

Designing an Effective and Usable Cognitive Behavioral Therapy (CBT) App

by

Jessica Hatfield

May 2022

Presented to the

Division of Science, Information Arts, and Technologies
University of Baltimore

In Partial Fulfillment
of the Requirements for the Degree of
Master of Science

Approved by: _____
[BrIDGET Blodgett, Thesis Advisor]

[Kathryn Summers, Committee Member]

Abstract

Cognitive behavioral therapy (CBT) mobile apps have risen in popularity; however, most are either hard-to-use or not clinically valid. To address this issue, researchers have published CBT app reviews, case studies, and guidelines to inform the development of an effective and usable CBT app. In this paper, I review CBT app literature and use the literature, along with mobile design best practices and interaction design principles, as the basis for an effective and usable CBT app design. To verify the design's usability, I conducted usability tests with a prototype of the app. Based on the results of the usability tests, I can conclude that my CBT app design is more usable than the average user interface. Additionally, study participants found the app's CBT exercise valuable, and completing the exercise was associated with an improvement in mood. However, common pain points experienced by participants during the usability tests warrant design changes, and as app development continues, I will continue to collaborate with clinicians, test with end-users, and iterate on the app's design. Overall, the positive results of this study are exciting because of the potential benefits that an effective and usable CBT app could bring to people's lives.

Acknowledgments

First, I'd like to thank my thesis advisor, Bridget Blodgett, and second reader, John Donahue, for their guidance during this process. Their feedback and suggestions helped me refine, evolve, and improve my thesis, and I'm appreciative of their time and effort.

Additionally, I'd like to thank my family and friends for motivating and encouraging me, acting as a sounding board, testing prototypes, and proofreading paragraphs for the better part of a year. Completing my thesis would have been a lot harder without them, and I'm grateful for their love and support.

Table of Contents

List of Tables	iii
List of Figures	iv
Chapter 1: Introduction	1
Chapter 2: Literature Review	2
CBT App Benefits	2
Apps vs. Websites	2
Popularity	2
Efficacy	2
CBT App Challenges	3
Clinical Validity	3
Usability	3
User Retention	4
Availability	4
CBT App Development Guidelines	5
CBT App Evaluation Techniques	6
Application of Literature	8
Chapter 3: Methodology	9
Design	9
Features	9
Prototyping	10
Testing	11
Methods	11
Recruiting	12
Analysis	13
Chapter 4: Results	15
Participant Demographics	15
Usability Tests	15
Pain Points	16

Positive Experiences	16
Feedback	17
System Usability Scale	17
Chapter 5: Discussion	18
Limitations	18
Design Changes	18
Next Steps	19
Conclusion	19
References	20
Appendix A: Beck Institute's <i>Testing Your Thoughts</i> Exercise	23
Appendix B: Survey Questions	24
Appendix C: Usability Test Tasks	25
Appendix D: SUS Questions	26

List of Tables

Table 1. Usability Test Task Completion Rates	15
Table 2. Participant's SUS Responses and Scores	17

List of Figures

Figure 1. Some wireframes in Adobe XD

10

Chapter 1: Introduction

In the past few years, mental health mobile apps, specifically those offering cognitive behavioral therapy, or CBT, have grown in popularity. CBT is an umbrella term for various cognitive and behavioral therapy techniques used to treat depression and anxiety, among other mental health conditions (Rathbone et al., 2017). Specifically, CBT techniques consist of three core principles: “cognitive activity affects behavior, cognitive activity may be monitored and altered, and desired behavior change can occur through cognitive change” (Rathbone et al., 2017, p. 2). By following these principles, people can use CBT to reevaluate their interpretations of situations and feelings in order to think and act more positively (Rathbone et al., 2017). When CBT is offered as a mobile app, users get cheap, quick access to CBT-based exercises and activities.

There are hundreds of CBT apps available for download; however, most either do not follow core CBT principles or are not user-friendly. Researchers have noticed this gap in the market, and they have compiled CBT app development best practices and written CBT app development guidelines. I believe that by learning from CBT app literature and following mobile app and interaction design best practices, I can design a CBT app that is both effective and usable to fill the gap in the CBT app market.

In this paper, I begin by reviewing existing CBT app literature in Chapter 2. In my literature review, I explore themes of CBT app efficacy, challenges, development guidelines, and evaluation techniques. In Chapter 3, I explain my app design process and testing methodologies. Finally, Chapters 4 and 5 outline the results of my study and discuss limitations and next steps, respectively.

CBT App Benefits

Apps vs. Websites

While CBT is most often used in traditional therapy settings, it has become more popular in e-Health settings as well (Marshall et al., 2020). e-Health is the “cost-effective and secure use of information and communications technologies in support of health and health-related fields” (Marshall et al., 2020, p. 21). Two popular e-Health mediums are websites and smartphone apps; however, several studies have proven the effectiveness of apps over websites. For example, Newton et al. (2020) argue that apps are better than websites because they can “take advantage of cross-functional device abilities” (p. 2), such as push notifications and calendar interactions. Additionally, Purkayastha et al. (2020) compared the usability of a CBT mobile app to a CBT website and found that the app was more usable and engaging. CBT app engagement is supported by Marshall et al. (2020), who found that CBT websites have a significantly higher dropout rate than CBT apps.

Popularity

CBT-based apps have become increasingly popular over the past decade, and there are hundreds available for download. Their popularity is due to several factors. First, apps are generally less expensive than traditional in-person therapy, making them more accessible to the general public (McGee, 2022). Also, apps can be used at users’ leisure, unlike traditional in-person therapy which requires traveling and scheduling (McGee, 2022). Finally, using a mental health app may not carry the same stigma that often comes with attending in-person therapy, which means that more people are likely to use them (Marshall et al., 2020; McGee, 2022).

Efficacy

Besides being more accessible and acceptable to users, mental health apps have also been proven to be as effective as in-person therapy, in most cases. For example, both Marshall et al. (2020) and Rathbone et al. (2017) found that mental health apps using

self-guided CBT techniques are effective at reducing symptoms of depression and anxiety. Additionally, Chandrashekhar (2018) found that for anxiety-focused CBT apps, there was no significant loss of treatment efficacy when replacing in-person therapy sessions with mobile apps.

There are limitations to CBT app effectiveness, however. Chandrashekhar (2018) states that CBT apps “yield the greatest benefit for individuals with mild to moderate, rather than major, depression” (p. 2), and Huckvale et al. (2020) found that the positive effects of CBT app use are “contingent on sustained engagement with app-delivered therapy” (p. 66). Also of note, Marshall et al. (2020) state that “the vast majority of the published research on the efficacy of mental health apps has been carried out by the individuals and organizations involved in that app’s development” (p. 24), which can lead to skewed or biased results.

CBT App Challenges

Clinical Validity

Although there are many CBT apps, they are not all usable and effective. One of the most prominent challenges that CBT apps face is clinical validity. According to Huckvale et al. (2020), CBT apps “appear to have low levels of adherence to core [CBT] components such as education and monitoring cognitions, sensations, and physical behaviors” (p. 66). In a 2016 study conducted by Huguet et al., researchers identified one hundred and seventeen apps claiming to use CBT, but only 10.26% of those apps were actually consistent with evidence-based CBT principles. Huguet et al. (2016) state that “there is a concerning lack of appropriate CBT … apps” (p. 1), which can be explained by non-research and non-clinician teams developing apps without regard to evidence-based guidelines (Huckvale et al., 2020). Huckvalue et al. (2020) warn that in extreme cases, clinically invalid CBT apps can include unsafe information that could degrade users’ mental health and create life-threatening conditions.

Usability

Another challenge to CBT app efficacy is usability, or the ability of users to understand, navigate, and engage with an app. Usability is particularly challenging for

CBT apps because, as Huguet et al. (2016) found, there is no relationship between a CBT app's clinical validity and its adherence to usability heuristics. For example, when using a clinically valid CBT app with bad usability, users "may interpret ineffectiveness as a treatment failure, when in fact, ineffectiveness may be the result of usability problems" (Huguet et al., 2016, p. 14). Conversely, if the app has good usability but is not clinically valid, users may enjoy using the app but find it ineffective.

User Retention

Even when CBT apps are both clinically valid and usable, many apps struggle with user retention. Huckvale et al. (2020) found that "only 4% of users who downloaded a mental health app opened it again after 15 days" (p. 67). This is consistent with Wasil et al. (2020), which state that within a week of installation, over 90% of users discontinue using mental health apps. Furthermore, although participants in Rathbone et al.'s 2017 study stated that they found CBT apps useful, "this was not reflected by longitudinal engagement" (p. 6).

Availability

Finally, finding a clinically valid, usable, and engaging CBT app is challenging. Unfortunately, many CBT apps are often developed for research purposes only (Marshall et al., 2020), and few evidence-based CBT apps are available for consumer download (Bakker et al., 2016). When CBT apps are available to the public, their popularity is heavily determined by user ratings. In fact, "the main way that individuals decide which mental health app to download is by using app ratings and reviews in app stores" (Marshall et al., 2020, p. 24). Ratings are influenced by usability, with the most engaging apps appearing in the first few hits (Wasil et al., 2020), and price, with lower-priced apps earning significantly higher ratings than higher-priced apps (Marshall et al., 2020). However, Neary and Schueller (2018) found that "consumer ratings do not seem to reflect clinical usefulness or utility" (p. 3), arguing that app ratings and reviews are not a sufficient way to help people identify clinically valid CBT apps.

CBT App Development Guidelines

To address CBT app challenges, such as clinical validity, usability, and retention, several researchers have created CBT app development guidelines. Most importantly, these guidelines address clinical validity by suggesting how best to apply CBT principles in a mobile app setting. For example, Bakker et al. (2016) suggest that CBT apps focus on three important functions: reporting of thoughts, feelings, or behaviors; recommended activities; and mental health information. These three functions adhere to the core components of a CBT approach to depression, which include education, monitoring, and behavioral and cognitive techniques (Huguet et al., 2016). However, in order to be clinically valid, it is not enough to simply use CBT principles. Marshall et al. (2020) argue that “clinicians have to be more involved in the app development process” (p. 25), and Rathbone et al. (2017) state that “future apps would benefit from input from multidisciplinary teams during the design and development stages” (p. 6). Clinician input is an important part of clinical validity because it ensures that CBT apps provide medically valid information and guidance (Rathbone et al., 2017).

While it is imperative that CBT apps are clinically valid, researchers warn that, in order to reach the largest audience, apps should be marketed to non-clinical users. For example, rather than marketing CBT apps as treatments, app developers should target the general public by marketing CBT apps as preventative (Bakker et al., 2016). Because of the social stigma surrounding mental health treatments, marketing the app as a preventative tool will increase the number of eligible and willing users (Bakker et al., 2016). This is supported by Bakker et al. (2016), which found that “self-help interventions were significantly more effective when recruitment occurred in non-clinical settings” (p. 7).

In addition to being clinically valid, researchers argue that CBT apps should adhere to usability standards. Huguet et al. (2016) state that app developers should address Nielsen’s usability heuristics in order to “optimize clinical benefits and make the app more usable” (p. 15). Nielsen’s usability heuristics are ten general principles for interaction design that user interface designers and developers, including CBT app

developers, are encouraged to follow (Nielsen, 2020). Of the ten heuristics, Chandrashekhar (2018) argues that CBT apps should prioritize those that reduce cognitive load by using pictures rather than text, reducing sentence lengths, and using inclusive, non-clinical language.

Because CBT app success is dependent on user retention, researchers include engagement recommendations in their development guidelines. User engagement can be addressed in two ways – by using interaction design techniques and by partnering with CBT providers. For example, interaction design techniques such as real-time engagement, usage reminders, and gamified interactions have been proven to be effective at increasing general app user retention, and they can be applied to CBT apps as well (Chandrashekhar, 2018). Additionally, Marshall et al. (2020) state that “the clinician’s perspective would appear to be vital in the successful uptake and wider dissemination of e-mental health resources” (p. 22). They encourage CBT providers to integrate CBT apps into their practice in the same way that they would use traditional homework assignments (Marshall et al., 2020).

CBT App Evaluation Techniques

While adhering to development guidelines is an excellent starting point in creating a CBT app, researchers argue that a CBT app’s effectiveness and usability should be backed by evidence-based research. Bakker et al. (2016) suggest measuring a CBT app’s effectiveness using assessments given before, during, and after app use and self-report questionnaires administered throughout app use. For example, Bakker et al. (2016) state that “it is recommended that follow-up data are collected at several different time points throughout MHapp [mental health app] intervention and after its use has been concluded” (p. 16). In practice, Kinderman et al. (2016) used mood assessments before and after app use to test whether or not their app was able to address mental health problems and improve mental wellness. Additionally, there are several well-known self-report questionnaires that give insight into users’ emotional well-being and intervention effectiveness. The most common questionnaires used are the 9-item Patient Health Questionnaire, the 7-item Generalized Anxiety Disorder scale, the ESA Scale, and the

Coping Self-Efficacy Scale (Bakker et al., 2016). Administering a combination of these scales can give a balanced assessment of users' emotional states and CBT app effectiveness (Bakker et al., 2016).

CBT app usability is often tested independently of its clinical validity, as was the case in Purkayastha et al. (2020). Common usability testing techniques include asking users to think aloud while using the app, administering the 10-item System Usability Scale, and evaluating the app against Nielsen's usability heuristics (Newton et al., 2020). When combined, these techniques can accurately measure a CBT app's usability.

The most popular scale to measure CBT app efficacy and usability is the Mobile App Rating Scale (MARS); however, several researchers and organizations have proposed their own CBT app rating guidelines (Neary & Schueller, 2018). For example, Huguet et al. (2016) suggest evaluating CBT apps on three dimensions: usefulness, usability, and integration. The American Psychiatric Association's app rating framework includes evaluating potential risk and harm, particularly surrounding privacy and data management; reviewing research evidence for efficacy; evaluating ease of use, or usability; and determining whether or not the app has the ability to share data with clinicians (Marshall et al., 2020; Neary & Schueller, 2018). Furthermore, Marshall et al. (2020) proposed their own framework for evaluating the effectiveness of mobile mental health tools called Mhabit. Although there are so many different CBT app rating frameworks, the "consensus among these varied guidelines is that multidimensional rating systems are needed" (Neary & Schueller, 2018, p. 4).

Because of the challenges associated with app store ratings, researchers have proposed app rating platforms that "act as clearinghouses for mobile apps in a particular domain" (p. 1). In their 2018 paper, Neary & Schueller promote their app rating platform called PsyberGuide, which aims to provide objective rating information on publicly available mental health apps. At the time, the authors claimed that PsyberGuide was the "most active and comprehensive [mental health mobile app rating] platform in the United States" (p. 4).

Application of Literature

In sum, although CBT apps have been proven to be effective at treating depression and anxiety, there are many challenges to developing a clinically valid, usable CBT app with high user retention. Researchers have addressed these challenges by creating sets of development guidelines, and there are still opportunities for app developers to capitalize on the gap in the CBT app market. Huguet et al. (2016) sum this up by stating:

A mobile app based on clinical best practice, that meets the most basic usability standards, that is evaluated scientifically, has a privacy policy, and deals with safety matters has the potential to remove barriers to care and alleviate suffering for a large number of people with depression at a modest cost. (p. 15)

Leveraging the lessons learned from CBT app literature and following app development guidelines, I will attempt to design a CBT app that is clinically valid, usable, and engaging.

Chapter 3: Methodology

After reviewing the literature, I designed my CBT app and tested its usability. In both the design and testing phases, I employed user-centered design principles by following interaction design best practices, soliciting early feedback from users and subject-matter experts, and using standard user research techniques.

Design

I began the design process by deciding which features my CBT app would include. Once I had the most important features defined, I started designing and prototyping my app.

Features

Because Bakker et al. (2016) state that the three most important features of a CBT app are reporting of thoughts, feelings, or behaviors; recommended activities; and mental health information, I prioritized these features. It was important to choose a clinician-approved CBT exercise to support thought, feeling, and behavior reporting, so I chose to adapt the Beck Institute's (2018) *Testing Your Thoughts* worksheet into a feature called "test a thought" (see Appendix A for the *Testing Your Thoughts* worksheet). As part of this exercise, users are asked to come up with an activity, and, as suggested by the literature, my app provides recommended activities, such as exercising or meditating, during this step. To satisfy the third most important feature, mental health information, I included an "articles" section in my app that gives users access to relevant mental health information and additional resources.

Outside of Bakker's (2016) feature recommendations, I included additional literature-recommended features in my CBT app. For example, because the American Psychiatric Association evaluates mental health apps based on their ability to share data with clinicians (Marshall et al., 2020), I included a "history" feature in my app that allows users to reference and share past completions of the "test a thought" exercise. To address user retention challenges described by Huckvale et al. (2020) and Rathbone et al. (2017), I used the interaction design principle of gamification to inform an

“achievements” feature, where users earn awards by completing “test a thought” exercises. To address privacy and data management concerns posed by Chandrashekhar (2018) and Huguet et al. (2016), I included a “frequently asked questions” feature in the app’s privacy policy and made the policy easy to find. Finally, to address Huckvale et al.’s (2020) clinical safety concerns, I added an “emergency resources” feature that provides national mental health hotline phone numbers.

Prototyping

With my main features defined, I began creating wireframes for my CBT app. In this stage, it was important to use a low-fidelity prototyping technique, such as wireframing, because it allowed for quicker design iteration. My initial designs were sketched on paper, and once I had defined the general layout and interaction of my app, I fleshed out the prototype in Adobe XD (see Figure 1). Before moving into the testing phase, I solicited feedback on my prototype from a clinician with CBT experience and a target end-user to ensure that it was clinically valid and usable, and I incorporated their suggestions into my design.

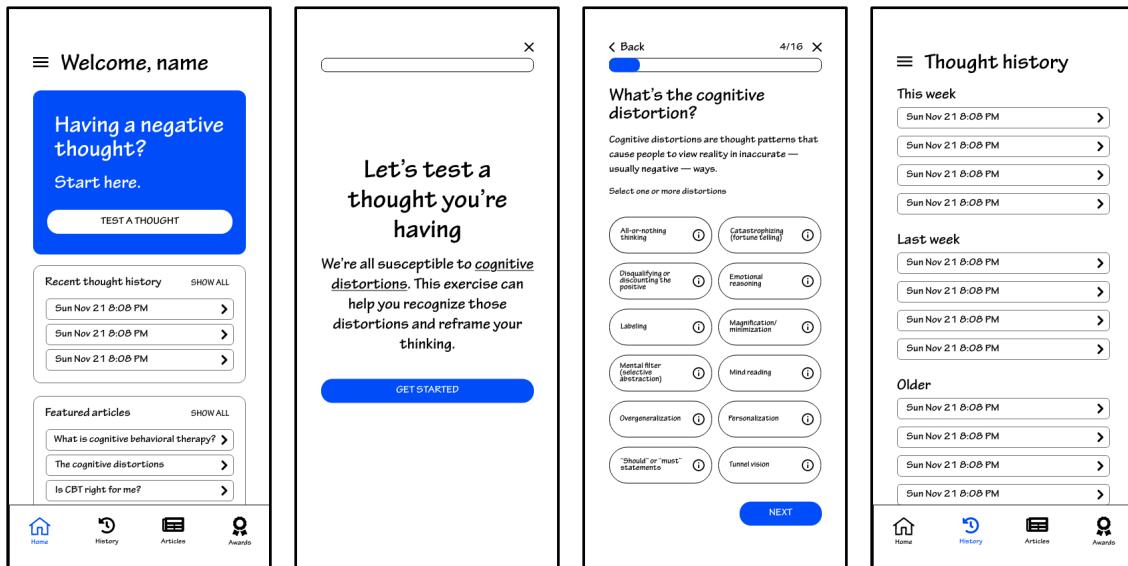


Figure 1. Some wireframes in Adobe XD.

Testing

Once my prototype was created, I tested it using standard usability testing techniques. The goals of this study were to test the usability of my CBT app and to gauge the value of the app to users. I intended to answer the following questions:

- Are users able to navigate the app?
- Are users able to successfully “test a thought”?
- Are users able to access previously “tested thoughts”?
- Are users able to find more information about CBT on the “articles” page?
- Are users able to find emergency resources?
- Do users find the app valuable?

Importantly, I chose not to test CBT efficacy and user retention because I believe that they warrant their own studies. Additionally, I chose not to use a randomized controlled trial (RCT) because Neary & Schueller (2018) state that “the time required to conduct an RCT and publish results does not align with the rapid development cycle of apps” (p. 2).

Methods

I used both qualitative and quantitative methods to test my app, including surveys, interviews, usability tests, and the System Usability Scale.

Surveys

Participants were asked to complete a survey prior to the start of the scheduled usability test session (see Appendix B for the full list of survey questions). The purpose of the survey was to obtain basic demographic information and gauge users’ experiences with mental wellness apps and CBT. I adapted the Nielsen Norman group’s “28 Tips for Great Qualitative Surveys” to create a clear, concise survey that generated meaningful responses (Farrell, 2016). The survey was created with an online survey tool and sent to participants’ email addresses.

Interviews

Interviews were conducted at both the beginning and end of each scheduled usability test session to learn about users’ preconceptions of mental wellness apps and

feelings about my app's design (Pernice, 2018). At the beginning of the session, I asked each participant follow-up questions based on their answers to the pre-test survey in order to learn more about their existing mental wellness app usage habits. At the end of the test session, participants and I discussed their experiences with and impressions of the app. Additionally, I used this time to ask follow-up questions about the tasks they completed, if necessary, and I asked general questions about their perceived value of the app. Participants were given an opportunity to ask me questions and give feedback. Interviews occurred either in person or over a video conference call.

Usability Tests

The bulk of my study consisted of conducting usability tests with participants. I chose to conduct usability tests because they are an industry-standard observation methodology used to uncover design problems (Moran, 2019). During the usability tests, participants were given the low-fidelity prototype of my app and asked to complete a set of tasks that directly related to my research questions (see Appendix C for the full list of tasks). As users completed the tasks, I observed and took notes. Usability tests were either completed in person or over a video conference call with screen sharing enabled. The prototype was given to participants using the “share” features of Adobe XD.

System Usability Scale

After completing the usability test tasks, participants filled out the System Usability Scale (SUS). The SUS is a well-known post-test survey that has been used to quantitatively measure usability since 1986 (Gallavin, 2014). It contains ten questions that ask users for their opinions on how easy a system is to use, and each question is answered with a five-item Likert scale, where the values range from strongly disagree to strongly agree (see Appendix D for the full list of SUS questions) (Gallavin, 2014). The survey was created with an online survey tool and given to participants during the scheduled test session.

Recruiting

I recruited three people to participate in my study. I chose to recruit only three people because, according to Neilsen (2012), usability tests only require a few

participants to find most usability problems. Testing with more than five participants generally exceeds the test's maximum cost-benefit ratio (Neilsen, 2012).

Participants were adult members of the general public that were recruited from a convenience sample. They were not pre-screened for or asked about existing mental health conditions, such as depression or anxiety, which are typically conditions targeted by CBT. I chose to recruit from the general public, rather than from a sample with pre-existing mental health conditions because the literature suggests that CBT apps are more successful when they are marketed as a preventative tool (Bakker et al., 2016).

Logistically, targeting the general public also expanded the pool of possible participants, and participants did not have to divulge sensitive health information in order to participate in the study.

Analysis

Because I collected qualitative and quantitative data, data analysis differed depending on the collection method.

Qualitative

Qualitative data were collected in the form of notes from user interviews and observations of usability test task completion. To derive themes from my data, I followed a grounded theory approach by using open, axial, and selective coding techniques (Delve et al., 2022). In the open coding stage of analysis, I labeled, or coded, my data to break it up into discrete pieces. Next, during the axial coding stage, I grouped similar codes together and drew connections between code groups. Finally, in the selective coding phase, I synthesized common themes from code groups.

Quantitative

Quantitative data were collected in the form of surveys, task completion rates, and the SUS. Survey response answers provided demographic metrics, and whether or not participants were able to complete usability test tasks yielded completion rates per task. As outlined in Smyk (2020), responses to the SUS were calculated per participant by

- summing responses to odd-numbered questions and subtracting by 5 to get X;
- summing responses to even-numbered questions and subtracting from 25 to get Y;

- multiplying X plus Y by 2.5

After calculating each participant's SUS score, I took the average and compared that number against the average SUS score of 68 (Sauro, 2011). A higher SUS score means that my app is more usable than the average application, and a lower SUS score means that my app is less usable than the average application.

Chapter 4: Results

Participant Demographics

Participants varied in age and gender. Age ranges included 18 to 24 years old, 25 to 39 years old, and 40 to 60 years old. Two participants identified as female, and one participant identified as male.

All three participants had heard of CBT before, and all had downloaded a mental wellness app, including journaling, therapy, or meditation apps. However, only two out of the three participants were actively using a mental wellness app at the time of the study. Participants' prior experiences with mental wellness apps varied. For example, one participant liked not having to leave their house for therapy services, but they did not like adding more screen time to their day. Another participant enjoyed using some mental wellness apps as part of their daily routine, but they felt like many mental wellness apps put too much pressure on them to be "happy." Finally, one participant had stopped using mental wellness apps because they felt like the apps' value had diminished over time.

Usability Tests

In general, participants completed all of the usability test tasks with little to no issues (see Table 1 for task completion rates). After analyzing the test session data, trends in pain points, positive experiences, and participant feedback emerged.

Table 1

Usability Test Task Completion Rates

Task	Participants			Rate
	P1 Completed?	P2 Completed?	P3 Completed?	
T1	Y	Y	Y	100%
T2	Y	Y	Y	100%
T3	Y	Y	Y	100%
T4	Y	Y	Y	100%
T5	Y	Y	Y	100%

Pain Points

While completing the usability test tasks, participants experienced a few common pain points. The most prominent pain point was the app's various navigation schemes. For example, participants had trouble differentiating between tab navigation, which is used for main pages, and drawer navigation, which is used for secondary information. On more than one occasion, participants looked for the "Home" button in the drawer navigation instead of the tab navigation. However, after using the app for a couple of minutes, participants adapted to the navigation structure.

Another common pain point that participants experienced occurred during the "test a thought" exercise. In one step of the exercise, participants are asked to choose which cognitive distortions they may be experiencing. The names of these cognitive distortions are a CBT standard, but they are not entirely intuitive. While the app defines each cognitive distortion, participants had trouble finding these definitions. Specifically, it took one participant a few moments to find the definitions, another participant needed prodding to find them, and the last participant did not find the definitions at all.

Finally, all participants expected emergency resources to be more easily accessible. When asked to find emergency resource information, all three participants scrolled to the bottom of the "Home" page; however, when they noticed that the information was not there, they checked the navigation drawer and found the link to the information. Participants commented that they thought that life-saving information like emergency resources should be more prominently displayed in the app.

Positive Experiences

Although participants experienced common pain points while using the application, they had common positive experiences as well. For example, after completing the "test a thought" exercise, two out of three participants were excited to see that they had earned an award. Additionally, all three participants found the "test a thought" exercise valuable, and they all felt better about their negative thought by the end of the exercise. In particular, they liked that the exercise was more than just reflecting or journaling; rather, they felt as though they had accomplished something by the end of the

exercise. This was reflected by the fact that all three participants believed their negative thought was less true by the end of the exercise. When asked whether or not they thought they would use the app if it was real, two out of three participants said that they would, and the remaining participant said that they would consider using the app in the future if they felt like they would benefit from CBT.

Feedback

Finally, during the interview portion of the study, participants gave general feedback and suggested app improvements. One participant suggested allowing therapists direct access to user entries, via a therapist-version of the app, to monitor progress. Another participant suggested using push notifications to remind users to use the app throughout the day. Finally, the same participant suggested designing the app with calming aesthetics to make the entire experience more pleasant and relaxing.

System Usability Scale

The app's usability was quantified by participants' responses to the System Usability Scale (see Table 2). I wrote a Python program to calculate the overall SUS score of 92, which is higher than the average of 68. This means that, based on the responses from my three participants, my app is more usable than the average app.

Table 2

Participants' SUS Responses and Scores

Participant	Questions										Score
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	
P1	5	1	4	1	5	1	5	1	5	1	97.5
P2	4	1	4	1	3	1	5	1	3	1	85.0
P3	4	1	5	1	4	1	4	1	5	1	92.5
Average											92.0

Chapter 5: Discussion

Limitations

Before analyzing the results of my study, it is important to understand its limitations. One of the main limitations was the fact that I used a convenience sample. Convenience sampling does not guarantee a diverse, representative sample, and, when using this sampling method, it is more difficult to generalize results to a larger population. Additionally, because means are highly impacted by outliers, the mean SUS scores that I calculated should be interpreted with caution, given my small sample size. Although I designed my app around the Beck Institute's (2018) *Testing Your Thoughts* exercise, I did not use this study to confirm its efficacy as a valid or engaging CBT technique. Because I tested an app of my design, both observation and confirmation bias were possible. Finally, because users knew that they were being observed while using the app, they were susceptible to the Hawthorne effect, which states that users change their behaviors during observation.

Design Changes

Based on the results of the study, there are several usability improvements to be made to the app. First, because there was some initial confusion between the drawer and tab navigation, either more distinction needs to be made between the two navigation schemes or a new navigation structure needs to be designed. Second, because users had a hard time finding the descriptions for each cognitive distortion, the buttons to view cognitive distortion definitions need to be clearer, and text explaining how to find the descriptions needs to be added to the “test a thought” exercise. Additionally, the app could feature a tutorial for first-time users to support the psychoeducation period in which users learn CBT terms and principles. Finally, because users’ instincts were to look for emergency resources at the bottom of the “Home” page, a link to emergency resources needs to be added there.

Next Steps

After making the above design changes, I will begin developing the app for production. Aside from the technical tasks that need to be completed to develop an app and make it available to users, as development progresses, more input from clinicians will be needed. For example, I will need to consult clinicians to source content for the “articles” section of my app. Additionally, in order to write a clear, comprehensive privacy policy and address possible HIPAA concerns, I will need to source legal help. To create a comprehensive set of engaging awards for the app’s “awards” feature, I will need to conduct more research on gamification. Additionally, I should conduct a diary study to test long-term user engagement. As the app matures, participant-suggested features such as therapist integration, push notifications, and better aesthetics will be added to the app.

Conclusion

In conclusion, it is clear from the literature that there is a market need for a usable and effective CBT app. By following CBT app development guidelines, collaborating with a clinician, and verifying the usability of my app with end-users, I believe that I have designed an app to fill this need. In the future, I hope that I will be able to fully develop this app and make it available to end-users so that I can put research into action.

References

- Bakker, D., Kazantzis, N., Rickwood, D., & Rickard, N. (2016). Mental health smartphone apps: review and evidence-based recommendations for future developments. *JMIR mental health*, 3(1), e4984.
- Beck Institute. (2018). Testing Your Thoughts [Worksheet]. Retrieved from <https://beckinstitute.org/wp-content/uploads/2021/08/Testing-Your-Thoughts-Worksheet.pdf>
- Chandrashekhar, P. (2018). Do mental health mobile apps work: evidence and recommendations for designing high-efficacy mental health mobile apps. *Mhealth*, 4.
- Delve, Ho, L., & Limpaecher, A. (2022, February 8). *How To Do Open, Axial, & Selective Coding in Grounded Theory*. Practical Guide to Grounded Theory. Retrieved May 19, 2022, from <https://delvetool.com/blog/openaxialselective>
- Farrell, S. (2016, September 25). *28 Tips for Creating Great Qualitative Surveys*. Nielsen Norman Group. Retrieved May 19, 2022, from <https://www.nngroup.com/articles/qualitative-surveys/>
- Gallavin, G. (2014, August 29). *System Usability Scale (SUS): Improving Products Since 1986*. Digital.Gov. Retrieved May 19, 2022, from <https://digital.gov/2014/08/29/system-usability-scale-improving-products-since-1986/>
- Huckvale, K., Nicholas, J., Torous, J., & Larsen, M. E. (2020). Smartphone apps for the treatment of mental health conditions: status and considerations. *Current opinion in psychology*, 36, 65-70.
- Huguet, A., Rao, S., McGrath, P. J., Wozney, L., Wheaton, M., Conrod, J., & Rozario, S. (2016). A systematic review of cognitive behavioral therapy and behavioral activation apps for depression. *PloS one*, 11(5), e0154248.
- Kinderman, P., Hagan, P., King, S., Bowman, J., Chahal, J., Gan, L., ... & Tai, S. (2016). The feasibility and effectiveness of Catch It, an innovative CBT smartphone app. *BJPsych open*, 2(3), 204-209.

- Marshall, J. M., Dunstan, D. A., & Bartik, W. (2020). Clinical or gimmickal: The use and effectiveness of mobile mental health apps for treating anxiety and depression. *Australian & New Zealand Journal of Psychiatry*, 54(1), 20-28.
- McGee, V. (2022, April 25). *Virtual Therapy vs. In-Person Therapy*. Psychology.Org. Retrieved May 19, 2022, from <https://www.psychology.org/resources/virtual-therapy-vs-in-person/>
- Moran, K. (2019, December 1). *Usability Testing 101*. Nielsen Norman Group. Retrieved May 19, 2022, from <https://www.nngroup.com/articles/usability-testing-101/>
- Neary, M., & Schueller, S. M. (2018). State of the field of mental health apps. *Cognitive and Behavioral Practice*, 25(4), 531-537.
- Newton, A., Bagnell, A., Rosychuk, R., Duguay, J., Wozney, L., Huguet, A., ... & Curran, J. (2020). A Mobile Phone-Based App for Use During Cognitive Behavioral Therapy for Adolescents With Anxiety (MindClimb): User-Centered Design and Usability Study. *JMIR mHealth and uHealth*, 8(12), e18439.
- Nielsen, J. (2012, June 3). How many test users in a usability study? Nielsen Norman Group. Retrieved April 15, 2022, from <https://www.nngroup.com/articles/how-many-test-users/>
- Nielsen, J. (2020, November 15). 10 usability heuristics for user interface design. Nielsen Norman Group. Retrieved December 28, 2021, from <https://www.nngroup.com/articles/ten-usability-heuristics/>
- Pernice, K. (2018, October 7). *User Interviews: How, When, and Why to Conduct Them*. Nielsen Norman Group. Retrieved May 19, 2022, from <https://www.nngroup.com/articles/user-interviews/>
- Purkayastha, S., Addepally, S. A., & Bucher, S. (2020). Engagement and usability of a cognitive behavioral therapy mobile app compared with web-based cognitive behavioral therapy among college students: randomized heuristic trial. *JMIR human factors*, 7(1), e14146.

- Rathbone, A. L., Clarry, L., & Prescott, J. (2017). Assessing the efficacy of mobile health apps using the basic principles of cognitive behavioral therapy: systematic review. *Journal of medical Internet research*, 19(11), e399.
- Sauro, J. (2011, February 3). *Measuring usability with the system usability scale (SUS)*. MeasuringU. Retrieved May 18, 2022, from <https://measuringu.com/sus/>
- Smyk, A. (2020, March 17). The system usability scale & how it's used in UX. Adobe XD Ideas. Retrieved January 24, 2022, from <https://xd.adobe.com/ideas/process/user-testing/sus-system-usability-scale-ux/>
- Wasil, A. R., Weisz, J. R., & DeRubeis, R. J. (2020). Three questions to consider before developing a mental health app. *World Psychiatry*, 19(2), 252.

Appendix A: Beck Institute's *Testing Your Thoughts* Exercise

1. What is the situation? You might be having thoughts about something that just happened in the environment or something that happened inside of you (e.g., an intense emotion, a painful sensation, an image, a daydream, a flashback or a stream of thoughts, such as thinking about my future).
2. What am I thinking or imagining?
3. What is the cognitive distortion?
4. What makes me think the thought is true?
5. What makes me think the thought is not true or not completely true?
6. What's another way to look at this?
7. If the worst happens, what could I do then?
8. What's the best that could happen?
9. What will probably happen?
10. What will happen if I keep telling myself the same thought?
11. What could happen if I changed my thinking?
12. What would I tell my friend or family member [think of a specific person] if this happened to him or her?
13. What would be good to do now?

Appendix B: Survey Questions

1. What is your name?
2. What is your age range?
 - a. 18 to 24
 - b. 25 to 39
 - c. 40 to 59
 - d. 60 plus
3. With which gender do you identify?
 - a. Male
 - b. Female
 - c. Non-binary
 - d. Other
4. Have you ever heard of cognitive behavioral therapy (CBT)?
 - a. Yes
 - b. No
 - c. Maybe
5. Have you ever downloaded a mental wellness mobile app, such as a journaling, therapy, or meditation app?
 - a. Yes
 - b. No
6. Are you actively using a mental wellness mobile app, such as a journaling, therapy, or meditation app?
 - a. Yes
 - b. No

Appendix C: Usability Test Tasks

1. You've just downloaded the app from the app store, and you're opening it for the first time. Show me how you would create and log in to an account.
2. You're in a situation where you're having a negative thought, and you want to use the app to help feel better. Show me how you would do this.
3. You want to revisit a thought that you've previously tested. For example, maybe you want to share it with your therapist, or maybe you'd like to reflect back on the progress you've made. Show me how you would do this.
4. You want to learn more about CBT. Show me how you would do this.
5. You're in a crisis situation, and you need to find information about emergency mental health resources. Show me how you would do this.

Appendix D: SUS Questions

1. I think that I would like to use this app.
2. I found the app unnecessarily complex.
3. I thought the app was easy to use.
4. I think that I would need the support of a technical person to be able to use this app.
5. I found the various functions in this app were well integrated.
6. I thought there was too much inconsistency in this app.
7. I would imagine that most people would learn to use this app very quickly.
8. I found the app very cumbersome to use.
9. I felt very confident using the app.
10. I needed to learn a lot of things before I could get going with this app.