of the app/e-tool?
    1 - ★ One of the worst apps/e-tools I've used
    2 - ★★
    3 - ★★★ Average
    4 - ★★★★
    5 - ★★★★★ One of the best apps/e-tools I've used

E. Subjective mean score = _____

SECTION F
Supplement Health Related App/e-tools**:** Questions to consider when using a health-related app/e-tool

23. Additional resources available. Does the app/e-tool provide up to date relevant offline/online resources to support the information presented?
    1 - No – provides no further resources
    2 - Provides few online OR offline resources
    3 - Somewhat – provides some offline and/or online resources, may be outdated
    4 - Provides adequate online and/or offline resources
    5 - Yes – provides abundant and up to date offline and online resources

24. Strategies: Does the app/e-tool recommend strategies that are non-tech based and linked to the problems you have reported?
    1 - No – none
    2 -
    3 - Somewhat – some information; may be too much, or too little resources
    4 -
    5 - Yes – adequate/plentiful but not overbearing

25. Solutions: Does it offer multiple solutions for one issue?
    1 - No – offers one solution to address one issue/health symptom
    2 -
    3 - Some – offers some solutions for the one issue/health symptom; offers solutions but they indirectly address the issue
    4 -
    5 - Yes – offers various related solutions to directly address the issue

26. Multiple health issues/symptoms: Does it address more than one symptom or health issue?
    1 - No – addresses one symptom/health issue only
    2 -
    3 - Some – addresses some symptoms/health issues; or considers many but only partly address them
    4 -
    5 - Yes – considers multiple symptoms/health issues and related ones, and sufficiently addresses them

27. Real time tracking: Can you use the app/e-tool in real time, as you're experiencing a health issue?
    1 - No – the app/e-tool is mainly useful for prevention or recovery
    2 -
    3 - The app/e-tool is useful for prevention, management and/or recovery of the health issue(s)

4 -

5 - Yes – the app/e-tool is useful for prevention, management, and recovery of the health issue(s)

28. Access to help: Easy/obvious to access health related help when needed?
   1 - No – Difficult to navigate or find related health information when needed
   2 - Can find needed information after a lot of time/effort
   3 - Can find needed information after some time/effort
   4 - Easy to understand/navigate needed information
   5 - Perfectly logical, easy, clear, and intuitive screen flow throughout, and/or has shortcuts to needed health information. Offline options are available.

F. Health-related information mean score = _____

Scoring
App/e-tool quality scores:

SECTION
A: Engagement Mean Score = _____
B: Functionality Mean Score = _____
C: Aesthetics Mean Score = _____
D: Information Mean Score = _____
   Quality mean Score = _____

E. Subjective quality Score = _____
F. Health-related quality Score = _____

Appendix B: American Psychiatric Association App Evaluation Model

American Psychiatric Association App Evaluation Model Screener
The below Model is comprehensive. This brief version of the Model extracts a sample of the most fundamental questions that should be asked before considering using an app, and can serve as a good "jumping off" point to get you started:

1. On which platforms/operating systems does the app work? Does it also work on a desktop computer?
2. Has the app been updated in the last 180 days?
3. Is there a transparent privacy policy that is clear and accessible before use?
4. Does the app collect, use, and/or transmit sensitive data? If yes, does it claim to do so securely?
5. Is there evidence of specific benefit from academic institutions, end user feedback, or research studies?
6. Does the app have a clinical/recovery foundation relevant to your intended use?
7. Does the app seem easy to use?
8. Can data be easily shared and interpreted in a way that's consistent with the stated purpose of the app?

The Comprehensive App Evaluation Model

Step 1: Access and Background
The first step of the model is to help ensure that as much useful background information about the app is known **before you evaluate it**. This information helps create a useful context in which you can consider using the app and provides a framework for your decision making as you progress through the model. Thus, the questions below will help you decide whether to proceed with the app evaluation. You do not need to have an answer for each question in order to proceed with evaluating an app.

1. Does the app identify ownership?
2. Does the app identify funding sources and conflicts of interest?
3. Does the app come from a trusted source?
4. Does it claim to be medical?
5. Are there additional or hidden costs?
6. Does the app work offline?
7. On which platforms/operating systems does it work?
   - Does it work on a desktop computer?
8. Does the app work with accessibility features of the iPhone/android?
   - Is it accessible for those with impaired vision or other disabilities?
9. Has the app been updated in the last 180 days?

Step 2: Privacy and Security
While nearly any measurement or intervention contains some risk (e.g., physical, psychological, legal, social, and economic), apps present some unique risks that may often be overlooked. Risks may include data costs associated with app use (i.e., depending on your contracted data plan with your wireless provider), social profiling, loss of insurance benefits or insurability—all of which are associated with privacy and

security. Digital privacy and security are not often high-level risk factors when prescribing a medication or conducting in-person therapy; when deciding on whether to use an app, however, they are extremely important and should be the first area evaluated.

The questions below are intended to help you consider many aspects of app security and privacy. Note that they are not all-inclusive, as there is currently no "gold standard" for rating apps' privacy and security. Many of your answers to these questions should be found in the app's privacy policy, or perhaps on the app developer's web site. If there is no privacy policy that you could readily find, you should consider whether the app is appropriate for collecting sensitive personal health information.

For certain questions, like what security measures are in place, it is necessary to take the app's description at face value at this time. There is no cut-off or score for this level of the model; instead, you and the patient will need to decide if—based on the answers to these questions—you feel the app meets your standards. However. If they are not addressed or are addressed in a way suggesting that patient privacy and security may be compromised, you should consider whether this is appropriate for collecting sensitive personal health information. The ultimate goal of this level is to ensure an app will not cause harm by violating patient safety, security, and privacy and that you and the patient understand the scope and limitations of privacy and security.

1. Is there a transparent privacy policy that is clear and accessible before use?
2. Does the app declare data use and purpose?
3. Does the app describe use of PHI?
   o Deidentified vs. anonymous?
4. Can you opt out of data collection or delete data?
5. Are data maintained in the device or on the web?
6. Does the app explain security systems used?
7. Does the app collect, use, and/or transmit sensitive data? If yes, does it claim to do so securely?
8. What third parties does the app share data with?
9. If appropriate, is the app equipped to respond to potential harms or safety concerns?

Step 3: Clinical Foundation

App developers often make many claims on their apps' clinical effectiveness or background—though there is often little data to support the claims. This does not mean that apps don't work, but rather that there is much we still do not know. If you decide that an app has sufficient privacy and security at Level 2, then your task at Level 3 is to evaluate any evidence for potential benefits.

While some apps' benefits have been documented in clinical studies, many—if not most—have not. Because of this, we recommend that you download and try the app to see what it is actually doing and if the content and information it offers appear at least reasonable and not harmful (i.e., evidence of "face validity"). Again, few apps will have a gold standard, randomized double blinded placebo controlled study to suggest they are

effective, so the questions presented below are designed to help you think of other ways you can make an informed decision about an app's evidence base.
1. Does the app appear to do what it claims to do?
2. Is the app content correct, well-written, and relevant?
3. What are the relevant sources or references supporting the app use cases?
4. Is there evidence of specific benefit from academic institutions, publications, end user feedback, or research studies?
5. Is there evidence of effectiveness/efficacy?
6. Was there an attempt to validate app usability and feasibility?
7. Does the app have a clinical/recovery foundation relevant to your intended use?

Step 4: Usability

To recap, if an app has satisfied criteria in Steps One and Two within this Model, then you may assume that:
1. It offers minimal risk in terms of digital safety and privacy.
2. It appears to have some benefit.

Thus, Step 4 helps you to evaluate usability, because an app is only as useful as determined by you and your patients, after taking it for a spin. Because of this, Usability is a more subjective category and so different people will have very different ideas about what this means to them. The questions below are, again, designed to help you think about the app's interface and overall functionality and then make an informed decision about how usable an app will be for the case and patient at hand.
1. What are the main engagement styles of the app?
2. Do the app and its features align with your needs and priorities?
3. Is it customizable?
4. Does the app clearly define functional scope?
5. Does the app seem easy to use?

Step 5: Data Integration towards Therapeutic Goal

Finally, the last step in the model is Data Integration towards Therapeutic Goal. This is the topmost level, as the ability to share data may only matter if this is an app that you and the patient want to use (based on background information in Step 1); if it is safe and secure (Step 2); has some evidence base (Step 3), and is easy to use (Step 4). The reason why data integration with the patient's therapeutic goal becomes important in this model, is because apps should not fragment care and the patient and psychiatrist should be able to share and discuss data or retrieve feedback from the app as appropriate.

In some cases, the ability for apps to share data may not be relevant. For other apps, however (e.g., mood trackers and medication management), ensuring that such data can be easily shared and accessed by those who need to see it is an important factor to consider. While the specifics of this level in the Model will vary for each patient (e.g., with respect to the devices they use, and your medical record system) the following questions can help you to think about whether an app's data or other output can be used in a clinically meaningful way.

1. Do you own your data?
2. Can data be easily shared and interpreted in a way that's consistent with the stated purpose of the app?
3. Can the app share data with EMR and other data tools (apple HealthKit, Fitbit)?
4. Is the app for individual use or to be used in collaboration with a provider?
5. If intended to be used with a provider, does the app have the ability to export or transfer data?
6. Does the app lead to any positive behavior change or skill acquisition?
7. Does the app improve therapeutic alliance between patient and provider?

Appendix C: THESIS

THESIS Detailed app rating criteria

Transparency
- Cost of app (purchase price, subscriptions, in-app purchases)
  - Are the prices, subscriptions, and in-app purchases accurately conveyed?
- Consent
  - What is the quality of the consent process, if any?
- Accuracy of app store description
  - How accurate is the app store description of the app's purpose?

Health content
- Appropriate measurement
  - Does the app appropriately measure what it claims to measure?
- Appropriate interpretation of data
  - Does the app appropriately interpret what it claims to interpret?
- Quality of information
  - How optimal is the quality of information?
- Potential for harm
  - Is the potential for harm minimized?
- Literacy level
  - How appropriate is the literacy level for the app's intended audience?
- Presentation of information
  - Is information presented in an optimal manner? For example, is scaffolding used?

Technical content
- Software performance/stability
  - Does the app run well with zero interface crashes or bugs?
- Interoperability
  - Is the app able to exchange information with EHRs and other apps?
- Bandwidth
  - Does the app require significant bandwidth to run?
    5: App does not require the use of cellular data; very few graphics used in the app
    4: Main function of the app doesn't require the internet or cellular data, however there is a decent number of images or animations used in the app
    3: App uses large images and animations, and only a few of its functions require the use of internet or cellular data
    2: Many of the app's function use the internet or need cellular service, however the user can use the app offline
    1: The main functions of the app require significant use of cellular data, access to the internet and/or location services; app includes a large amount of images, animations, and/or videos. User cannot use app without internet (no offline version).

- Application size
  - Does the app require significant storage capacity?
    5: <10 MB
    4: between 10 to 20 MB
    3: between 20-30 MB
    2: between 30-40 MB
    1: >40 MB

Security/privacy
- Protection against theft, viruses, etc.
  - Does the app follow best practices in security with optimal anti-virus and safeguards against breaches?

- Authentication
  - Is the authentication procedure optimal?

- Data sharing
  - When sharing information, does the app use best practices?

- Maintenance
  - Does the app have regular cycles to update and patch its security?
    5: last update occurred during the month of rating or during the month before; also, if the update schedule is very consistent
    4: last update occurred 2 months before the time of rating; update schedule is generally consistent
    3: last update occurred between 3-5 months before the time of rating; update schedule is a little inconsistent
    2: last update occurred between 6 months to a year before time of rating; update schedule is completely inconsistent
    1: last update was occurred more than 1 year ago
- Signaling of breaches
  - If a breach occurs, does the app have a method to notify its users?
- Anonymization
  - Does the app appropriately anonymize individuals?

Usability
- Installation and setup
  - How would you rate are installation and setup?
- Functionality: ease of use, navigation, gestural design, help/instructions
  - Quality of ease of use, navigation, gestural design, help/instructions?
- Aesthetics: layout, graphics, visual appeal, image readability
  - Quality of layout, graphics, visual appeal, and image readability?
- Customization/tailoring
  - Ability to customize and tailor to the specific user's needs?
- Ease of use for users with low literacy and numeracy
  - Is the app usable by users with low literacy and numeracy?

5: The medical information provided by the app does not use a significant amount of medical terms; complex conditions are explained using laymen's terms. The medical information is also supplemented with simple images or short animations. Lastly, if the app allows users to input their symptoms, they can use laymen's terms, instead using medical terms, and the app will output the possible conditions the patient may have.

4: Overall, the medical information in the app is sufficiently explained using laymen's terms, and any graphics used to supplement this information is simple and clear. There are very few instances where medical terms are not explained.

3: Some of the medical information in the app uses medical terms that are not sufficiently explained or supplemented by diagrams or images. Not all the information is complex, patient or their caregiver can still navigate through the app.

2: Very few laymen's terms are used to describe medical conditions. Medical terminology is not explained. Any images or graphics meant to help explain medical conditions require some medical familiarity.

1: The language used in the app is complex, and very difficult to understand if the user does not have any prior medical knowledge. The description of medical conditions does not use any laymen's terms. No resources such as images or graphics to help explain medical conditions.

- Available in multiple languages
  - Is the app available in multiple languages?
    5: >16
    4: 12-16
    3: 8-12
    2: 4-8
    1: <4

Subjective rating
- Recommend app
  - Would you recommend this app?

- Overall star rating
  - What is your overall rating of this app?

Appendix D: ABACUS

App Behavior Change Scale

1. Knowledge and information

1.1 Does the app have the ability to customize and personalize some features?
Definition: Elements of the app can be personalized through specific tools or functions that are specific to the individual using the app.

1.2 Was the app created with expertise and/or does the app provide information that is consistent with national guidelines?
Definition: This would be found in the about section or generally in the app.

1.3 Does the app ask for baseline information?
Definition: This includes BMI, weight, smoking rate, exercise, or drinking behaviors

1.4 Does the app provide instruction on how to perform the behavior?
Definition: The app is clear in telling the person how to perform a behavior or preparatory behaviors, either verbally, through video, or in written form.

1.5 Does the app provide information about the consequences of continuing and/or discontinuing behavior?
Definition: The app gives the user information about the consequences of behavior in general, this includes information about the relationship between the behavior and its possible or likely consequences in the general case. This information can be general or personalized.

2. Goals and planning

2.1 Does the app ask for willingness for behavior change?
Definition: Is there a feature during setup where you describe how ready you are for behavior change?

2.2 Does the app allow for the setting of goals?
Definition: The person is encouraged to make a behavioral resolution. The person is encouraged to set a general goal that can be achieved by behavioral means. This includes subgoals or preparatory behaviors and/or specific contexts in which the behavior will be performed. The behavior in this technique will be directly related to or be a necessary condition for the target behavior.

2.3 Does the app have the ability to review goals, update, and change when necessary?
Definition: Involves a review or analysis of the extent to which previously set behavioral goals (regardless of short or long) were achieved.

3. Feedback and monitoring

3.1 Does the app give the user the ability to quickly and easily understand the difference between current action and future goals?
Definition: Allows user to see how they are tracking against a goal and to see the difference between what they want to do and what they are currently doing. This will give some feedback on where they are at and what they need to change to get to where they want to be.

3.2 Does the app have the ability to allow the user to easily self-monitor behavior?
Definition: The app allows for a regular monitoring of the activity.

3.3 Does the app have the ability to share behaviors with others (including social media or forums) and/or allow for social comparison?
Definition: The app allows the person to share his or her behaviors on social media or in forums. This could also include a buddy system or a leaderboard.

3.4 Does the app have the ability to give the user feedback—either from a person or automatically?
Definition: The app is able to provide the person with feedback, comments, or data about their own recorded behavior. This might be automatic or could be personal.

3.5 Does the app have the ability to export data from app?
Definition: The app allows for the export of information and progress to an external user.

3.6 Does the app provide a material or social reward or incentive?
Definition: App provides rewards for attempts at achieving a behavioral goal. This might include efforts made toward achieving the behavior or progress made in preparatory steps toward the behavior or in achieving a goal.

3.7 Does the app provide general encouragement?
Definition: The app provides general encouragement and positive reinforcement on actions leading to the goal.

4. Actions

4.1 Does the app have reminders and/or prompts or cues for activity?
Definition: The app prompts the user to engage in the activity. The app has the ability to give notifications or reminders to cue the behavior.

4.2 Does the app encourage positive habit formation?
Definition: The app prompts explicit rehearsal and repetition of the behavior–not just tracking or logging.

4.3 Does the app allow or encourage for practice or rehearsal, in addition to daily activities?
Definition: App does not have a lock on activities or a number that you cannot exceed daily.

4.4 Does the app provide opportunity to plan for barriers?
Definition: The app encourages the person to think about potential barriers and identify ways of overcoming them.

4.5 Does the app assist with or suggest restructuring the physical or social environment?
Definition: The app prompts the person to alter the environment in ways so that it is more supportive of the target behavior.

4.6 Does the app assists with distraction or avoidance?
Definition: The app gives suggestions and advice on how the person can avoid situations or distract themselves when trying to reach their goal.

APPENDIX E: MIND

M-Health Index and Navigation Database Questions

App Origin
Who is the Developer?
- Does it come from the government?
- Does it come from a for-profit company or developer?
- Does it come from a non-profit company?
- Does it come from a trusted healthcare company?
- Does it come from an academic institution?

App Functionality
App Store Attributes
- Does it work on Apple (iOS)?
- What is the Apple version?
- What is the oldest iOS version supported?
- What was the Apple release date?
- Has the Apple version been updated in the last 180 days?
- Number of reviews on Apple store?
- Rating (number of stars) on Apple store?
- App size on iOS?
- Does it work on Android?
- What is the Android version?
- What is the oldest Android version supported?
- What was the Google Play store release date?
- When was the last Android update?
- Has the android version been updated in the last 180 days?
- Number of reviews on Google Play store?
- Rating (number of stars) on Google Play store?
- App size on Android?

Accessibility
- Does the app work offline?
- Does it have at least one accessibility feature (like adjust text size, text to voice, or colorblind color scheme adjuster)?
- Does it work with Spanish?
- Does it work with a language other than English or Spanish?
- Is the app totally free?
- What is the cost up front?
- Are there in-app purchases?
- Is it a subscription (recurrent/monthly/annually)?

Privacy & Security
- Is there a privacy policy?

- Does the app declare data use and purpose?
- Does the app report security measures in place?
- Is PHI shared?
- Is de-identified data shared?
- Is anonymized/aggregate data shared?
- Can you opt out of data collection?
- Can you delete your data?
- Is the user data stored only on the device?
- Is the user data stored on a server?
- Does the app have a crisis management feature?
- Does the app claim it meets HIPAA?
- Reading level of the privacy policy (what grade reading level)?
  - https://readabilityformulas.com/free-readability-formula-tests.php (copy and paste privacy policy in)
- Does the app use 3rd party vendors (i.e., Google analytics, etc.)

App Inputs
- Input: surveys – Does the app enable a user to enter surveys such as mood or symptom surveys?
- Input: diary – Does the app have a journaling, diary, or free writing feature?
- Input: geolocation – Does the app enable location services from the phone?
- Input: camera – Do any features of the app utilize camera input? So profile picture? Or photo diary feature? Or video chat?
- Input: microphone – Does the app allow a user to record using the phone microphone?
- Input: external devices (e.g., a wearable sending direct data) – Does the app connect with an external device such as a smart watch or heart rate monitor?
- Input: social network – Connection to social media. Does the app allow you to input social media information? For example, do you connect it to your Facebook to log in? Or do you connect with social media contacts through the app?

App Outputs
- Output: notifications – Does the app send notifications? These notifications could be incoming messages, reminders from the app, or alerts.
- Output: psychoeducational references/information – Does the app provide psychoeducational references or information? (Note: this means the exact same thing as the question about features: psychoeducation)
- Output: social network – Can you post information from the app to social media? Does the app connect to social media for posting purposes?
- Output: reminders – Does the app allow you to set reminders? (Oftentimes these reminders will then generate notifications)
- Output: graphs of data – Does the app allow a user to see graphically depicted data?
- Output: summary of data (in text or numbers) – Does the app provide written summaries of data (description of data apart from a graph)?

- Output: link to formal care/coaching – Does the app connect a user with a healthcare provider? A licensed therapist or clinician?

Clinical Foundation
- Does the app appear to do what it claims to do?
- Is the app patient facing?
- How many feasibility/usability studies?
- What is the highest feasibility impact factor?
- How many evidence/efficiency studies?
- What is the highest efficacy impact factor?
- Can the app cause harm?
- Does the app provide any warning for use?

App Features
- Features: mood tracking – Does the app provide surveys where a user can enter their mood data?
- Features: medication tracking – Designated medication tracking feature?
- Features: sleep tracking – Does the app track sleep, either in conjunction with a wearable or through user-entered information?
- Features: physical exercise tracking – Does it allow a user to track duration or content of physical exercise?
- Features: psychoeducation – Does it provide definitions, explanations, or different diagnoses? Is it didactic?
- Features: journaling – Is there a place for the user to journal or free write?
- Features: picture gallery/hope board – Does the app allow a user to curate a gallery of saved and searched images and quotes?
- Features: mindfulness – Any mindfulness exercises? May include deep breathing, but not necessarily.
- Features: deep breathing – Does the app offer exercises in deep breathing?
- Features: iCBT or sleep therapy – Does the app offer sleep therapy of any kind (including iCBT, a targeted sleep intervention)?
- Features: CBT – Does the app provide cognitive-behavioral therapy?
- Features: ACT – Does the app provide Acceptance and Commitment Therapy?
- Features: DBT – Does the app provide dialectical behavior therapy?
- Features: peer support – Does the app offer connection to peer specialists or individuals with lived experience?
- Features: connection to coach/therapist – The app has a built-in way to connect with a provider or coach.
- Features: biodata – Does the app collect heart rate or skin conductance?
- Features: goal setting/habits – Productivity feature allowing user to set and check in on goals.
- Features: physical health exercises – Something like 7 minute workout that actually gives a workout (this is a recommendation of exercises, NOT tracking).
- Features: Bbot interaction (like with virtual character) – The app allows a user to interact with a virtual character.

- Features: biofeedback with sense data (EEG, HRB, skin conductance, etc.) – The app uses biodata to provide feedback/recommendations (an app that will recommend more breathing exercises to respond to high heart rate, for example).

App Engagement Style
- Engagement style: chat/message based – User can send and receive messages
- Engagement style: is it a screener/assessment – Examples include PHQ9, GAS7, etc.
- Engagement style: real time response – Someone will reply to your chat right away
- Engagement style: asynchronous response – There are no immediate responses to chats, responses come at predetermined intervals (once a day, every four hours, etc.)
- Engagement style: gamification (points, badges) – User can win points and prizes for engaging with the app
- Engagement style: videos – App includes videos user can view
- Engagement style: audio/music/scripts – Does the app provide music or audio experiences? Some meditation apps, for example, utilize audio sessions
- Engagement style: AI support – Interaction is not with a real person but with a bot
- Engagement style: peer support – Peer is defined as a person with lived experience and support involves actually communicating (so not just watching a video)
- Engagement style: network support – Network is defined as someone (like family or friend) who is actually known by the user outside the app. An example is an app that allows a user to communicate with family members about relevant health information.
Engagement style: collaborative with provider/other – Does it allow for direct collaboration with a provider or clinician? Beyond just being able to share your data.

Interoperability and Data Sharing
App Use
- Is it a self-help/self-management tool?
- Is it a reference app?
- Is it intended for hybrid use with a clinician in conjunction with treatment plan?
Interoperability and Data Sharing
- Do you own your data?
- Can you email or export your data?
- Can you send your data to a medical record?

One Mind PsyberGuide App Scoring Criteria

Credibility

- Consumer Ratings
  Possible Points: 2.
  Note: Ratings may come from app store or other sources. Total number counted must be from a single source
  0 – Fewer than 30 user rating OR an average rating below 3.5
  1 – Ratings exist from 31-1500 users with an average rating of 3.5+
  2 – Ratings exist from >1500 users with an average rating of 3.5+
- Proposed Goal
  Possible Points: 2.
  0 – No clear goals
  1 – Product describes non-specific or hard to measure mental health goals (e.g., improve your life, improve your wellbeing)
  2 – Product describes at least one mental health goal which is specific, measurable, and achievable (e.g., reduce stress, reduce symptoms of PTSD)
- Evidence-Based Content
  Possible Points: 1.
  0 – The app does not use evidence-based practices to achieve its goals (or there are no goals described)
  1 – The app uses evidence-based practices to achieve its goals
- Research Base
  Possible Points: 3.
  0 – No research
  1 – Other research (e.g., single case designs, quasi-experimental methods demonstrating efficacy, or preliminary analyses)
  2 – Some research support for the product (at least one experiment that shows efficacy or effectiveness)
  3 – Strong research support for the product (at least two between-group design experiments that show efficacy or effectiveness)
- Software Updates
  Possible Points: 2.
  0 – The application has not been revised or was revised more than 12 months ago
  1 – The application has been revised within the last 12 months
  2 – The application has been revised within the last 6 months
- Clinical Input in Development
  Possible Points: 1.
  0 – No clinical leader with mental health expertise involved in development
  1 – Clinical leader with mental health expertise involved in development
- Research on Development Process
  Possible Points: 1.

0 – No pilot, feasibility and acceptability data AND no evidence of stakeholder engagement
1 – Pilot, feasibility and acceptability data OR evidence of stakeholder engagement in development

- Efficacy of Other Products
Possible Points: 1.
0 – No other mental health technological interventions demonstrating efficacy have been developed by this team
1 – Developer/development team has developed other mental health interventions delivered via technological medium which demonstrate efficacy

- Research Independence & Review
Possible Points: 2.
0 – No information about source of funding for the research AND No published, peer-reviewed papers
1 – All research funded primarily by for-profit organizations or combined funding sources OR one article published in a peer-reviewed journal
2 – At least one research paper funded by government agency (e.g., NIH) or non-profit organization OR two articles published in peer-reviewed journals

User Experience
Mobile Application Rating Scale (MARS)

- App Classification
The Classification section is used to collect descriptive and technical information about the app. Please review the app description in iTunes / Google Play to access this information.

App Name: _____
Rating this version: _____
Rating all versions: _____
Developer: _____
N ratings this version: _____
N ratings all versions: _____
Version: _____
Last update: _____
Cost - basic version: _____
Cost - upgrade version: _____
Platform: iPhone   iPad   Android

Brief description:
_____
_____

- Focus: what the app targets (select all that apply)

Increase Happiness/Well-being	Mindfulness/Meditation/Relaxation
Reduce negative emotions	Depression
Anxiety/Stress	Anger

Behavior Change	Alcohol /Substance Use
Goal Setting	Entertainment
Relationships	Physical health
Other _____	

- Theoretical background/Strategies (all that apply)

Assessment	Feedback
Information/Education	Monitoring/Tracking
Goal setting	Advice /Tips /Strategies /Skills training
CBT - Behavioral (positive events)	CBT – Cognitive (thought challenging)
ACT - Acceptance commitment therapy	Mindfulness/Meditation
Relaxation	Gratitude
Strengths based	Other _____

- Affiliations:

Unknown   Commercial   Government   NGO   University

- Age group (all that apply)

Children (under 12)	Adolescents (13-17)
Young Adults (18-25)	Adults
General	

- Technical aspects of app (all that apply)

Allows sharing (Facebook, Twitter, etc.)	Has an app community
Allows password-protection	Requires login
Sends reminders	Needs web access to function

APP QUALITY RATINGS

The Rating scale assesses app quality on four dimensions. All items are rated on a 5-point scale from "1. Inadequate" to "5. Excellent". Circle the number that most accurately represents the quality of the app component you are rating. Please use the descriptors provided for each response category.

- SECTION A - Engagement – fun, interesting, customizable, interactive (e.g., sends alerts, messages, reminders, feedback, enables sharing), well-targeted to audience

  1. Entertainment: Is the app fun/entertaining to use? Does it use any strategies to increase engagement through entertainment (e.g., through gamification)?
      1 – Dull, not fun or entertaining at all
      2 – Mostly boring
      3 – OK, fun enough to entertain user for a brief time (< 5 minutes)
      4 – Moderately fun and entertaining, would entertain user for some time (5-10 minutes total)
      5 – Highly entertaining and fun, would stimulate repeat use

2. Interest: Is the app interesting to use? Does it use any strategies to increase engagement by presenting its content in an interesting way?

    1 – Not interesting at all

    2 – Mostly uninteresting

    3 – OK, neither interesting nor uninteresting; would engage user for a brief time (< 5 minutes)

    4 – Moderately interesting; would engage user for some time (5-10 minutes total)

    5 – Very interesting, would engage user in repeat use

3. Customization: Does it provide/retain all necessary settings/preferences for apps features (e.g., sound, content, notifications, etc.)?

    1 – Does not allow any customization or requires setting to be input every time

    2 – Allows insufficient customization limiting functions

    3 – Allows basic customization to function adequately

    4 – Allows numerous options for customization

    5 – Allows complete tailoring to the individual's characteristics/preferences, retains all settings

4. Interactivity: Does it allow user input, provide feedback, contain prompts (reminders, sharing options, notifications, etc.)? Note: these functions need to be customizable and not overwhelming in order to be perfect.

    1 – No interactive features and/or no response to user interaction

    2 – Insufficient interactivity, or feedback, or user input options, limiting functions

    3 – Basic interactive features to function adequately

    4 – Offers a variety of interactive features/feedback/user input options

    5 – Very high level of responsiveness through interactive features/feedback/user input options

5. Target group: Is the app content (visual information, language, design) appropriate for your target audience?

    1 – Completely inappropriate/unclear/confusing

    2 – Mostly inappropriate/unclear/confusing

    3 – Acceptable but not targeted. May be inappropriate/unclear/confusing

    4 – Well-targeted, with negligible issues

    5 – Perfectly targeted, no issues found

A. Engagement mean score = _____

- SECTION B – Functionality – app functioning, easy to learn, navigation, flow logic, and gestural design of app

6. Performance: How accurately/fast do the app features (functions) and components (buttons/menus) work?

    1 – App is broken; no/insufficient/inaccurate response (e.g., crashes/bugs/broken features, etc.)

    2 – Some functions work, but lagging or contains major technical problems

    3 – App works overall. Some technical problems need fixing/Slow at times

    4 – Mostly functional with minor/negligible problems

5 – Perfect/timely response; no technical bugs found/contains a 'loading time left' indicator

7. Ease of use: How easy is it to learn how to use the app; how clear are the menu labels/icons and instructions?
    1 – No/limited instructions; menu labels/icons are confusing; complicated
    2 – Useable after a lot of time/effort
    3 – Useable after some time/effort
    4 – Easy to learn how to use the app (or has clear instructions)
    5 – Able to use app immediately; intuitive; simple

8. Navigation: Is moving between screens logical/accurate/appropriate/ uninterrupted; are all necessary screen links present?
    1 – Different sections within the app seem logically disconnected and random/confusing/navigation is difficult
    2 – Usable after a lot of time/effort
    3 – Usable after some time/effort
    4 – Easy to use or missing a negligible link
    5 – Perfectly logical, easy, clear, and intuitive screen flow throughout, or offers shortcuts

9. Gestural design: Are interactions (taps/swipes/pinches/scrolls) consistent and intuitive across all components/screens?
    1 – Completely inconsistent/confusing
    2 – Often inconsistent/confusing
    3 – OK with some inconsistencies/confusing elements
    4 – Mostly consistent/intuitive with negligible problems
    5 – Perfectly consistent and intuitive

B. Functionality mean score = _____

- SECTION C – Aesthetics – graphic design, overall visual appeal, color scheme, and stylistic consistency

10. Layout: Is arrangement and size of buttons/icons/menus/content on the screen appropriate or zoomable if needed?
    1 – Very bad design, cluttered, some options impossible to select/locate/see/read device display not optimized
    2 – Bad design, random, unclear, some options difficult to select/locate/see/read
    3 – Satisfactory, few problems with selecting/locating/seeing/reading items or with minor screen size problems
    4 – Mostly clear, able to select/locate/see/read items
    5 – Professional, simple, clear, orderly, logically organized, device display optimized. Every design component has a purpose

11. Graphics: How high is the quality/resolution of graphics used for buttons/icons/menus/content?
    1 – Graphics appear amateur, very poor visual design - disproportionate, completely stylistically inconsistent
    2 – Low quality/low resolution graphics; low quality visual design – disproportionate, stylistically inconsistent

3 – Moderate quality graphics and visual design (generally consistent in style)
4 – High quality/resolution graphics and visual design – mostly proportionate, stylistically consistent
5 – Very high quality/resolution graphics and visual design - proportionate, stylistically consistent throughout

12. Visual appeal: How good does the app look?
1 – No visual appeal, unpleasant to look at, poorly designed, clashing/mismatched colors
2 – Little visual appeal – poorly designed, bad use of color, visually boring
3 – Some visual appeal – average, neither pleasant, nor unpleasant
4 – High level of visual appeal – seamless graphics – consistent and professionally designed
5 – As above + very attractive, memorable, stands out; use of color enhances app features/menus

C. Aesthetics mean score = _____

- SECTION D – Information – Contains high quality information (e.g., text, feedback, measures, references) from a credible source. Select N/A if the app component is irrelevant.

13. Accuracy of app description (in app store): Does app contain what is described?
1 – Misleading. App does not contain the described components/functions. Or has no description
2 – Inaccurate. App contains very few of the described components/functions
3 – OK. App contains some of the described components/functions
4 – Accurate. App contains most of the described components/functions
5 – Highly accurate description of the app components/functions

14. Goals: Does app have specific, measurable, and achievable goals (specified in app store description or within the app itself)?
N/A – Description does not list goals, or app goals are irrelevant to research goal (e.g., using a game for educational purposes)
1 – App has no chance of achieving its stated goals
2 – Description lists some goals, but app has very little chance of achieving them
3 – OK. App has clear goals, which may be achievable.
4 – App has clearly specified goals, which are measurable and achievable
5 – App has specific and measurable goals, which are highly likely to be achieved

15. Quality of information: Is app content correct, well written, and relevant to the goal/topic of the app?
N/A – There is no information within the app
1 – Irrelevant/inappropriate/incoherent/incorrect
2 – Poor. Barely relevant/appropriate/coherent/may be incorrect
3 – Moderately relevant/appropriate/coherent/and appears correct
4 – Relevant/appropriate/coherent/correct

5 – Highly relevant, appropriate, coherent, and correct

16. Quantity of information: Is the extent coverage within the scope of the app; and comprehensive but concise?

 N/A – There is no information within the app
 1 – Minimal or overwhelming
 2 – Insufficient or possibly overwhelming
 3 – OK but not comprehensive or concise
 4 – Offers a broad range of information, has some gaps or unnecessary detail; or has no links to more information and resources
 5 – Comprehensive and concise; contains links to more information and resources

17. Visual information: Is visual explanation of concepts – through charts/graphs/images/videos, etc. – clear, logical, correct?

 N/A – There is no visual information within the app (e.g., it only contains audio, or text)
 1 – Completely unclear/confusing/wrong or necessary but missing
 2 – Mostly unclear/confusing/wrong
 3 – OK but often unclear/confusing/wrong
 4 – Mostly clear/logical/correct with negligible issues
 5 – Perfectly clear/logical/correct

18. Credibility: Does the app come from a legitimate source (specified in app store description or within the app itself)?

 1 – Source identified but legitimacy/trustworthiness of source is questionable (e.g., commercial business with vested interest)
 2 – Appears to come from a legitimate source, but it cannot be verified (e.g., has no webpage)
 3 – Developed by small NGO/institution (hospital/center, etc.) /specialized commercial business, funding body
 4 – Developed by government, university or as above but larger in scale
 5 – Developed using nationally competitive government or research funding (e.g., Australian Research Council, NHMRC)

19. Evidence base: Has the app been trialed/tested; must be verified by evidence (in published scientific literature)?

 N/A – The app has not been trialed/tested
 1 – The evidence suggests the app does not work
 2 – App has been trialed (e.g., acceptability, usability, satisfaction ratings) and has partially positive
outcomes in studies that are not randomized controlled trials (RCTs), or there is little or no contradictory evidence.
 3 – App has been trialed (e.g., acceptability, usability, satisfaction ratings) and has positive outcomes in studies that are not RCTs, and there is no contradictory evidence.
 4 – App has been trialed and outcome tested in 1-2 RCTs indicating positive results
 5 – App has been trialed and outcome tested in > 3 high quality RCTs indicating positive results

D. Information mean score = _____ *
* Exclude questions rated as "N/A" from the mean score calculation.

<u>APP SUBJECTIVE QUALITY</u>
- SECTION E
  20. Would you recommend this app to people who might benefit from it?
     - 1 – Not at all I would not recommend this app to anyone
     - 2 – There are very few people I would recommend this app to
     - 3 – Maybe There are several people whom I would recommend it to
     - 4 – There are many people I would recommend this app to
     - 5 – Definitely I would recommend this app to everyone
  21. How many times do you think you would use this app in the next 12 months if it was relevant to you?
     - 1 – None
     - 2 – 1-2
     - 3 – 3-10
     - 4 – 10-50
     - 5 – >50
  22. Would you pay for this app?
     - 1 – No
     - 2 –
     - 3 – Maybe
     - 4 –
     - 5 – Yes
  23. What is your overall star rating of the app?
     - 1 – «   One of the worst apps I've used
     - 2 – « «
     - 3 – « « «   Average
     - 4 – « « « «
     - 5 – « « « « «   One of the best apps I've used

Scoring App quality scores for SECTION
   A: Engagement Mean Score = _____
   B: Functionality Mean Score = _____
   C: Aesthetics Mean Score = _____
   D: Information Mean Score = _____
   App quality mean Score = _____
   App subjective quality Score = _____

<u>APP-SPECIFIC</u>
These added items can be adjusted and used to assess the perceived impact of the app on the user's knowledge, attitudes, intentions to change as well as the likelihood of actual change in the target health behavior.

- SECTION F

1. Awareness: This app is likely to increase awareness of the importance of addressing [insert target health behavior]
   1– Strongly disagree
   2
   3
   4
   5 – Strongly Agree
2. Knowledge: This app is likely to increase knowledge/understanding of [insert target health behavior]
   1– Strongly disagree
   2
   3
   4
   5 – Strongly Agree
3. Attitudes: This app is likely to change attitudes toward improving [insert target health behavior]
   1– Strongly disagree
   2
   3
   4
   5 – Strongly Agree
4. Intention to change: This app is likely to increase intentions/motivation to address [insert target health behavior]
   1– Strongly disagree
   2
   3
   4
   5 – Strongly Agree
5. Help seeking: Use of this app is likely to encourage further help seeking for [insert target health behavior] (if it's required)
   1– Strongly disagree
   2
   3
   4
   5 – Strongly Agree
6. Behavior change: Use of this app is likely increase/decrease [insert target health behavior]
   1– Strongly disagree
   2
   3
   4
   5 – Strongly Agree

Transparency
- Rating: Acceptable

- o A product that has been scored as acceptable has an acceptable level of data transparency; the privacy policy of the product provides sufficient and easily accessible information on the policies related to data collection, storage, and exchange. The information provided conforms to standards for collection, storage, and exchange of health information.
- Rating: Questionable
  - o A product that has been scored as questionable has a privacy policy that is unclear or lacking specific details of policies surrounding data collection, storage, and exchange or is questionable in its adherence to standards on collection, storage, and exchange of health information.
- Rating: Unacceptable
  - o A product that has been scored as unacceptable either a) does not have a privacy policy, b) has a privacy policy that excludes important information about data privacy, collection, storage, or exchange, or c) has a privacy policy that outlines practices for data privacy, collection, storage, or exchange that do not conform to standards for health information.