USER ENGAGEMENT WITH APPS FOR DEPRESSION

Danielle Rae Schweitzer
Rowan University

Follow this and additional works at: https://rdw.rowan.edu/etd

Part of the Clinical Psychology Commons

Recommended Citation
Schweitzer, Danielle Rae, "USER ENGAGEMENT WITH APPS FOR DEPRESSION" (2023). Theses and Dissertations. 3151.
https://rdw.rowan.edu/etd/3151

This Thesis is brought to you for free and open access by Rowan Digital Works. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of Rowan Digital Works. For more information, please contact graduateresearch@rowan.edu.
USER ENGAGEMENT WITH APPS FOR DERPESSION

Danielle Rae Schweitzer

A Thesis

Submitted to the
Department of Psychology
College of Science and Mathematics
In partial fulfillment of the requirement
For the degree of
Master of Arts in Clinical Psychology
at
Rowan University
June 21, 2023

Dissertation Chair: Jim A. Haugh, Ph.D., Associate Professor & Director of Clinical Training, Department of Psychology

Committee Members:
Dustin Fife, Ph.D., Associate Professor, Department of Psychology
Steven Brunwasser, Ph.D., Assistant Professor, Department of Psychology
Acknowledgements

I would like to express appreciation toward Jim A. Haugh, Ph.D. for his continued mentorship and support throughout this research. The skills and knowledge I have learned will aid me throughout my future education and I look forward to future academic endeavors. I would like to thank Dustin Fife, Ph.D. and Steve Brunwasser, Ph.D. for being active members of my thesis committee and for their guidance with data analysis. I would like to thank Ninelle Edenne, Natalie Policella, Madison Giordano, Erin O’Donnell, and the rest of our research team for their assistance with recruitment efforts for this project. Finally, I would like to express my appreciation to my friends and family for their unwavering support throughout this process.
Abstract

Danielle Schweitzer

USER ENGAGEMENT WITH APPS FOR DEPRESSION

2022-2023

Jim A. Haugh, Ph.D.

Master of Arts in Clinical Psychology

Depression is a debilitating and prevalent psychiatric condition, however, few individuals with depression seek mental health treatment. Mental health apps (MHapps) represent one treatment delivery option that may reduce gaps in mental health care. Current evidence suggests that MHapps are efficacious for reducing depressive symptoms, yet little is known about how MHapps are effective. Thus, the aims of the present study were 1) to examine whether two MHapps were effective for reducing depressive symptoms and 2) to determine whether variables such as user engagement, guidance, outcome expectancies, preferences, and coping skill development were associated with effectiveness. A community sample (N=20) was recruited and completed measures over a two week period. Results suggested that symptoms of depression decreased within the sample across both app conditions and were impacted by the reception of outside mental health treatment. Additionally, increases in mindfulness skills were associated with decreases in depressive symptoms across app conditions. Clinical implications of findings and future directions will be explored.
# Table of Contents

Abstract ............................................................................................................................... iv

List of Figures .................................................................................................................... viii

List of Tables ..................................................................................................................... x

Chapter 1: Introduction ..................................................................................................... 1

User Engagement .............................................................................................................. 4

Guidance ............................................................................................................................. 9

User Expectations and Preferences ................................................................................ 11

Coping Skills ..................................................................................................................... 14

Aims of the Present Study ................................................................................................. 15

Chapter 2: Method .......................................................................................................... 19

Participants ....................................................................................................................... 19

Inclusion and Exclusion Criteria ..................................................................................... 22

Recruitment ....................................................................................................................... 23

Procedure .......................................................................................................................... 24

Intervention ....................................................................................................................... 25

Measures ............................................................................................................................ 27

The Patient Health Questionnaire-8 (PHQ-8) ................................................................ 27

Mobile Application Usage ............................................................................................... 27

User Engagement Scale – Short Form (UES-SF) ............................................................ 27
Table of Contents (Continued)

Depression Change Expectancy Scale (DCES) ................................................................. 28

Treatment Preferences ................................................................................................. 28

Five Facet Mindfulness Questionnaire-Short Form (FFMQ-SF) ................................. 28

Cognitive-Behavioral Therapy Skills Questionnaire. (CBTSQ) ................................... 29

The Coping Self-Efficacy Scale (CSE Scale) ................................................................. 29

Data Analysis .................................................................................................................. 29

Chapter 3: Results ........................................................................................................ 33

Effectiveness for Depressive Symptoms ................................................................. 33

Condition and Treatment ............................................................................................. 38

Engagement and Effectiveness ..................................................................................... 45

Guidance and Effectiveness ......................................................................................... 52

User Expectations, Preferences, and Effectiveness ................................................. 53

User Expectations ......................................................................................................... 53

Preferences .................................................................................................................. 55

Coping Skills and Effectiveness .................................................................................. 58

Mindfulness Skills ....................................................................................................... 58

Cognitive Behavioral Skills ......................................................................................... 64

Coping Self-Efficacy ..................................................................................................... 70

Chapter 4: Discussion .................................................................................................. 75
### Table of Contents (Continued)

- Areas of Focus and Future Directions ................................................................. 75
- Effectiveness for Depressive Symptoms ............................................................... 75
- Engagement and Effectiveness .............................................................................. 77
- Guidance and Effectiveness .................................................................................. 77
- User Expectations, Preferences, and Effectiveness ............................................. 78
- Coping Skills and Effectiveness ............................................................................ 78
- Clinical Implications .............................................................................................. 80
- Limitations ............................................................................................................. 81
- Conclusion ............................................................................................................. 83
- References ............................................................................................................ 84
- Appendix A: Diagnostics ....................................................................................... 92
- Appendix B: R Script ............................................................................................. 102
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1. Participant Flow</td>
<td>20</td>
</tr>
<tr>
<td>Figure 2. Model Comparison of PHQ Scores Over Time</td>
<td>35</td>
</tr>
<tr>
<td>Figure 3. PHQ Scores Over Time</td>
<td>37</td>
</tr>
<tr>
<td>Figure 4. PHQ Scores and Condition Model Comparison</td>
<td>41</td>
</tr>
<tr>
<td>Figure 5. PHQ Scores, Condition, and Treatment Model Comparison</td>
<td>42</td>
</tr>
<tr>
<td>Figure 6. PHQ Scores, Condition, and Treatment</td>
<td>45</td>
</tr>
<tr>
<td>Figure 7. PHQ Scores, Minutes, and Condition Model Comparison</td>
<td>49</td>
</tr>
<tr>
<td>Figure 8. PHQ Scores, UES Scores, and Condition Model Comparison</td>
<td>50</td>
</tr>
<tr>
<td>Figure 9. PHQ Scores, UES Scores, and Condition</td>
<td>52</td>
</tr>
<tr>
<td>Figure 10. PHQ Score and Expectancy Model Comparison</td>
<td>55</td>
</tr>
<tr>
<td>Figure 11. PHQ Score and Preference Model Comparison</td>
<td>57</td>
</tr>
<tr>
<td>Figure 12. PHQ Scores, FFMQ Scores, and Condition Model Comparison</td>
<td>62</td>
</tr>
<tr>
<td>Figure 13. PHQ Scores, FFMQ Scores, and Condition</td>
<td>64</td>
</tr>
<tr>
<td>Figure 14. PHQ Scores, CBTSQ Scores, Condition Model Comparison</td>
<td>67</td>
</tr>
<tr>
<td>Figure 15. PHQ Scores, CBTSQ Scores, and Condition</td>
<td>69</td>
</tr>
<tr>
<td>Figure 16. PHQ Scores, CSE Scores, and Condition Model Comparison</td>
<td>72</td>
</tr>
<tr>
<td>Figure</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>Figure 17. PHQ Scores, CSE Scores, and Condition</td>
<td>74</td>
</tr>
</tbody>
</table>

List of Figures (Continued)
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1. Sociodemographic Characteristics of Participants</td>
<td>21</td>
</tr>
<tr>
<td>Table 2. Mean PHQ Scores Over Time</td>
<td>34</td>
</tr>
<tr>
<td>Table 3. Model Comparison Between the Baseline and Time (With Time) Model</td>
<td>34</td>
</tr>
<tr>
<td>Table 4. Fixed and Random Effects for the Time Model</td>
<td>36</td>
</tr>
<tr>
<td>Table 5. Model Comparisons Between 1) the Reduced (With Condition and Time) and Baseline Model and 2) the Reduced (With Condition and Time) and Full (With Condition, Treatment, and Time) Model</td>
<td>40</td>
</tr>
<tr>
<td>Table 6. Fixed and Random Effects for the Full Model, Which Includes Depression, Time, Condition, and Treatment</td>
<td>43</td>
</tr>
<tr>
<td>Table 7. Mean Engagement Scores Over Time Separated by Condition</td>
<td>46</td>
</tr>
<tr>
<td>Table 8. Model Comparisons Between 1) the Reduced (With Condition and Time) and Minutes Model (With Minutes, Condition, and Time) and 2) the Reduced (With Condition and Time) and UES Model (UES Score, Condition, and Time)</td>
<td>48</td>
</tr>
<tr>
<td>Table 9. Fixed and Random Effects for the UES Model, Which Includes Depression, Time, Condition, and User Engagement</td>
<td>51</td>
</tr>
<tr>
<td>Table 10. Model Comparison Between Baseline (With Time) and Expectancy (With Expectations and Time) Model</td>
<td>54</td>
</tr>
<tr>
<td>Table 11. Model Comparison Between Baseline (With Time) and Preference (With Preferences and Time) Model</td>
<td>57</td>
</tr>
<tr>
<td>Table 12. Mean FFMQ Scores Over Time Separated by Condition</td>
<td>59</td>
</tr>
</tbody>
</table>
List of Tables (Continued)

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 13. Model Comparison Between Reduced (With Condition and Time) and Mindfulness (With FFMQ Scores, Condition, and Time) Model</td>
<td>61</td>
</tr>
<tr>
<td>Table 14. Fixed and Random Effects for the Full Model, Which Includes Depression, Time, Condition, and Mindfulness</td>
<td>63</td>
</tr>
<tr>
<td>Table 15. Mean CBTSQ Scores Over Time Separated by Condition</td>
<td>65</td>
</tr>
<tr>
<td>Table 16. Model Comparison Between the Baseline (With Condition and Time) and CBT Model (With CBTSQ Scores, Condition, and Time)</td>
<td>66</td>
</tr>
<tr>
<td>Table 17. Fixed and Random Effects for the Full Model, Which Includes Depression, Time, Condition, and CBTSQ Skills</td>
<td>68</td>
</tr>
<tr>
<td>Table 18. Mean CSE Scores Over Time Separated by Condition</td>
<td>70</td>
</tr>
<tr>
<td>Table 19. Model Comparison Between the Baseline (With Condition and Time) and CSE (With CSE Score, Condition, and Time) Model</td>
<td>71</td>
</tr>
<tr>
<td>Table 20. Fixed and Random Effects for the Full Model, Which Includes Depression, Time, Condition, and Coping Self-Efficacy</td>
<td>73</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

Depression continues to be one of the most debilitating and prevalent psychiatric disorders. Recent reports suggest that depressive disorders occur in approximately 8.4% of the U.S. adult population (National Institute of Mental Health, 2022). In addition to being prevalent, depression causes significant dysfunction and is often tied to functional impairments in work, relationships, and health (Kessler, 2012). Decreased workplace productivity and increased utilization of healthcare subsequently contribute to the high economic burden created by depression (Greenberg et al., 2003). Due to the dysfunction that depression causes, it is a leading cause of global disability that burdens both the individual and society (World Health Organization, 2021).

Despite the large number of depressed individuals and the great impact that depression has on functioning, it is estimated that less than one-third of Americans experiencing depression seek therapy (Olfson et al., 2016). These low rates of treatment seeking may be associated with barriers that individuals face when seeking treatment. Environmental factors may include the high cost of treatment and lack of available treatment providers, while internal factors may include reduced motivation associated with depression and stigma surrounding mental health treatment (Donker et al., 2013; Mohr et al., 2010). Given the gap between the prevalence of depression and treatment-seeking rates, alternatives to traditional treatment delivery should be considered.

One alternative approach to traditional treatment delivery is the use of mental health applications (MHapps). MHapps are smartphone applications that are developed to teach coping skills, monitor symptoms, and/or provide information about mental health
Concerns. Compared to traditional treatment delivery methods, MHapps have many advantages. First, MHapps are more accessible than traditional treatment. Second, MHapps are less expensive than psychotherapy, as many MHapps are available at little to no cost. Third, MHapps offer greater flexibility because they can be used by the individual as needed. Fourth, MHapps can be used in isolation and thus minimize potential stigma associated with seeking treatment (Donker et al., 2013). Fifth, app users can track their moods and symptoms in real-time, which may provide insight into possible triggers of mood states (Proudfoot et al., 2010). Finally, MHapps can be personalized and tailored to match user preferences. MHapps may offer various features, including psychoeducation, meditation, and journaling. Overall, these advantages suggest that MHapps have the potential to expand access to care.

Along with these advantages, evidence suggests that MHapps may be efficacious interventions for depression. Firth and colleagues (2017) conducted a meta-analysis of 18 RCTs to examine the efficacy of MHapps for treating depression. Results indicated that the use of MHapps had a moderate positive effect on symptoms of depression (g=0.56), compared to waitlist control conditions. Specific app features including feedback, the presence of mindfulness or CBT skills, and mood monitoring were also examined in relation to effect size. Of these features, only the presence of feedback impacted study effect sizes. Feedback features were defined as progress scores or summary statistics derived from user data. MHapps with feedback features had moderate effect sizes (g=0.53), which was greater than the small effect sizes of MHapps that did not provide feedback to users (g=0.27).
Linardon and colleagues (2019) examined the effectiveness of MHapps to improve a number of outcomes, including depression, anxiety, stress, and quality of life. Fourteen trials specifically focused on the efficacy of MHapps for treating depressive symptoms. Results suggested that MHapps have a small positive effect on depressive symptoms ($g=0.28$) when compared to control conditions. However, effect sizes varied based on the type of control condition that was included in the trial. Trials that included placebo control conditions (e.g., gaming apps, listening to music) had a smaller effect size ($g=0.12$), while trials using informational ($g=0.39$) and waitlist controls ($g=0.32$) had moderate effect sizes. Within each type of control condition, effect sizes remained statistically significant.

Taken together, these two meta-analyses suggest that MHapps can be effective in helping individuals decrease their depressive symptoms. However, these findings provide little information about the processes through which MHapps are effective. Within psychotherapy literature, researchers have focused on intervention outcomes and the process of therapeutic change separately. Increasingly, there have been calls to investigate questions related to both process and outcome in intervention studies (Pachankis & Goldfried, 2007). As evidence indicates that MHapps are effective for depression, examining the processes through which they are effective remains paramount for increasing potential benefit. Paralleling the psychotherapy literature, it remains crucial to continue examining not only if MHapps are effective, but also how MHapps are effective.

Several key questions are salient for examining the process through which MHapps are effective. First, how can the delivery of MHapps be adapted to optimize
effectiveness? Both the level of user engagement necessary to achieve clinically significant benefit and the incorporation of guidance may impact effectiveness. Second, how can MHapps be tailored to meet user needs? User-level variables, such as preferences and expectancies for symptomatic change, may indicate for whom these interventions may be most effective. Third, why are MHapps effective? MHapps may facilitate the development of specific types of coping skills (e.g., mindfulness, cognitive-behavioral skills) that lead to symptomatic reduction. Changes in coping self-efficacy may be another variable that affects symptomatic change. Closer observation of each of these variables may elucidate the process through which MHapps can be effective for depressive symptoms.

**User Engagement**

User engagement is one variable related to the delivery of MHapps that may impact their effectiveness for depression. Engagement has been defined as the interaction between a user and an app. Subjectively, users may experience a feeling of interest, enjoyment, and involvement while using their app and engaged users experience a state of focused attention (O’Brien, 2016; Stoyanov et al., 2015). User engagement can also be influenced by user perceptions of the aesthetics and novelty of apps (O’Brien, 2016).

While some operationalizations of engagement have included user adherence, user engagement and adherence have also been conceptualized separately across the literature. Unlike engagement, user adherence refers solely to the continued use of an app. Metrics for user adherence include the number of minutes that a person has used an
app, the number of logins, or the proportion of content within an app that has been accessed.

User engagement with MHapps has been measured variably in the literature. Ng and colleagues (2019) reviewed 40 studies that focused on user engagement with MHapps for depression, anxiety, bipolar disorder, and schizophrenia. The authors defined usability, feasibility, acceptability, and satisfaction as indicators of user engagement. Eighty percent of the studies focused on multiple indicators of user engagement. Additionally, the authors differentiated between objective and subjective criteria used to measure engagement. Objective criteria included indicators of user adherence to the intervention, while subjective criteria included surveys and interviews that focused on user satisfaction or usability. Results indicated that across 60% of the studies, 71 objective measures of user adherence were included. Common measures included the frequency of logins, usage of specific features, proportion of app content that was completed, survey completion, and duration of app usage. Ninety percent of studies incorporated subjective engagement criteria, including 371 questions. Subjective measures included questions regarding ease of use, acceptability, and user satisfaction. Only 13 studies utilized existing assessment tools to develop these questions. Across both objective and subjective measures of engagement, various thresholds were used to define positive indicators of user engagement. In spite of discrepancies between thresholds across studies, the authors of all studies concluded that most users were engaged with their app.

Sieverink and colleagues (2017) conducted a systematic review of how adherence has been operationalized for mental and physical health apps. Their review included a
total of 62 studies that examined an e-health intervention. Fifty-nine percent of studies defined adherence as “the more use, the better” and provided no specific threshold for use. Within the remaining studies, adherence was defined as meeting a specific threshold. Yet few studies provided justification for a threshold via theory, evidence, or rationale. These findings highlight the importance of providing scientific justification for making decisions on the thresholds that constitute adherence, which may follow exploratory investigations of adherence for specific MHapps. Taken together, the literature suggests that operationalizations of user engagement and adherence vary. Subsequently, this may create difficulties in comparing findings across studies.

Along with variability in the measurement of these constructs, evidence for a relationship between user engagement and adherence appears mixed. Baumel and colleagues (2018) included expert ratings of engagement within an analysis of user data from 52 MHapps. Engagement was defined as the degree to which the app design would attract users as measured by ratings of content presentation, interactivity, customization, personalization, and ability to capture user interest. Of these engagement variables, content presentation, personalization, and interesting areas of the app were significantly associated with increased app usage. An additional analysis of user data within 56 MHapps suggested that there was not a relationship between expert-rated engagement and app usage (Kaveladze et al., 2022). Specifically, the engagement scale of the Mobile App Rating Scale (MARS) was not found to be correlated with continued app usage. These findings may be attributed to the utilization of expert-rated, rather than self-reported, measures of engagement.
Preliminary evidence exists in favor of a relationship between user engagement and outcomes, although much of the current evidence solely focuses on user adherence. Bakker and Rickard (2018) found that self-reported app engagement ratings were predictive of decreases in depressive and anxious symptoms for a self-monitoring app. Additionally, Zhang and colleagues (2019) examined a suite of apps for depression and anxiety. App usage was divided into four quartiles for data analysis, with the third quartile being described as moderate use. Results indicated that moderate usage of skill-building and goal-setting features was predictive of decreased depressive symptoms. Furthermore, they found that greater overall app usage and usage of a tracking feature predicted reduced depressive symptoms. Ludtke and colleagues (2018) found minimal differences between the intervention and waitlist control groups within an RCT focusing on an eclectic self-help app. However, participants in the intervention group that used the app at least several times per week experienced larger reductions in depressive symptoms compared to those in the control group.

In contrast, results of other studies have failed to support a relationship between user adherence and outcomes. While reductions in depressive and anxious symptoms were significant within the intervention group in two RCTs, overall app usage was not associated with outcomes (Graham et al., 2020; Moberg et al., 2019). Moberg and colleagues (2019) examined the usage of specific app features (e.g., thought record, meditation exercises) in relation to outcomes. No usage of specific features was associated with reduced depressive symptoms. Within a suite of apps used in primary care, no association was found between changes in depressive symptoms and total number of app logins, days the apps were used, and total length of time between first and
last app use (Graham et al., 2020). While MHapps may have the capability to reduce depressive symptoms, the relationship between MHapp usage and outcomes remains unclear.

A closer examination of app usage may allow for the discovery of dose-response relationships. Within these relationships, a specific amount of app usage would be associated with better outcomes. A focus on dose-response relationships appears particularly salient as data indicates that many individuals discontinue their app use within several weeks of downloading an app. An analysis of app store user data, including 93 MHapps, indicated that 80% of users stop using the app within 10 days (Baumel et al., 2019). When recommended by a healthcare provider, MHapps have the lowest 30-day retention rate compared to other health and wellness apps (Agarwal et al., 2015). Within RCTs, decreased rates of app use have been reported in as little as two weeks (Anguera et al., 2015; Arean et al., 2016; Roepke et al., 2015). According to a recent meta-analysis, dropout rates within 18 RCTs examining MHapps for depressive symptoms were estimated to be 47.8%, which was higher than dropout within non-app active control groups (14.2%; Torous et al., 2020).

In summary, user engagement has been conceptualized in various ways across previous literature. User adherence has been examined as a part of user engagement, as well as a separate construct. Evidence for a relationship between engagement and outcomes appears mixed, yet app usage tends to decrease within a brief period. It is therefore crucial to further examine the relationship between user engagement and outcomes in a brief period, which may lead to the discovery of dose-response relationships.
Guidance

Guidance is another delivery-related factor that might impact MHapp effectiveness for depression. Guidance involves support that is provided to users by a researcher, therapist, or coach and is typically delivered to users via text message or phone call. Potential goals of guidance may involve addressing questions, delivering feedback, providing emotional support, enhancing motivation, and developing recommendations for users. Subsequently, the format and purpose of guidance vary across studies.

Within meta-analyses, evidence for a relationship between guidance and outcomes appears unclear. In a meta-analysis of RCTs investigating MHapps, larger effect sizes were present within studies that offered guidance and reminders to engage with the app (Linardon et al., 2019). However, these results may not extend to depression apps. According to a meta-analysis examining depression apps, studies incorporating feedback from researchers had a nonsignificant effect size ($g=0.14$), while studies without in-person feedback had a moderate positive effect size ($g=0.47$; Firth et al., 2017).

Similarly, weak evidence suggests that there may be a relationship between guidance and MHapp effectiveness. An eclectic suite of 14 apps for depression and anxiety has been examined within primary care (Graham et al., 2020) and community samples with guidance (Mohr et al., 2017). Specific apps were recommended to users during coaching sessions. Within the community sample (Mohr et al., 2017), all participants were provided with the intervention. Over 8 weeks, participants experienced a significant reduction in depressive symptoms with a large effect size ($d=1.4$). Within
the primary care sample (Graham et al., 2020), participants were randomized to an intervention arm or a waitlist control arm. Participants in the intervention arm had lower depressive symptoms compared to those in the waitlist control after 8 weeks ($d=0.78$). Results indicated that there was a weak correlation between guidance and depressive symptoms ($r=-0.07$), suggesting that guidance had a small effect on efficacy.

Evidence from the broader scope of digital health interventions for depression may further inform the potential relationship between guidance and the efficacy of MHapps. Digital health interventions that incorporate therapeutic guidance ($g=0.63$) have been found to be more efficacious than standalone interventions ($g=0.34$) according to a meta-analytic review of 83 RCTs (Moshe et al., 2021). Furthermore, moderate effect sizes were seen within 9 studies of digital interventions for depression that involved therapist contact, while studies without therapist involvement only reached small effect sizes (Garrido et al., 2019).

Overall, while MHapps are effective as standalone self-help, delivery with guidance has the potential to further improve outcomes. Currently, mixed evidence exists regarding the role of guidance in MHapp efficacy. In part, these inconclusive findings may be attributed to the various goals, formats, and providers of guidance. Previous studies have indicated that the delivery of guidance with digital interventions enhances outcomes, yet it remains unclear whether this relationship extends to MHapps. Thus, the relationship between guidance and app effectiveness will be examined in the present study.
User Expectations and Preferences

Psychotherapy research focusing on key predictors of client outcomes may extend to MHapps, indicating how these interventions can be tailored to maximize effectiveness. Within the literature on psychotherapy, findings demonstrate that client expectancy for improvement is predictive of symptomatic reduction (Kim et al., 2015; Thiruchselvam et al., 2019; Thompson-Hollands et al., 2014; Tsai et al., 2014). Client expectancy effects account for an estimated 15% of variance in treatment outcomes (Norcross & Lambert, 2011). Preliminary evidence also indicates that user expectations may influence outcomes for digital and non-digital self-help (Boettcher et al., 2013; Cavanagh et al., 2009; Cludius et al., 2018). Results from a randomized trial of an internet-based program for social anxiety disorder indicated that higher expectations for improvement predicted symptomatic change and treatment adherence (Boettcher et al., 2013). Moreover, Cludius and colleagues (2018) examined treatment expectancies within self-help manuals for depression by altering the framing of the study. Participants that were informed the study would focus on a therapeutic treatment experienced a significant reduction in depressive symptoms, which was not present in the cognition framing condition. Expectancy for symptomatic improvement may therefore be predictive of MHapp efficacy.

Additionally, client preferences have a modest effect on outcomes within psychotherapy. Across meta-analyses, outcomes were compared between participants that were matched with their preferred treatment and those that remained unmatched. Swift and Callahan (2009) conducted a meta-analysis and found a small significant effect of treatment preferences on mental health outcomes (r=0.15). Measured treatment preferences included therapeutic approach (e.g., mindfulness), treatment type (e.g.,
psychotherapy, pharmacotherapy), and therapy format (e.g., individual, group).

Participants who were matched with their preferred treatment demonstrated a 58% chance of improvement. Within a meta-analysis of 9 RCTs, treatment preferences between psychotherapy and pharmacotherapy were found to affect outcomes (ES=0.23) for those experiencing mental health conditions (Delevry & Le, 2019). Lindheim and colleagues (2014) conducted a meta-analysis of 32 clinical trials that focused on treatment preferences or patient choice of treatment. Results indicated that matched treatment preferences or choices may be associated with improved treatment completion (OR=1.37), treatment satisfaction (d=0.34), and outcomes (d=0.15). Thus, matched treatment preferences improve outcomes in psychotherapy.

Preference matching has also been associated with treatment adherence for primary care patients with depression. Raue and colleagues (2009) examined preference matching within sixty primary care patients experiencing at least mild depressive symptoms. Participants were asked to rank-order a variety of potential treatments, then were randomized to either remain matched or unmatched to their preferred treatment. Treatments offered to participants included antidepressant medication and brief interpersonal psychotherapy. Patients with a greater strength of preference toward their assigned treatment demonstrated higher treatment adherence rates after 12 weeks. However, matching to a preferred treatment was unrelated to reductions in depressive symptoms and strength of preference was associated with higher levels of depressive symptoms after 12 weeks.

Preliminary evidence regarding the relationship between user preferences and digital health appears promising. Johansson and colleagues (2013) conducted a pilot
preference trial examining internet-based interventions for depression. Participants (N=44) were asked whether they preferred to receive Internet-based psychodynamic or cognitive behavioral therapy, along with the strength of their preference. Each participant was then matched to their preferred intervention. Exploratory analyses indicated that preference matching could predict outcomes and may affect treatment adherence.

However, the measurement of preferences may differ when applied to MHapps due to variations in intervention characteristics. Previously, preferences have been observed between traditional types of treatment such as psychotherapy or pharmacotherapy. However, preferences relevant to MHapps may include treatment elements (e.g., CBT, mindfulness, acceptance and commitment) and intervention modality (e.g., mobile app, computerized program, book). These differences in the operationalization of preference should be considered when examining whether preferences may play a role in MHapp effectiveness.

In summation, user preferences and expectations are factors that may influence the process through which MHapps are effective due to their association with psychotherapy outcomes. Broader applications to digital health and self-help have provided promising preliminary evidence for a relationship between these user-level variables and outcomes, yet differences in the operationalization of treatment preferences may impact whether these findings extend to MHapps. Further investigation of user preferences and expectations for symptomatic improvement appears justified, especially given the potential for these variables to inform provider decisions and tailor MHapps to meet user needs.
**Coping Skills**

The development of coping skills may provide explanatory value for the effectiveness of MHapps. Changes in cognitive behavioral skills have not yet been observed for MHapps, while several studies have observed mindfulness skills as an outcome variable. Within a randomized controlled trial of two mindfulness apps, increases in mindfulness and decreases in depressive symptoms were observed in college students (Flett et al., 2018). Similarly, a randomized controlled trial of a mindfulness app found increases in mindfulness and decreases in depressive and anxiety symptoms in individuals with compulsory internet behaviors (Quinones & Griffiths, 2019). However, depressive symptoms and mindfulness skills have solely been examined as separate outcome variables. It remains unclear whether decreases in depressive symptoms are associated with increases in mindfulness skills.

Along with the development of specific coping skills, app efficacy may also be associated with broader increases in coping self-efficacy. Qualitative evidence indicates that users retain coping skills to practice independently of their app (Pung et al., 2018). Indeed, coping self-efficacy significantly mediated app efficacy in reducing depressive symptoms within an RCT comparing cognitive behavioral apps (Bakker et al., 2018). Increases in self-efficacy have also been observed within a randomized controlled trial comparing an integrative app for depression and anxiety to a waitlist control condition (Moberg et al., 2019). Further examination of the relationship between engagement, self-efficacy, and symptomatic reduction therefore appears warranted.
Aims of the Present Study

The first aim is to replicate previous literature by examining the effectiveness of a mindfulness and a cognitive behavioral app in the treatment of depressive symptoms. Mindfulness was selected because guided and unguided mindfulness exercises can readily be adapted to a digital format, and mindfulness MHapps are being developed frequently (Mani et al., 2015). Furthermore, preliminary evidence supports the efficacy of mindfulness apps for the treatment of depression. For example, Gal and colleagues (2021) conducted a meta-analysis of 34 RCTs examining mindfulness apps for treating a variety of mental health conditions. Examination of the 15 RCT’s that examined depression found that apps had small, yet significant effects ($g=0.33$). Effect sizes were similar across studies that included waitlist controls ($g=0.35$) and active psychological controls ($g=0.28$). Therefore, mindfulness-based apps may be efficacious self-help interventions.

The second app was based on CBT, which was selected because CBT is also easily adaptable to a mobile delivery format and there is significant evidence to indicate that CBT is an effective treatment for depression (Cuijpers et al., 2013). Within a recent systematic review of 8 studies that examined MHapps developed using cognitive behavioral principles, findings across studies suggested that these interventions are efficacious in treating symptoms of depression (Rathbone et al., 2017).

In addition to examining the question of whether apps, regardless of their orientation, are effective at treating depression, we will also examine whether these two apps are differentially effective. Meta-analytic evidence suggests that there are not differences in outcomes for depression between apps with CBT and Mindfulness features
(Firth et al., 2017; Linardon et al., 2019). Given this evidence, it is hypothesized that depressive symptoms will decrease over time within the present study regardless of the type of app being used.

The effect of traditional treatment on outcomes will be additionally be observed given that this may affect intervention effectiveness. It is hypothesized that individuals who are receiving traditional care will experience greater reductions in depressive symptoms.

The **second aim** is to examine the relationship between user engagement and app effectiveness. Previous literature indicates that user engagement decreases over time, emphasizing the importance of investigating possible dose-response effects of MHapps (Baumel et al., 2019). Moreover, using a combination of self-report and objective measures to define engagement may clarify this relationship and address current limitations within the literature. Within the present study, user engagement will be measured comprehensively using both an objective measure of user adherence and a subjective measure of feelings of engagement with the app. Due to limited evidence regarding sound scientific justification for thresholds of engagement, no specific cutoff will determine whether a user is engaged within the present study. Rather, levels of engagement will be observed in relation to reductions in depressive symptoms. Based on previous research, it is hypothesized that greater engagement levels will be associated with decreases in depressive symptoms (Bakker & Rickard, 2018; Zhang et al., 2019).

The **third aim** is to investigate whether guidance is related to reductions in depressive symptoms and user engagement. Previous studies have demonstrated that
guidance is often associated with improved outcomes (Linardon et al., 2019) and lower dropout rates (Torous et al., 2020). Within the present study, optional guidance will be offered to participants. Based on this literature, it is hypothesized that the utilization of guidance will result in increased user engagement and decreases in depressive symptoms.

The fourth aim is to examine whether user expectations and preferences are related to reductions in depressive symptoms and user engagement. These variables have held explanatory value for client change and treatment adherence within psychotherapy. Thus, user expectations and preferences could be related to symptomatic reduction for mental health app users. The present study will be the first to investigate whether these findings may extend to MHapps. Based on preliminary evidence within examinations of internet-based programs, it is hypothesized that greater user expectancy for change and matched preferences will be associated with decreases in depressive symptoms (Boettcher et al., 2013; Johansson et al., 2013).

The final aim is to examine whether coping skills are related to decreases in depressive symptoms. To date, no studies have examined the relationship between depressive symptoms and mindfulness or cognitive-behavioral skills development using MHapps. The present study will include measures of mindfulness and cognitive-behavioral skills to investigate whether changes in coping skills are associated with reduced depressive symptoms. Additionally, increases in coping self-efficacy have been found to be related to greater MHapp efficacy (Bakker et al., 2018). Within the present study, the relationship between mindfulness, cognitive-behavioral skills, coping self-efficacy, and depressive symptoms will be investigated. Based on this literature, it is
hypothesized that increases in each of these coping skills will be related to greater symptomatic reduction over time.
Chapter 2

Method

Participants

Recruitment flow is displayed in Figure 1. 589 individuals completed the pre-screen survey, 137 of which were eligible to participate. The most common reasons for exclusion were suspected bot activity \((n=432)\), citizenship outside of the U.S. \((n=16)\), being under 18 years of age \((n=2)\), previous use of an app included in the study \((n=1)\), and current pregnancy \((n=1)\). Twenty participants were recruited in the final sample. Thirteen participants were assigned to the mindfulness app condition and 7 participants were assigned to the CBT condition. Two participants were lost to follow-up, both of whom were assigned to the mindfulness app condition. Seven participants completed the follow-up survey. Sample characteristics are displayed in Table 1. The sample was predominantly white \((n=13)\), non-Hispanic/Latino(a) \((n=13)\), and female \((n=17)\). Eighty percent of participants had received at least some college education or higher. Participants ranged in age from 18 to 40 \((M=24.70; \ SD=6.837)\).
**Figure 1**

*Participant Flow*

- 589 individuals completed online prescreen
  - Excluded through screening
    - Bot-like activity \(n=432\)
    - Not a U.S. citizen \(n=16\)
    - Under 18 years old \(n=2\)
    - Use of study app \(n=1\)
    - Currently pregnant \(n=1\)
  - 137 individuals were eligible
  - 20 participants were randomized
    - CBT App \(n=7\)
    - Mindfulness App \(n=13\)
      - Lost to post-assessment \(n=2\)
<table>
<thead>
<tr>
<th>Baseline Characteristic</th>
<th>Mindfulness</th>
<th></th>
<th>CBT</th>
<th></th>
<th>Full Sample</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>11</td>
<td>85</td>
<td>6</td>
<td>85</td>
<td>17</td>
<td>85</td>
</tr>
<tr>
<td>Non-binary/Genderqueer/Gender non-conforming</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>14</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>Male</td>
<td>1</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>8</td>
<td>62</td>
<td>5</td>
<td>71</td>
<td>13</td>
<td>65</td>
</tr>
<tr>
<td>Black or African American</td>
<td>3</td>
<td>23</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>Asian American</td>
<td>1</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>14</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic/Latino(a/x)</td>
<td>3</td>
<td>23</td>
<td>2</td>
<td>29</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>Non-Hispanic/Latino(a/x)</td>
<td>9</td>
<td>69</td>
<td>4</td>
<td>57</td>
<td>13</td>
<td>65</td>
</tr>
<tr>
<td><strong>Highest educational level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school</td>
<td>2</td>
<td>15</td>
<td>2</td>
<td>29</td>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td>Some college</td>
<td>4</td>
<td>31</td>
<td>2</td>
<td>29</td>
<td>6</td>
<td>30</td>
</tr>
<tr>
<td>Associate degree</td>
<td>3</td>
<td>23</td>
<td>1</td>
<td>14</td>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td>Baseline Characteristic</td>
<td>Mindfulness</td>
<td></td>
<td>CBT</td>
<td></td>
<td>Full Sample</td>
<td></td>
</tr>
<tr>
<td>----------------------------</td>
<td>-------------</td>
<td>-----------</td>
<td>-----</td>
<td>-----------</td>
<td>-------------</td>
<td>-----------</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>2</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>Master’s degree</td>
<td>1</td>
<td>7</td>
<td>2</td>
<td>29</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>Professional degree</td>
<td>1</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Employment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-time</td>
<td>4</td>
<td>31</td>
<td>1</td>
<td>14</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>Part-time</td>
<td>4</td>
<td>31</td>
<td>3</td>
<td>43</td>
<td>7</td>
<td>35</td>
</tr>
<tr>
<td>Unemployed</td>
<td>4</td>
<td>31</td>
<td>3</td>
<td>43</td>
<td>7</td>
<td>35</td>
</tr>
<tr>
<td>Income</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than $14,999</td>
<td>3</td>
<td>23</td>
<td>3</td>
<td>43</td>
<td>6</td>
<td>30</td>
</tr>
<tr>
<td>$15,000-$24,999</td>
<td>2</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>$25,000-$49,999</td>
<td>4</td>
<td>31</td>
<td>1</td>
<td>14</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>$50,000-$74,999</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>14</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>More than $75,000</td>
<td>1</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>5</td>
</tr>
</tbody>
</table>

**Inclusion and Exclusion Criteria**

To be included in the study, participants: a) owned a mobile device and have access to the internet; b) were at least 18 years of age; c) were English-speaking; and d)
were a U.S. citizen. These eligibility criteria were selected to facilitate recruitment efforts and allow for participant compensation in compliance with university policies. Individuals that are currently pregnant were excluded as this population requires additional protections within research studies.

**Recruitment**

Participants were recruited using several strategies. First, nearly twice as many college students experience depression when compared to the general population (American College Health Association, 2018). Therefore, undergraduates enrolled in an Essentials of Psychology course at Rowan University were recruited using an online participant pool management system. Second, participants were recruited via online advertisements on social media platforms (e.g., Reddit, Facebook), postings on the Rowan University student daily announcements, and flyers posted around the Rowan University main campus. Study recruitment occurred from February 2023 to May 2023.

To determine eligibility, individuals completed an online pre-screen questionnaire with the Patient Health Questionnaire-8 (PHQ-8; Kroenke et al., 2010). Eligible individuals were contacted via email to schedule a virtual meeting with a research assistant on WebEx, a video conferencing platform. Each participant was provided with information regarding the purpose of the study, rights and expectations for participation, and potential risks and benefits. Informed consent was obtained with a written or digital signature.
Procedure

Participants met with a research assistant in a baseline session to first complete a questionnaire online, which included questions regarding demographic information, treatment preferences, current treatment status, previous usage of MHapps, the PHQ-8, the Depression Change Expectancy Scale (DCES; Eddington et al., 2014), the Five Facet Mindfulness Questionnaire-Short Form (FFMQ-SF; Bohlmeijer et al., 2011), the Cognitive-Behavioral Therapy Skills Questionnaire (CBTSQ; Jacob et al., 2011), and the Coping Self-Efficacy Scale (CSE Scale; Chesney et al., 2006). The PHQ-8 was selected to measure symptoms due to good sensitivity for the detection of Major Depressive Disorder (MDD) and widespread use in healthcare systems. The DCES was included to measure the degree to which participants believe their depressive symptoms will change. The FFMQ-SF and CBTSQ were included to measure potential mechanisms of change that could be targeted by an app. While the FFMQ-SF allowed for changes to be detected in mindfulness, the CBTSQ can act as a control measure as these skills are expected to remain unchanged. Finally, the CSE scale will allow for changes in coping self-efficacy to be measured, which may mediate symptomatic change (Bakker et al., 2018).

After taking the baseline survey, participants were instructed to download their assigned app and provided with a brief video tutorial with instructions about how to use the app. At the end of the meeting, participants were asked to use the app however much they naturally would over the following two weeks in order to align with naturalistic patterns of app usage. If participants did not own a smartphone with a built-in usage tracker, they were asked to download StayFree (StayFree, n.d.), a publicly available app usage tracker. An app usage tracker allowed for the objective measurement of app usage.
As guidance can increase the efficacy of self-help, participants had the option to schedule a phone call with a research assistant weekly. During these phone calls, participants could ask any questions they may have about their app or the study. If they preferred, they could complete the survey while speaking with the researcher. One week after baseline (T2), participants completed a survey including the PHQ-8, User Engagement Scale-Short Form (UES-SF; O’Brien et al., 2018), FFMQ-SF, CBTSQ, CSE Scale, questions about daily app usage, treatment status, and qualitative questions about their experience. Given that the UES-SF demonstrated a theoretical rationale for the operationalization of user engagement, it was selected to allow participants to self-report their feelings of engagement toward an app. One week after T2 (T3), participants scheduled a second phone call and completed an identical survey. Following the completion of the third survey, participants had the opportunity to enter a raffle for a $50 gift card. Two gift cards were awarded to participants. One month after completing the T3 survey, participants had the opportunity to complete a follow-up survey including the PHQ-8, UES-SF, FFMQ-SF, CBTSQ, CSE Scale, and questions about their experience with the app.

**Intervention**

Mindfulness and cognitive behavioral therapy (CBT) are empirically supported treatments for depression as defined by the American Psychological Association Society for Clinical Psychology website (Society of Clinical Psychology, 2022). Mindfulness Coach is an app developed by the Veteran’s Administration that was selected based on consistency with this approach. The app contains 12 guided mindfulness exercises, progress tracking, goal setting, and information about mindfulness. Guided mindfulness
exercises include awareness of the senses, breathing exercises, loving-kindness meditation, mindful eating, mindful walking, mindful listening, and progressive muscle relaxation. Users can practice unguided mindfulness with the ability to customize the length of exercise time. The app additionally contains a gamification component that allows users to receive badges while progressing through levels of mindfulness mastery.

MoodTools is an app developed by Inquiry Health that was selected for consistency with CBT. The app contains psychoeducation, monitoring, cognitive restructuring, behavioral activation, and safety planning features. In the behavioral activation section of the app, users can select preset activities (e.g., exercise, socializing) or custom activities to complete. Users additionally have the option to have the app choose a mood-boosting activity for them. Users rate their mood before and after completing each activity on a scale from 0 (Horrible) to 10 (Ecstatic). Within the cognitive restructuring section of the app, users complete a digital thought diary. Users begin by describing the situation, logging their emotions, and rating their level of distress. Next, users describe their negative thoughts, indicate which cognitive distortions might be present in their negative thoughts, and challenge their negative thoughts. The app provides descriptions of each cognitive distortion and examples of challenges to negative thoughts to guide users through completion of the thought diary. Lastly, users log alternative thoughts about the situation and rate their distress.
Measures

The following demographic information was collected from each participant: age, gender, race, ethnicity, income, and treatment status. Along with demographic information, the following variables were measured:

*The Patient Health Questionnaire-8 (PHQ-8)*

The PHQ-8 is an 8-item self-report questionnaire that assesses 8 of the 9 DSM-5 symptom criteria for Major Depressive Disorder (MDD). Each item is rated on a 4-point scale from 0 (Not at all) to 3 (Nearly every day). Scores on the PHQ-8 may range from 0 to 24 and a cut-off score of 10 is used to indicate a probable DSM-5 diagnosis of MDD. The PHQ-8 demonstrates good sensitivity and specificity for the detection of MDD (Kroenke et al., 2010).

*Mobile Application Usage*

Length and frequency of mobile application usage will be assessed via a 14-item self-report questionnaire designed by the research team. Each participant will report how many times they opened the app and how many minutes they used the app each day over the last week. This data will be self-reported by referencing an app usage tracker installed on participants’ smartphones.

*User Engagement Scale – Short Form (UES-SF)*

The UES-SF is a 12-item self-report questionnaire that measures the level of user engagement with technology. Each item is rated on a 5-point scale ranging from 1 (Strongly disagree) to 5 (Strongly agree). There is support for a four-factor model of the UES, with the subscales including aesthetic appeal, focused attention, perceived usability, and reward (O’Brien et al., 2018). The four-factor model of the UES has specifically
been replicated within a mobile health context (Holdener et al., 2020). Subscale scores are calculated via mean item ratings, which can then be summed to determine an overall engagement score (O’Brien et al., 2018). Reliability for each of the 4 subscales was good and each subscale correlated well with the remaining items in the original UES (O’Brien et al., 2018).

**Depression Change Expectancy Scale (DCES)**

The DCES is a 20-item self-report questionnaire that measures the extent to which a person believes their depression can change. Each item is rated on a 5-point scale ranging from 1 (*Strongly disagree*) to 5 (*Strongly agree*). Scores may range from 20 to 100, with a higher score indicating greater expectations for change. The DCES demonstrates excellent internal consistency, good convergent validity, good discriminant validity, and differential predictive validity of short-term outcomes (Eddington et al., 2014).

**Treatment Preferences**

Treatment preferences will be assessed using 6 items developed by the research team. Each participant will report their preference for types of treatment (e.g., psychotherapy, pharmacotherapy, self-help), modality of self-help (e.g., app, bibliotherapy), and evidence-based approach to treatment (e.g., mindfulness, cognitive-behavioral). Strength of treatment preferences will be rated on a 5-point scale ranging from 1 (Not Strong) to 5 (Very Strong).

**Five Facet Mindfulness Questionnaire-Short Form (FFMQ-SF)**

The FFMQ-SF is a 24-item self-report questionnaire measuring the five facets of mindfulness, which include observing, describing, acting with awareness, non-judging, and nonreactivity. Items are rated on a 5-point scale ranging from 1 (never or very rarely
true) to 5 (very often or always true). Confirmatory factor analysis demonstrated that the FFMQ-SF fit a correlated five-factor structure (Bohlmeijer et al., 2011). The FFMQ-SF has adequate internal consistency, good convergent validity, and good discriminant validity (Bohlmeijer et al., 2011).

**Cognitive-Behavioral Therapy Skills Questionnaire (CBTSQ)**

The CBTSQ is a 16-item self-report questionnaire measuring client skills developed during behavioral activation and cognitive restructuring interventions. Items are rated on a 5-point Likert scale ranging from 1 (I don’t do this) to 5 (I always do this). Confirmatory factor analysis supported a two-factor structure, with Behavioral Activation and Cognitive Restructuring subscales. The CBTSQ demonstrated good internal consistency and criterion validity (Jacob et al., 2011).

**The Coping Self-Efficacy Scale (CSE Scale)**

The CSE scale is a 13-item self-report questionnaire that measures perceived self-efficacy for coping with challenges. Items are rated on an 11-point Likert scale ranging from 0 (Cannot do at all) to 10 (Certain can do). There is support for a three-factor structure of the CSE Scale, which include using problem-focused coping, stopping unpleasant emotions and thoughts, and getting support from friends and family. The CSE scale demonstrates high internal consistency, high test-retest reliability, and good concurrent, convergent, discriminant, and predictive validity (Chesney et al., 2006).

**Data Analysis**

Descriptive analyses were conducted in SPSS 28 and further analyses were conducted in R Studio. Data with two bootstrapped cases sampled from the preliminary dataset (N=18) was used to develop a data analytic plan. The analysis was conducted
using a rough confirmatory approach. Hypotheses were specified in advance, with room for flexibility in analytic decisions based on patterns that emerged from the dataset (Fife & Rodgers, 2019).

Each study aim was investigated using mixed models (hierarchical linear models), which natively handle nonindependent data, in the lme4 package (Bates et al., 2015). Models were compared using nested model comparisons (Rodgers, 2010). With the exception of the first aim, models for each variable of interest were compared to baseline model. A combination of model selection criteria (e.g., AIC, BIC, Bayes factor) and visual inspection with the Flexplot package (Fife, 2021) were used to evaluate model fit. Within each model, fixed slopes and random intercepts were fit. Across the models, the ICC was 0.519 and the design effect was 1.857.

Before performing the analysis, the assumptions of the statistical models were evaluated using visuals from the R package flexplot. According to residual plots (Appendix A), the residuals appear to meet the assumption of normality and heteroscedasticity for all models. For all models beyond those created in the first aim, non-linearity may be present. The addition of non-linear terms to the models did not reduce non-linearity and were not favored by model comparison statistics. Therefore, non-linear terms were not included in the final models for each aim.

First, a model was constructed with PHQ scores. A baseline model was then created by including time of measurement and PHQ scores to determine whether depressive symptoms changed over the duration of the study. To evaluate the first aim, a reduced model containing PHQ scores, time of measurement, and condition assignment (mindfulness or CBT) was constructed. A full model including PHQ scores, time of
measurement, condition, treatment, and interactions between condition and time, treatment and time, and condition and treatment was constructed. Model comparisons were conducted between baseline, reduced, and full models.

To examine the second aim, a model of objective engagement was created using PHQ scores, time of measurement, minutes of app usage, and the interaction between minutes of app usage and time. One participant was excluded from the analysis in this model as an outlier. Additionally, a model of subjective engagement was created using PHQ scores, time of measurement, UES score, and the interaction between UES scores and time. Model comparisons were conducted between each model and the reduced model. No analysis was conducted for the third aim as no participants opted to receive guidance.

For the fourth aim, a model of preference was constructed using PHQ scores, time of measurement, strength of preference for treatment approach (e.g., mindfulness, CBT), and the interaction between preference and time. A model of expectancy for changes in depressive symptoms was also created using PHQ scores, time of measurement, total DCES scores, and the interaction between DCES scores and time. Model comparisons were conducted between each model and the baseline model.

To investigate the fifth aim, a model of mindfulness skills was created using PHQ scores, time of measurement, total FFMQ scores, and the interaction between FFMQ scores and time. A model of CBT skills included PHQ scores, time of measurement, total CBTSQ scores, and the interaction between CBTSQ scores and time. A model of coping skills was constructed using PHQ scores, time of measurement, CSE scores, and the
interaction between CSE scores and time. Model comparisons were conducted using each skill model and the reduced model.
Chapter 3

Results

Effectiveness for Depressive Symptoms

Descriptive statistics for PHQ scores at each time of measurement are located in Table 2. Across the sample, the mean PHQ score decreased by approximately 4 points from T1 to T3. Fifty percent ($n=10$) of participants experienced a decrease in PHQ score of at least five points or higher. Participants were additionally asked how much their depressive symptoms interfered with or caused difficulty in their lives. Impairment scores fluctuated, with the mean rating decreasing from T1 to T2, then increasing from T2 to T3. Mean scores for total PHQ score and impairment ratings remained similar through follow-up measurement. Before determining whether additional variables accounted for changes in depressive symptoms over time, a baseline model was created to examine whether participants were experiencing symptomatic reduction over time. PHQ scores were included in the Baseline model and PHQ scores and time of measurement were included in the Time model.
Table 2

*Mean PHQ Scores Over Time*

<table>
<thead>
<tr>
<th></th>
<th>Time 1 M (SD)</th>
<th>Time 2 M (SD)</th>
<th>Time 3 M (SD)</th>
<th>Follow-Up M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total PHQ-8 Score</td>
<td>12.80 (5.123)</td>
<td>11.13 (6.365)</td>
<td>8.82 (6.274)</td>
<td>7.86 (3.933)</td>
</tr>
<tr>
<td>Impairment</td>
<td>3.00 (0.858)</td>
<td>2.56 (0.892)</td>
<td>2.76 (1.091)</td>
<td>2.29 (0.488)</td>
</tr>
</tbody>
</table>

Table 3

*Model Comparison Between the Baseline and Time (With Time) Model*

<table>
<thead>
<tr>
<th></th>
<th>aic</th>
<th>bic</th>
<th>bayes factor</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>332.168</td>
<td>338.079</td>
<td>0.059</td>
<td>0.006</td>
</tr>
<tr>
<td>Time</td>
<td>322.551</td>
<td>332.403</td>
<td>17.085</td>
<td></td>
</tr>
</tbody>
</table>
Figure 2

Model Comparison of PHQ Scores Over Time

*Note.* The blue line shows the fit for the Time model (Depression and time), while the red line shows the fit for the Baseline model (Depression only).

Table 3 shows the statistical estimates for the model comparisons, while Figure 2 plots the predictions of the two models. The model comparison statistics appear to favor the Time model over the Baseline model. The Bayes factor suggests that the evidence
favoring the Time model is approximately 17 times as strong as that favoring the Baseline model and the aic additionally factors the Time model. The p-value suggests that there is a difference between the models. Compared to the Baseline model, the Time model explains 22.4% more of the residual variance and no additional slope variance. Figure 2 shows the relationship between depressive symptoms and time of measurement for 10 randomly selected participants. The Time plot appears to demonstrate a difference in model fit, with depressive symptoms decreasing steadily over time.

Table 4

*Fixed and Random Effects for the Time Model*

<table>
<thead>
<tr>
<th>Source</th>
<th>Fixed</th>
<th>Random</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>10.817</td>
<td>4.556</td>
</tr>
<tr>
<td>Time (L;Q)*</td>
<td>-2.987;</td>
<td>-0.316</td>
</tr>
<tr>
<td>Residual</td>
<td>3.706</td>
<td></td>
</tr>
</tbody>
</table>

*L = linear; Q = quadratic*

Table 4 displays the fixed and random effects for the Time model. Over time, there was a -2.99 fixed linear effect and a -0.316 fixed quadratic effect for depressive symptoms. Figure 3 shows a visual of the model and displays the relationship between
depressive symptoms and time for 10 randomly sampled clusters. The fixed effects within the figure appear to demonstrate that there is a decrease in depressive symptoms from T1 to T3. The random intercept effect indicates that on average, there is an 4.556 point deviation between intercepts of residual variance between participants.

Figure 3

*PHQ Scores Over Time*

*Note.* Each dotted line represents the relationship for a randomly sampled cluster.
**Condition and Treatment**

Participants were asked whether they were receiving treatment from an outside provider, such as psychotherapy or medication. Twelve participants received outside treatment during their study enrollment, while eight did not. Of those who received treatment, 6 participants received psychotherapy, 2 participants received medication, and 4 participants received a combination of both approaches.

The Time model was used as a baseline model to evaluate whether condition and treatment were associated with changes in depressive symptoms. The reduced model included PHQ scores, condition assignment (mindfulness or CBT), and time of measurement. The full model included PHQ scores, condition assignment, whether outside treatment was received during study participation (e.g., psychotherapy, medication), time of measurement, and an interaction between condition and treatment. Compared to the baseline model, the reduced model does not explain any additional residual variance and the full model explains 3% more of the residual variance. The full model explains 24.7% of the residual variance and no additional slope variance.

Table 5 shows the statistical estimates for the model comparisons, while Figure 4 and Figure 5 plot the predictions of the two models. Within the model comparison between the baseline and reduced models, the statistics appear ambiguous. The Bayes factor is close to one for both models, suggesting that there is comparable evidence for both models. Additionally, the p-value suggests that the models are similar. However, the aic favors the reduced model. Figure 4 shows the relationship between study condition and depressive symptoms for ten randomly selected participants. There appear to be slight differences in the fit of the model, with the CBT group experiencing a greater overall reduction in depressive symptoms and greater reductions from T2 to T3 in the
reduced plot. The reduced plot also appears to show that the mindfulness group experienced greater reductions in symptoms from T1 to T2. While it appears unclear which model is favored, condition will be included in additional model comparisons and readers may interpret estimates accordingly.

The model comparison statistics appear to favor the full model over the reduced model. The Bayes factor suggests that the evidence favoring the full model is approximately 10 times as strong as that favoring the reduced model and the aic favors the full model. However, there is only a small difference between the models based on the p-value. Figure 5 shows the relationship between condition, treatment, and depressive symptoms over time for ten randomly selected participants. The full model plot demonstrates a difference in model fit based on treatment when compared to the reduced model plot in most panels. Individuals who did not receive outside treatment appeared to experience greater reductions in depressive symptoms over time.
Table 5

Model Comparisons Between 1) the Reduced (With Condition and Time) and Baseline Model and 2) the Reduced (With Condition and Time) and Full Model (With Condition, Treatment, and Time)

<table>
<thead>
<tr>
<th></th>
<th>aic</th>
<th>bic</th>
<th>bayes factor</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reduced x Baseline</td>
<td>322.551</td>
<td>332.403</td>
<td>0.994</td>
<td>0.50</td>
</tr>
<tr>
<td>Baseline</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>316.629</td>
<td>332.391</td>
<td>1.006</td>
<td></td>
</tr>
<tr>
<td>Reduced x Full</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>316.629</td>
<td>332.391</td>
<td>0.098</td>
<td>0.168</td>
</tr>
<tr>
<td>Condition_Treatment</td>
<td>304.111</td>
<td>327.754</td>
<td>10.160</td>
<td></td>
</tr>
</tbody>
</table>
Figure 4

*PHQ Scores and Condition Model Comparison*

*Note.* The blue line shows the fit for the reduced model (condition and time), while the red line shows the fit for the baseline model (time only).
Figure 5

PHQ Scores, Condition, and Treatment Model Comparison

Note. The red line shows the fit for the reduced model (condition), while the blue line shows the fit for the full model (condition and treatment).
Table 6

**Fixed and Random Effects for the Full Model, Which Includes Depression, Time, Condition, and Treatment**

<table>
<thead>
<tr>
<th>Source</th>
<th>Fixed*</th>
<th>Random</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>11.533</td>
<td>4.600</td>
</tr>
<tr>
<td>Treatment</td>
<td>-3.562</td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>1.952</td>
<td></td>
</tr>
<tr>
<td>Time (L;Q)**</td>
<td>-3.198; -1.837</td>
<td></td>
</tr>
<tr>
<td>Condition x Time (L;Q)</td>
<td>2.268; 1.409</td>
<td></td>
</tr>
<tr>
<td>Treatment x Time (L;Q)</td>
<td>-2.596; 1.394</td>
<td></td>
</tr>
<tr>
<td>Condition x Treatment</td>
<td>-1.274</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td></td>
<td>3.650</td>
</tr>
</tbody>
</table>

*Only fixed effects were modeled for condition and treatment, as these variables did not change over time.

**L = linear; Q = quadratic**

Table 6 displays the fixed and random effects for the reduced model. According to the table, relative to those participants who were receiving treatment, participants who were not receiving treatment had an average difference of 3.562 in PHQ scores. Relative to those in the CBT app condition, those in the mindfulness app condition had an average 1.952 point difference in PHQ score. Figure 6 shows a visual of the model and displays the relationship between condition, treatment, and depressive symptoms for 10 randomly sampled clusters. The fixed effects within the figure appear to demonstrate that participants all experienced a decrease in depressive symptoms. However, those who did not receive outside treatment experienced a greater reduction in symptoms. While
participants in the mindfulness app condition who did not receive treatment experienced a greater reduction in depressive symptoms from T1 to T2, participants in the CBT app condition who did not receive treatment experienced consistent reductions in depressive symptoms across timepoints. Participants who were receiving treatment in the mindfulness app condition appeared to experience the lowest reduction in depressive symptoms, whereas participants in the CBT app condition appeared to experience a greater overall decrease in symptoms between T2 and T3.
Engagement and Effectiveness

Descriptive statistics for user engagement at each time of measurement are presented in Table 7. Within the Mindfulness app condition, the mean minutes of app usage remained consistent across timepoints, while the mean number of logins decreased from T2 to T3. In the CBT app condition, the mean number of minutes increased from T2 (M=11.0) to T3 (M=48.5), while the mean number of logins decreased by approximately
3. Within each study condition, the means of total UES scores remained consistent. However, the mean UES score for participants in the Mindfulness app condition was higher (M=3.49, 3.5) than the mean UES score for participants in the CBT app condition (M=3.0, 3.14). At each time of measurement, the UES aesthetic appeal and reward subscales appear to exhibit the largest differences in mean UES subscale scores between study conditions.

### Table 7

*Mean Engagement Scores Over Time Separated by Condition*

<table>
<thead>
<tr>
<th></th>
<th>Time 2 M (SD)</th>
<th>Time 3 M (SD)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>CBT</td>
<td>M</td>
<td>CBT</td>
</tr>
<tr>
<td>Minutes of App Usage</td>
<td>39.77 (22.74)</td>
<td>11.0 (7.67)</td>
<td>40.1 (39.07)</td>
<td>48.57 (83.82)</td>
</tr>
<tr>
<td>Number of App Logins</td>
<td>9.40 (9.19)</td>
<td>6.50 (3.27)</td>
<td>5.90 (6.09)</td>
<td>5.71 (2.69)</td>
</tr>
<tr>
<td>UES Total Score</td>
<td>3.49 (0.57)</td>
<td>3.0 (0.41)</td>
<td>3.50 (0.79)</td>
<td>3.14 (0.65)</td>
</tr>
<tr>
<td>UES Perceived Usability</td>
<td>3.77 (0.80)</td>
<td>4.0 (0.67)</td>
<td>4.27 (0.60)</td>
<td>4.0 (0.67)</td>
</tr>
<tr>
<td>UES Focused Attention</td>
<td>3.10 (0.67)</td>
<td>2.17 (0.78)</td>
<td>2.67 (0.98)</td>
<td>2.52 (0.79)</td>
</tr>
<tr>
<td>UES Aesthetic Appeal</td>
<td>3.50 (0.67)</td>
<td>2.78 (0.89)</td>
<td>3.27 (1.18)</td>
<td>2.81 (0.74)</td>
</tr>
<tr>
<td>UES Reward</td>
<td>3.60 (0.90)</td>
<td>3.05 (0.74)</td>
<td>3.8 (0.93)</td>
<td>3.24 (0.98)</td>
</tr>
</tbody>
</table>
To evaluate whether engagement was related to outcomes, two models were created. The Minutes model included the objective measure of minutes of app usage, time of measurement, and the interaction between minutes of app usage and time. The UES model included the subjective measure of total UES score, time of measurement, and the interaction between UES score and time. Each model was compared to the Condition model, which included study condition, time, and their interaction. Compared to the Condition model, neither the Minutes nor UES model explained any additional residual variance. The UES model explained an additional 2.85% of slope variance compared to the Condition model. Table 8 shows the statistical estimates for the model comparisons, while Figure 7 and Figure 8 plot the predictions of the two models.

Within the model comparison between the baseline and Minutes models, the statistics appear to favor the Condition model. The Bayes factor suggests the evidence favoring the Condition model is approximately 3,835 times stronger than the evidence favoring the Minutes model. Similarly, the aic favors the Condition model. However, the p-value suggests that the models are similar. Figure 7 shows the relationship between minutes of app usage, study condition, and depressive symptoms for ten randomly selected participants. There appears to be little difference between the plots of the Baseline and Minutes models across study conditions.

The model comparison statistics for the UES and Condition models appear ambiguous. The Bayes factor is close to 1, suggesting that there is nearly equal evidence favoring each model and the aic favors the UES model. There is additionally only a small difference between the models based on the p-value. Figure 8 shows the relationship between UES scores and depressive symptoms over time for ten randomly selected
participants. There appears to be a difference in model fit, with individuals with higher UES scores experiencing greater initial levels of depressive symptoms across study conditions in the UES model plot.

Table 8

Model Comparisons Between 1) the Reduced (With Condition and Time) and Minutes Model (With Minutes, Condition, and Time) and 2) the Reduced (With Condition and Time) and UES Model (With UES Score, Condition, and Time)

<table>
<thead>
<tr>
<th></th>
<th>aic</th>
<th>bic</th>
<th>bayes factor</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minutes x Baseline</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>294.628</td>
<td>309.924</td>
<td>3835.112</td>
<td>0.855</td>
</tr>
<tr>
<td>Minutes</td>
<td>307.308</td>
<td>326.428</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>UES x Baseline</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>316.629</td>
<td>332.391</td>
<td>1.024</td>
<td>0.595</td>
</tr>
<tr>
<td>UES</td>
<td>312.736</td>
<td>332.439</td>
<td>0.976</td>
<td></td>
</tr>
</tbody>
</table>
Note. The red line shows the fit for the reduced model (condition), while the blue line shows the fit for the Minutes model (condition and minutes).
**Figure 8**

*PHQ Scores, Condition, and UES Scores Model Comparison*

*Note.* The red line shows the fit for the reduced model (condition), while the blue line shows the fit for the UES model (condition and UES score).

As the model comparison statistics were ambiguous, model visuals and effects will be displayed to allow for readers to reach their own conclusions about findings.

Table 9 displays the fixed and random effects for the UES model. According to the table, each standard deviation increase in UES score is associated with an average 1.208 point increase in PHQ score. Figure 9 shows a visual of the model and displays the relationship...
between UES scores and depressive symptoms for 15 randomly sampled clusters. Within each condition, the fixed effects appear similar. Those with higher UES scores in each condition appeared to exhibit greater symptoms of depression, but the size of symptomatic reduction from T2 to T3 does not appear to have changed.

**Table 9**

*Fixed and Random Effects for the UES Model, Which Includes Depression, Time, Condition, and User Engagement*

<table>
<thead>
<tr>
<th>Source</th>
<th>Fixed</th>
<th>Random</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>7.842</td>
<td>4.567</td>
</tr>
<tr>
<td>UES</td>
<td>1.208</td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>0.032</td>
<td></td>
</tr>
<tr>
<td>Time (L;Q)*</td>
<td>-7.241; -0.607</td>
<td></td>
</tr>
<tr>
<td>UES x Time (L;Q)</td>
<td>0.392</td>
<td></td>
</tr>
<tr>
<td>Condition x Time (L;Q)</td>
<td>-0.573; 0.110</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>3.813</td>
<td></td>
</tr>
</tbody>
</table>

*L = linear; Q = quadratic*
Figure 9

*PHQ Scores, UES Scores, and Condition*

![Graph showing PHQ and UES scores over time across different conditions.](image)

*Note.* Each dotted line represents the relationship for a randomly sampled cluster.

**Guidance and Effectiveness**

No participants opted to receive guidance via a phone call from a member of the research team. Therefore, analyses were not conducted to determine whether a relationship exists between guidance and reductions in depressive symptoms. However, 70% (n=14) of participants indicated that if they were to receive self-help, they would
prefer for it to be guided. Strength of preferences were rated on a 5-point scale ranging from 1 *(Not Strong)* to 5 *(Very Strong)*. Participants who preferred guided self-help had a moderately strong preference (M=3.36).

**User Expectations, Preferences, and Effectiveness**

**User Expectations**

At baseline, participants had a mean total score of 70.9 (SD=12.91) on the DCES. While variability was present, this mean score suggests that participants overall endorsed the belief that their depressive symptoms could change. There was a mean of 38.85 (SD=7.77) for total endorsement of pessimistically-worded items and 31.65 (SD=5.51) for total endorsement of optimistically-worded items.

To determine if user expectations for improvement are associated with outcomes, an Expectancy model was created using PHQ scores, time of measurement, total DCES scores, and the interaction between DCES scores and time. Compared to the baseline model, the Expectancy model does not explain any additional residual variance, but explains 5.7% more of the slope variance.

Table 10 shows the statistical estimates for the model comparisons, while Figure 10 plots the predictions of the two models. Within the model comparison between the Baseline and Expectancy models, the statistics favor the Baseline model. The Bayes factor suggests the evidence favoring the Baseline model is 15,031 times stronger than the evidence favoring the Expectancy model. Similarly, the aic favors the Baseline model. The p-value indicates that there is not a large difference between the models. Figure 10 shows the relationship between expectancy for improvement and depressive symptoms for ten randomly selected participants. There appears to be a difference in
model fit for individuals with lower and higher expectations, while expectations for improvement in the middle has a similar fit to the baseline model. The Expectancy model appears to show that increasing expectations for improvement are associated with lower initial levels of symptoms and smaller decreases in symptom change over time. Given the model comparison statistics and small sample size, it is likely that these differences in model fit are negligible. As the Baseline model was favored within the model comparison statistics, fixed and random effects were not included for the variable of user expectancy.

Table 10

*Model Comparison Between the Baseline (With Time) and Expectancy Model (With Expectations and Time)*

<table>
<thead>
<tr>
<th></th>
<th>aic</th>
<th>bic</th>
<th>bayes factor</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>322.551</td>
<td>332.403</td>
<td>15031.5</td>
<td>0.430</td>
</tr>
<tr>
<td>Expectancy</td>
<td>335.876</td>
<td>351.638</td>
<td>0.000</td>
<td></td>
</tr>
</tbody>
</table>
Figure 10

*PHQ Score and Expectancy Model Comparison*

![Graph showing PHQ scores and expectancy model comparison](image)

*Note.* The red line shows the fit for the Baseline model (time of measurement), while the blue line shows the fit for the Expectancy model (time of measurement and expectancy).

**Preferences**

To determine if user preferences are associated with outcomes, a Preference model was created using PHQ scores, time of measurement, and strength of preference score for the treatment approach (i.e., CBT, mindfulness). Preference scores range from -5 to 5, with negative scores indicating that participants were not matched with their
preferred treatment approach. Eight participants were matched with their preferred
treatment approach, while 12 remained unmatched. Those who were unmatched with
their preferred approach had a mean score of -2.92 (SD=1.16) and those who were
matched had a mean score of 4 (SD=0.93). Compared to the baseline model, the
Preference model explains 2.47% more of the residual variance and 0.48% more of the
slope variance.

Table 11 shows the statistical estimates for the model comparisons, while Figure
11 plots the predictions of the two models. Within the model comparison between the
Baseline and Preference models, the statistics primarily favor the Baseline model. The
Bayes factor suggests the evidence favoring the Baseline model is 153 times stronger
than the evidence favoring the Preference model. Similarly, the aic favors the Baseline
model. The p-value suggests that the models are similar. Figure 11 shows the relationship
between preferences and depressive symptoms for ten randomly selected participants.
There appear to be slight differences in model fit. In the Preference model, while
individuals with strong matched preferences and weaker matched or unmatched
preferences experience larger improvements from T1 to T2, those with stronger
unmatched preferences do not appear to experience reductions in symptoms during that
time and instead experience a larger decrease in symptoms from T2 to T3. Given the
model comparison statistics and small sample size, it is likely that these differences in
model fit are minimal. As the Baseline model was favored within the model comparison
statistics, fixed and random effects were not included for the variable of user preferences.
Table 11

Model Comparison Between the Baseline (With Time) and Preference Model (With Preferences and Time)

<table>
<thead>
<tr>
<th></th>
<th>aic</th>
<th>bic</th>
<th>bayes factor</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>322.551</td>
<td>332.403</td>
<td>153.290</td>
<td>0.232</td>
</tr>
<tr>
<td>Preference</td>
<td>326.705</td>
<td>342.467</td>
<td>0.007</td>
<td></td>
</tr>
</tbody>
</table>

Figure 11

PHQ Score and Preference Model Comparison

Note. The red line shows the fit for the Baseline model (time of measurement), while the blue line shows the fit for the Preference model (time of measurement and preference).
Coping Skills and Effectiveness

Mindfulness Skills

Descriptive statistics for mindfulness at each time of measurement are presented in Table 12. Across both study groups, increases in the means of total FFMQ scores were seen. However, this increase was greater for participants in the CBT app condition. Within the mindfulness app condition, increases were seen in the mean scores of the describing, activating with awareness, and nonjudging subscales. Decreases were seen in the mean scores of the nonreacting and observing subscales for participants in the mindfulness app condition. In the CBT app condition, increases were seen in the mean scores of the describing, acting with awareness, nonjudging, and nonreacting subscales. A decrease was seen in the mean scores of the observing subscale for those in the CBT app condition.
Table 12

<table>
<thead>
<tr>
<th></th>
<th>Time 1 M (SD)</th>
<th>Time 2 M (SD)</th>
<th>Time 3 M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>CBT</td>
<td>M</td>
</tr>
<tr>
<td>Total FFMQ</td>
<td>68.30</td>
<td>(10.50)</td>
<td>66.20</td>
</tr>
<tr>
<td>Describe</td>
<td>17.00</td>
<td>(4.60)</td>
<td>17.30</td>
</tr>
<tr>
<td>Acting with</td>
<td>13.23</td>
<td>(5.53)</td>
<td>13.80</td>
</tr>
<tr>
<td>Awareness</td>
<td>13.57</td>
<td>(3.95)</td>
<td>14.00</td>
</tr>
<tr>
<td>Nonjudge</td>
<td>12.62</td>
<td>(4.23)</td>
<td>13.50</td>
</tr>
<tr>
<td>Nonreact</td>
<td>12.46</td>
<td>(3.48)</td>
<td>10.70</td>
</tr>
<tr>
<td>Observe</td>
<td>13.00</td>
<td>(3.46)</td>
<td>10.90</td>
</tr>
</tbody>
</table>

To determine if mindfulness skills are related to effectiveness, a Mindfulness model was created using PHQ scores, study condition, time of measurement, total FFMQ scores, and interactions between time and study condition, time and FFMQ scores, and FFMQ scores and study condition. The FFMQ variable was centered by converting the FFMQ total scores to z-scores to allow for the modeling of random slopes for this variable. Within a model comparison, including only fixed slopes for FFMQ scores was favored. This model was compared to the Condition model, which included PHQ scores, study condition, time of measurement, and the interaction between study condition and
time. The Mindfulness model explained 16.4% more of the residual variance and no additional slope variance when compared to the Condition model.

Table 13 shows the statistical estimates for the model comparisons and Figure 12 plots the predictions of the two models. Within the model comparison between the Baseline and Mindfulness models, the statistics favor the Mindfulness model. The Bayes factor suggests the evidence favoring the Mindfulness is model 2.8 times stronger than the evidence favoring the Condition model. Similarly, the aic favors the Mindfulness model and the p-value suggests that there is a large difference between the models.

Figure 12 displays the relationship between mindfulness skills and depressive symptoms for ten randomly selected participants. Across study conditions, there appears to be a difference in model fit for individuals with lower and higher mindfulness skills, while mindfulness scores in the middle have a similar fit to the baseline model. The Mindfulness model appears to show that while all participants appear to experience a similar decrease in depressive symptoms regardless of mindfulness skills, those with lower initial mindfulness skills experience higher levels of initial depressive symptoms. Similarly, those with higher initial mindfulness skills exhibit lower initial levels of depressive symptoms and appear to experience greater benefit from T1 to T2 than those with different FFMQ scores. However, the model fit appears similar for participants with FFMQ scores closer to the mean.
Table 13

*Model Comparison Between the Reduced (With Condition and Time) and Mindfulness Model (With FFMQ scores, Condition, and Time)*

<table>
<thead>
<tr>
<th></th>
<th>aic</th>
<th>bic</th>
<th>bayes factor</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition</td>
<td>316.629</td>
<td>332.391</td>
<td>0.350</td>
<td>0.025</td>
</tr>
<tr>
<td>Mindfulness</td>
<td>306.648</td>
<td>330.292</td>
<td>2.856</td>
<td></td>
</tr>
</tbody>
</table>
Figure 12

**PHQ Scores, FFMQ Scores, and Condition Model Comparison**

*Note.* The red line shows the fit for the reduced model (condition and time of measurement), while the blue line shows the fit for the Mindfulness model (time of measurement, condition, and FFMQ scores).

Table 14 displays the fixed and random effects for the Mindfulness model. For each standard deviation increase in FFMQ score, PHQ scores would decrease by 2.016 points. Figure 13 shows a visual of the model and displays the relationship between FFMQ scores, condition, and depressive symptoms for 10 randomly sampled clusters.
The fixed effects within the figure appear to demonstrate that across conditions, those with higher FFMQ scores demonstrated lower initial levels of depressive symptoms and greater reductions in depressive symptoms over time. At each level of FFMQ score, participants in the CBT app condition experienced greater reductions in depressive symptoms. However, these reductions occurred from T2 to T3, whereas participants in the mindfulness condition experienced steady reductions across all three times of measurement.

Table 14

*Fixed and Random Effects for the Full Model, Which Includes Depression, Time, Condition, and Mindfulness*

<table>
<thead>
<tr>
<th>Source</th>
<th>Fixed</th>
<th>Random</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>10.164</td>
<td>4.659</td>
</tr>
<tr>
<td>FFMQ</td>
<td>-2.016</td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>1.432</td>
<td></td>
</tr>
<tr>
<td>Time (L;Q)*</td>
<td>-3.846; -1.912</td>
<td></td>
</tr>
<tr>
<td>FFMQ x Time (L;Q)</td>
<td>-0.948; -0.394</td>
<td></td>
</tr>
<tr>
<td>Condition x Time (L;Q)</td>
<td>1.553; 2.266</td>
<td></td>
</tr>
<tr>
<td>FFMQ x Condition</td>
<td>-0.856</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td></td>
<td>3.406</td>
</tr>
</tbody>
</table>

*L = linear; Q = quadratic*
Figure 13

PHQ Scores, FFMQ Scores, and Condition

Note. Each dotted line represents the relationship for a randomly sampled cluster.

Cognitive Behavioral Skills

Descriptive statistics for CBT skills at each time of measurement can be found in Table 15. Across both study conditions, slight increases were observed in total CBTSQ mean scores. These small increases in mean scores were also present for the cognitive restructuring and behavioral activation subscales. The CBTSQ variable was centered by
converting the CBTSQ total scores to z-scores. Modeling the CBTSQ variable with random slopes resulted in singular fit, so CBTSQ was modeled solely using fixed slopes.

**Table 15**

*Mean CBTSQ Scores Over Time Separated by Study Condition*

<table>
<thead>
<tr>
<th></th>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
</tr>
<tr>
<td>Total CBTSQ</td>
<td>47.30 (13.1)</td>
<td>46.43 (6.37)</td>
<td>47.70 (14.54)</td>
</tr>
<tr>
<td>Cognitive</td>
<td>25.40 (8.31)</td>
<td>24.86 (7.22)</td>
<td>25.80 (9.20)</td>
</tr>
<tr>
<td>Restructuring</td>
<td>21.90 (5.63)</td>
<td>21.57 (4.04)</td>
<td>21.90 (5.82)</td>
</tr>
<tr>
<td>Behavioral</td>
<td>21.90 (5.63)</td>
<td>21.57 (4.04)</td>
<td>21.90 (5.82)</td>
</tr>
<tr>
<td>Activation</td>
<td>21.90 (5.63)</td>
<td>21.57 (4.04)</td>
<td>21.90 (5.82)</td>
</tr>
</tbody>
</table>

To determine if cognitive behavioral skills are related to effectiveness, a CBT model was created using PHQ scores, time of measurement, and total CBTSQ scores. Table 16 shows the statistical estimates for the model comparisons and Figure 14 plots the predictions of the two models. Within the model comparison between the Baseline and CBT models, the statistics appear ambiguous. The Bayes factor suggests the evidence favoring the Baseline model is approximately 12 times stronger than the evidence favoring the CBT model. However, the aic favors the CBT model. The p-value suggests that the models are similar. Figure 14 shows the relationship between CBT skills, study condition, and depressive symptoms for ten randomly selected participants. There appear
to slight differences in model fit within the CBT condition. While the fit is similar for
individual with CBTSQ scores close to the mean, those with lower initial CBTSQ scores
appear to exhibit higher initial levels of depressive symptoms and those with higher
initial CBTSQ scores appear to have lower levels of depressive symptoms. The size of
reductions in depressive symptoms appear similar regardless of CBTSQ score. Within the
Mindfulness app condition, the Baseline and CBT model plots appear similar. For
individuals with lower CBTSQ scores in the first panel, it appears that greater
improvement was experienced from T2 to T3 than T1 to T2 in the CBT model plot. The
CBT model explained no additional residual or slope variance when compared to the
Condition model.

Table 16

Model Comparison Between the Baseline (With Condition and Time) and CBT
Model (With CBTSQ Scores, Condition, and Time)

<table>
<thead>
<tr>
<th></th>
<th>aic</th>
<th>bic</th>
<th>bayes factor</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>316.629</td>
<td>332.391</td>
<td>12.566</td>
<td>0.77</td>
</tr>
<tr>
<td>CBT</td>
<td>313.809</td>
<td>337.453</td>
<td>0.080</td>
<td></td>
</tr>
</tbody>
</table>
Figure 14

PHQ Scores, CBTSQ Scores, and Condition Model Comparison

Note. The red line shows the fit for the Baseline model (time of measurement), while the blue line shows the fit for the CBT model (time of measurement and CBTSQ scores).

As the model comparison statistics were ambiguous, model visuals and effects will be displayed to allow for readers to reach their own conclusions about findings. Table 17 displays the fixed and random effects for the CBT model. For each standard deviation increase in CBTSQ score, PHQ scores would decrease by 2.616 points. Figure
15 shows a visual of the model and displays the relationship between CBTSQ scores, condition, and depressive symptoms for 10 randomly sampled clusters. The fixed effects within the figure appear to demonstrate that within the CBT condition, those with higher initial CBTSQ scores had lower initial levels of depressive symptoms and experienced greater reductions in depressive symptoms. Within the mindfulness condition, changes in depressive symptoms do not appear to be related to CBTSQ scores.

**Table 17**

*Fixed and Random Effects for the Full Model, Which Includes Depression, Time, Condition, and CBT Skills*

<table>
<thead>
<tr>
<th>Source</th>
<th>Fixed</th>
<th>Random</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>10.469</td>
<td>4.924</td>
</tr>
<tr>
<td>CBTSQ</td>
<td>-2.616</td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>0.558</td>
<td></td>
</tr>
<tr>
<td>Time (L;Q)*</td>
<td>-3.977; -1.663</td>
<td></td>
</tr>
<tr>
<td>CBTSQ x Time (L;Q)</td>
<td>-0.573; 0.110</td>
<td></td>
</tr>
<tr>
<td>Condition x Time (L;Q)</td>
<td>1.683; 2.148</td>
<td></td>
</tr>
<tr>
<td>CBTSQ x Condition</td>
<td>2.122</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td></td>
<td>3.774</td>
</tr>
</tbody>
</table>

*L = linear; Q = quadratic*
Figure 15

*PHQ Scores, CBTSQ Scores, and Condition*

*Note.* Each dotted line represents the relationship for a randomly sampled cluster.
Coping Self-Efficacy

Table 18

Mean CSE Scores Over Time Separated by Condition

<table>
<thead>
<tr>
<th></th>
<th>Time 1 M (SD)</th>
<th>Time 2 M (SD)</th>
<th>Time 3 M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>CBT</td>
<td>M</td>
</tr>
<tr>
<td>Total CSE</td>
<td>63.54</td>
<td>46.29</td>
<td>58.6</td>
</tr>
<tr>
<td>M (SD)</td>
<td>(21.33)</td>
<td>(19.69)</td>
<td>(23.7)</td>
</tr>
</tbody>
</table>

Descriptive statistics for coping self-efficacy at each time of measurement can be found in Table 18. Within both study conditions, mean CSE scale scores increased across times of measurement. The CSE variable was centered by converting the CSE total scores to z-scores. Modeling the CSE variable with random slopes resulted in singular fit, so CSE was modeled solely using fixed slopes.

To determine if coping self-efficacy is related to effectiveness, a CSE model was created using PHQ scores, time of measurement, and total CSE scores. Table 19 shows the statistical estimates for the model comparisons and Figure 16 plots the predictions of the two models. Within the model comparison between the Baseline and CSE models, the statistics appear ambiguous. The Bayes factor suggests the evidence favoring the Baseline model is approximately 5 times stronger than the evidence favoring the CSE model. However, the aic favors the CSE model. The p-value suggests that there are minimal differences between the models. Figure 16 shows the relationship between CSE scores and depressive symptoms for ten randomly selected participants. The Baseline and
CSE model plots appear similar within the CBT condition. In the Mindfulness condition, there appears to be a difference in model fit in the left panel. It appears that individuals with lower coping self-efficacy experienced higher levels of depressive symptoms at each time point. The CSE model explained no additional residual or slope variance when compared to the Condition model.

Table 19

*Model Comparison Between the Baseline (With Condition and Time) and CSE Model (With CSE Scores, Condition, and Time)*

<table>
<thead>
<tr>
<th></th>
<th>aic</th>
<th>bic</th>
<th>bayes factor</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>316.629</td>
<td>332.391</td>
<td>4.994</td>
<td>0.338</td>
</tr>
<tr>
<td>CSE</td>
<td>311.964</td>
<td>335.607</td>
<td>0.200</td>
<td></td>
</tr>
</tbody>
</table>
**Figure 16**

*PHQ Scores, CSE Scores, and Condition Model Comparison*

```
<table>
<thead>
<tr>
<th>CSE:</th>
<th>CSE:</th>
<th>CSE:</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.8-(-0.8)</td>
<td>(-0.8)-0.26</td>
<td>0.26-1.6</td>
</tr>
</tbody>
</table>

```

*Note.* The red line shows the fit for the Baseline model (time of measurement), while the blue line shows the fit for the CSE model (time of measurement and CSE scores).

As the model comparison statistics were ambiguous, model visuals and effects will be displayed to allow for readers to reach their own conclusions about findings. Table 20 displays the fixed and random effects for the CSE model. For each standard deviation increase in CSE score, PHQ scores would decrease by 1.111 points. Figure 17 shows a visual of the model and displays the relationship between CSE scores, condition,
and depressive symptoms for 10 randomly sampled clusters. The fixed effects within the figure appear to demonstrate that within the CBT condition, those with higher initial CSE scores had lower initial levels of depressive symptoms and experienced greater reductions in depressive symptoms. Within the mindfulness condition, changes in depressive symptoms do not appear to be related to CSE scores.

Table 20

*Fixed and Random Effects for the Full Model, Which Includes Depression, Time, Condition, and Coping Self-Efficacy*

<table>
<thead>
<tr>
<th>Source</th>
<th>Fixed</th>
<th>Random</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>10.250</td>
<td>4.65</td>
</tr>
<tr>
<td>CSE</td>
<td>-1.111</td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>1.080</td>
<td></td>
</tr>
<tr>
<td>Time (L;Q)*</td>
<td>-4.051; -2.599</td>
<td></td>
</tr>
<tr>
<td>CSE x Time (L;Q)</td>
<td>-0.576; -1.441</td>
<td></td>
</tr>
<tr>
<td>Condition x Time (L;Q)</td>
<td>1.825; 3.432</td>
<td></td>
</tr>
<tr>
<td>CSE x Condition</td>
<td>-0.152</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td></td>
<td>3.774</td>
</tr>
</tbody>
</table>

*L = linear; Q = quadratic*
Figure 17

*PHQ Scores, CSE Scores, and Condition*

*Note.* Each dotted line represents the relationship for a randomly sampled cluster.
Chapter 4

Discussion

Areas of Focus and Future Directions

Effectiveness for Depressive Symptoms

The primary aim of the study was to investigate variables that may influence the effectiveness of apps for depression. Within our sample, participants experienced a decrease in depressive symptoms over two weeks as hypothesized based on meta-analytic evidence (Firth et al., 2017; Linardon et al., 2019). However, the size and rate of decreases in depressive symptoms depended on both the study condition to which participants were assigned and whether they were engaged in outside treatment.

The model comparison results were ambiguous in determining whether study condition alone was associated with changes in depressive symptoms. Visually, it appears that participants assigned to the CBT app condition experienced greater reductions in depressive symptoms than those in the mindfulness app condition. Across the psychotherapy literature, mixed findings exist on the efficacy of CBT in comparison to other therapeutic approaches for the treatment of depression. While some meta-analyses indicate that comparable efficacy exists, others suggest that CBT is more efficacious than other techniques (Hofmann et al., 2012). It is additionally possible that efficacy of specific therapeutic approaches is related to the initial severity of depressive symptoms. Within a comparison of a Behavioral Activation (BA) app and a Mindfulness app, the BA app was more effective for individuals with more severe initial depressive symptoms and the mindfulness app was more effective for individuals with lower initial depressive symptoms (Ly et al., 2014). Continuing to examine the effectiveness of MHapps with
specific therapeutic techniques may continue to inform who may benefit most from these approaches.

Contrary to our hypotheses, individuals receiving treatment experienced smaller reductions in depressive symptoms. Individuals receiving outside care may have experienced less benefit given that the standard of care is more efficacious for the treatment of depression than MHapps (Firth et al., 2017; Linardon et al., 2019; De Maat et al., 2006; Munder et al., 2019). MHapp effectiveness for the treatment of depression has additionally been moderated by psychotropic medication, with individuals taking psychotropic medication experiencing less intervention benefit (Moberg et al., 2019). When directly compared to traditional psychotherapy alone, MHapps used in a blended treatment have not been demonstrated to be more effective (Ly et al., 2015). Thus, offering MHapps in addition to traditional care may not produce additional benefit in relation to outcomes. Future research may continue to investigate whether the addition of MHapps may contribute to the effectiveness of traditional treatment.

The rate of symptomatic improvement also appeared to be related to study condition and outside treatment. Participants in the mindfulness app condition experienced greater improvements in the first week of using their app that plateaued during the second week, regardless of their engagement with outside treatment. Those in the CBT app condition who received outside treatment experienced greater improvements in the second week, while participants who did not receive outside treatment experienced similar decreases in symptoms during both weeks. Future investigations may continue to examine rates of symptomatic reduction, especially in relation to dose-response effects.
**Engagement and Effectiveness**

Within our second aim, model comparisons did not indicate that a relationship was present between minutes of app usage and effectiveness. Additionally, it was ambiguous whether subjectively rated engagement was associated with symptomatic reduction beyond the effect of study condition. Visually, our data suggest that lower levels of subjectively rated engagement may have been associated with higher levels of depressive symptoms. There also did not appear to be an association between the size of symptomatic reduction and subjectively related engagement. In the wider literature on MHapps that were not included in the present study, findings regarding a relationship between engagement and depressive symptoms remain mixed (Bakker & Rickard, 2018; Graham et al., 2020; Moberg et al., 2019; Zhang et al., 2019). The relationship between engagement and MHapp effectiveness can continue to be investigated in future studies to determine whether dose-response effects may exist.

**Guidance and Effectiveness**

As no participants opted to receive guidance, there was not sufficient data that could be used to investigate the third aim. However, the majority of participants (n=14) indicated that if they were to use self-help, they would prefer to receive it in a guided format. Guidance was presented to participants as optional, which may have led to less uptake. Participants had access to an app video tutorial designed by the research team throughout the duration of the study and therefore may not have needed further guidance about using their app. The format of guidance offered via a phone call may have impacted uptake as well. It is possible that participants may have preferred written guidance initiated by researchers. Future studies may consider incorporating different
levels and formats of guidance to determine how this variable may relate to effectiveness, especially as findings on the impact of guidance on the effectiveness of MHapps for the treatment of depression remain mixed (Firth et al., 2017; Graham et al., 2020; Linardon et al., 2019).

**User Expectations, Preferences, and Effectiveness**

Results did not support the existence of a relationship between user expectations, preferences, and effectiveness within the fourth aim. In each model comparison, the baseline models were favored. These user-level variables may not affect outcomes. This study remains the first to examine whether these variables were related to MHapp effectiveness. One limitation may have included a narrow operationalization of preference that only consisted of treatment approach. Different dimensions of preference, such as modality of self-help and other forms of intervention (e.g., traditional treatment) may represent different ways that this variable could be operationalized. Additionally, expectations for improvement may have been elevated within the sample due to self-selection bias and the presentation of MHapps as interventions that could be used to reduce depressive symptoms. Future investigations of preferences applied to the effectiveness of MHapps may compare different types of preferences that are specific to MHapps, frame interventions in different ways to examine expectancy effects, and include populations in more naturalistic settings.

**Coping Skills and Effectiveness**

Within our fifth aim, findings are the first to our knowledge that directly demonstrate an association between changes in mindfulness skills and the effectiveness
of MHapps. However, contrary to our hypothesis, increases in mindfulness skills were observed across study conditions. Additionally, model comparison results were ambiguous in determining whether CBT skills were related to outcomes. These results may parallel trends within the psychotherapy literature. Wampold (2001) differentiated between common (i.e., therapeutic relationship) and specific (i.e., treatment approach) factors as mechanisms of change in the contextual model. Across the psychotherapy literature, common factors have a stronger association with therapeutic change compared to specific treatment approaches or ingredients (Wampold, 2015). Therefore, explanatory mechanisms for MHapp effectiveness may span across therapeutic approaches. It is recommended that skill development continue to be examined to better elucidate the processes through which MHapps are effective.

Model comparison results were ambiguous in determining whether a relationship exists between coping self-efficacy and MHapp effectiveness. Coping self-efficacy has been conceptualized as a part of client expectations in psychotherapy, which is a common factor that influences outcomes (Wampold, 2015). Coping self-efficacy is influenced by the belief that the therapeutic approach will be beneficial for the presenting concern that is influenced by the therapist. Yet not all common factors may directly translate from the psychotherapy literature to internet-based approaches (Mogoșe et al., 2016). MHapps may not present a tailored rationale that indicates why a particular approach may be helpful for a user, thereby influencing coping self-efficacy. While coping self-efficacy has been associated with improvements in depressive symptoms, the existence of this relationship may depend on the explanation of the therapeutic approach used within MHapps (Bakker et al., 2018). Future research may continue to investigate whether a
relationship is present between coping skill development and symptomatic improvement, especially while considering how the presentation of the intervention may impact changes in coping self-efficacy.

**Clinical Implications**

While the generalizability of results to clinical practice remains limited given the sample size, there are several ways in which findings may have practical implications. First, findings suggest that rate of symptomatic change may vary based on the therapeutic approach used in an MHapp. Clinicians may consider recommending that patients use CBT-based apps for at least two weeks prior to re-evaluating outcomes. Second, our data suggest that incorporating MHapps into existing treatment may provide limited benefit compared to the delivery of MHapps alone. Our findings may imply that the integration of MHapps into traditional treatment may not have an additive effect on efficacy. It is possible that clinician-led integration of MHapps into care may produce different outcomes. Not only may clinicians provide guidance on how to use an MHapp, but techniques presented within MHapps can be blended with the clinician’s approach. For participants who received psychotherapy, it remains unclear whether the techniques presented in the MHapps were complementary or contradictory to the approaches used by therapists. Future research may begin to investigate the processes through which MHapps may best be integrated into clinical practice, as data does not yet exist on the optimal way to integrate these practices into care based on clinician and client perceptions.
Limitations

Several limitations exist within the present study. First, the generalizability of our conclusions is limited. A small sample size was collected partially due to funding constraints, meaning that results should be interpreted with caution. This sample size limits both the extent of the conclusions that can be drawn from analyses and the generalizability of results to MHapp users. In our analysis, many nested model comparisons did not yield substantial differences between models. It is possible that using larger samples could demonstrate larger differences between models and relationships that differ from those found in our sample. Demographically, the sample was predominantly female, had a young mean age, and was highly educated. It remains unclear whether results may generalize to different genders, adults older than 40, and individuals with a high school level of education at most. The present study also solely focuses on two apps and thus has limited generalizability to the app marketplace. While both apps used in the study are publicly available, determining whether results may generalize to additional MHapps appears pertinent. Overall, replication of the present design within a larger sample appears necessary.

Second, the internal validity of the study was low. A control group was not present, meaning that improvement could be attributed to spontaneous remission of depressive symptoms. A control group was not included in the study as the existing evidence suggests that MHapps are efficacious interventions for depression when compared to control conditions (Firth et al., 2017; Linardon et al., 2019). Results may also be interpreted in terms of the magnitude of symptomatic change on the PHQ-8. While the mean decrease in depressive symptoms did not reach the clinically significant threshold of 5 points (Jacobson & Truax, 1991), the mean score for the sample was above
the clinical cut-off score for depression at T1 and below the cut-off score at T3. Half of the sample also experienced clinically significant change with a reduction of 5 points or greater. Similarly to the existing literature, our findings appear to support that individuals who use MHapps may experience benefits. Within one randomized controlled trial, participants in the waitlist control condition experienced a 0.5 point decrease in PHQ score over four weeks (Moberg et al., 2019). However, the relative benefit experienced to a control condition for our sample remains unknown. Furthermore, broad sample inclusion criteria could also limit our ability to determine whether changes in depressive symptoms could be attributed to the adoption of an MHapp. Other variables (e.g., the presence of bipolar or psychotic disorders) may have instead influenced outcomes. Participants were eligible regardless of whether they were receiving outside treatment, which makes it difficult to determine if symptomatic reductions were attributed to their treatment or the MHapps. It is thus recommended that replication efforts contain additional controls.

Finally, the measurement of depression and engagement in our study may present another limitation. Depression was operationalized using a sum score on the PHQ-8 given that this measure is widely used within healthcare settings, yet specific symptoms may have differential impacts on functioning and the measurement of individual symptoms may instead provide informative results (Fried & Nesse, 2015). In the context of mobile health, mood and functional impairment may be investigated as outcomes. Operationalizing depression in this way may also require different methods of measurement. By measuring participant symptoms and engagement at more frequent time points (e.g., daily) using methods such as ecological momentary assessment, the
relationship between depressive symptoms and engagement may be better understood. While our study included both an objective and subjective measure of user engagement, heterogeneity in the operationalization of user engagement exists across the scientific literature and creates difficulty in contextualizing findings (Ng et al., 2019). Future literature should strive to begin achieving a consensus on how to define engagement, especially by using theory to inform study design.

**Conclusion**

In summary, MHapps are an accessible and effective alternative delivery method for the treatment of depression. Our results appear consistent with the wider MHapp literature as participants experienced a decrease in depressive symptoms. Trends in our data suggest that effectiveness was related to assigned study condition and the reception of external treatment. Additionally, subjective user engagement experiences may have influenced outcomes, while no relationship was seen between objective engagement and improvement. Increases in mindfulness skills were associated with reductions in depressive symptoms regardless of study condition; this relationship may have been present for CBT skills and coping self-efficacy. Hopefully, continued investigation of these promising interventions may optimize the delivery of MHapps for the treatment of depression.
References


Appendix A

Diagnostics

Figure A1

*Time Model*
Figure A2

Condition Model

Histogram of Residuals

S-L Plot
Figure A3

Condition & Treatment Model

Histogram of Residuals

S-L Plot
Figure A4

Minutes Model
Figure A5

UES Model
Figure A6

Preference Model

Histogram of Residuals

Residual Dependence Plot

S-L Plot
Figure A7

*Expectancy Model*

![Histogram of Residuals](Image)

![Residual Dependence Plot](Image)

![S-L Plot](Image)
Figure A8

Mindfulness Model

Histogram of Residuals

Residual Dependence Plot

S-L Plot

Absolute Value of Residuals

fitted
Figure A9

*CBT Skill Model*

![Histogram of Residuals](image1)

![Residual Dependence Plot](image2)

![S-L Plot](image3)
Figure A10

Coping Self-Efficacy Model
Appendix B

R Script

# importing datasets

library(haven)

JH_2022_1_Baseline_April_27_2023_12_40 <- read_sav("C:/Users/Danielle/Downloads/JH+2022-1+Baseline_April+27,+2023_12.40/JH 2022-1 Baseline_April 27, 2023_12.40.sav")

JH_2022_1_T2_5_9 <- read_sav("C:/Users/Danielle/Downloads/JH+2022-1+T2_May+9,+2023_09.34/JH 2022-1 T2 5.9.sav")

JH_2022_1_T3_5_9 <- read_sav("C:/Users/Danielle/Downloads/JH+2022-1+T3_May+9,+2023_10.51/JH 2022-1 T3 5.9.sav")

# tidying datasets

library(tidyverse)

Merge1 = merge(JH_2022_1_Baseline_April_27_2023_12_40, JH_2022_1_T2_5_9, by = "ID", all=TRUE, all.x = TRUE)

Merge2 = merge(Merge1,JH_2022_1_T3_5_9,by = "ID", all=TRUE, all.x=TRUE)

Merge3 = merge(JH_2022_1_T2_5_9,JH_2022_1_T3_5_9,by = "ID", all=TRUE, all.x=TRUE)
PHQ = Merge2 %>%
  select(ID, Treatment, Condition, Preference, DCES_Tot, starts_with("PHQ")) %>%
  pivot_longer(cols=contains("PHQ"), names_to = "Time", values_to = "PHQ",
  values_drop_na = FALSE) %>%
  mutate(Time=gsub("PHQ_T1", "1", Time))%>%
  mutate(Time=gsub("PHQ_T2", "2", Time))%>%
  mutate(Time=gsub("PHQ_T3", "3", Time))%>
  mutate(Time=factor(Time, ordered=T))

GAD = Merge2 %>%
  select(ID, Treatment, Condition, Preference, DCES_Tot, starts_with("GAD")) %>%
  pivot_longer(cols=contains("GAD"), names_to = "Time", values_to = "GAD",
  values_drop_na = FALSE) %>%
  mutate(Time=gsub("GAD_T1", "1", Time))%>%
  mutate(Time=gsub("GAD_T2", "2", Time))%>%
  mutate(Time=gsub("GAD_T3", "3", Time))%>
  mutate(Time=factor(Time, ordered=T))
CSE = Merge2 %>%

  select(ID, Treatment, Condition, Preference, DCES_Tot, starts_with("ZCSE_T")) %>%

  pivot_longer(cols=contains("CSE"), names_to = "Time", values_to = "CSE",
  values_drop_na = FALSE) %>%

  mutate(Time=gsub("ZCSE_T1", "1", Time))%>

  mutate(Time=gsub("ZCSE_T2", "2", Time))%>

  mutate(Time=gsub("ZCSE_T3", "3", Time))%>

  mutate(Time=factor(Time, ordered=T))

FFMQ = Merge2 %>%

  select(ID, Treatment, Condition, Preference, DCES_Tot, starts_with("ZFFMQ_T"))

  pivot_longer(cols=contains("FFMQ"), names_to = "Time", values_to = "FFMQ",
  values_drop_na = FALSE) %>%

  mutate(Time=gsub("ZFFMQ_T1", "1", Time))%>

  mutate(Time=gsub("ZFFMQ_T2", "2", Time))%>
CBTSQ = Merge2 %>%
  select(ID, Treatment, Condition, Preference, DCES_Tot, starts_with("ZCBTSQ_T")) %>%
  pivot_longer(cols=contains("CBTSQ"), names_to = "Time", values_to = "CBTSQ", values_drop_na = FALSE) %>%
  mutate(Time=gsub("ZCBTSQ_T1", "1", Time)) %>%
  mutate(Time=gsub("ZCBTSQ_T2", "2", Time)) %>%
  mutate(Time=gsub("ZCBTSQ_T3", "3", Time)) %>%
  mutate(Time=factor(Time, ordered=T))

Mins = Merge2 %>%
  select(ID, Treatment, Condition, Preference, DCES_Tot, starts_with("Min")) %>%
  pivot_longer(cols=contains("Min"), names_to = "Time", values_to = "Min", values_drop_na = FALSE) %>%
mutate(Time=gsub("Mins_T1", "1", Time))%>%

mutate(Time=gsub("Mins_T2", "2", Time))%>%

mutate(Time=gsub("Mins_T3", "3", Time)) %>%

mutate(Time=factor(Time, ordered=T))

UES = Merge2 %>%

select(ID, Treatment, Condition, Preference, DCES_Tot, starts_with("UES_T")) %>%

pivot_longer(cols=contains("UES"), names_to = "Time", values_to = "UES",
values_drop_na = FALSE) %>%

mutate(Time=gsub("UES_T1", "1", Time))%>%

mutate(Time=gsub("UES_T2", "2", Time))%>%

mutate(Time=gsub("UES_T3", "3", Time))%>%

mutate(Time=factor(Time, ordered=T))

Join1 = full_join(PHQ, GAD, by = c("ID", "Time", "Treatment", "Condition",
"Preference", "DCES_Tot"))
Join2 = full_join(Join1, CSE, by = c("ID", "Time", "Treatment", "Condition", "Preference", "DCES_Tot"))

Join3 = full_join(Join2, FFMQ, by = c("ID", "Time", "Treatment", "Condition", "Preference", "DCES_Tot"))

Join4 = full_join(Join3, Mins, by = c("ID", "Time", "Treatment", "Condition", "Preference", "DCES_Tot"))

Join5 = full_join(Join4, UES, by = c("ID", "Time", "Treatment", "Condition", "Preference", "DCES_Tot"))

d = full_join(Join5, CBTSQ, by = c("ID", "Time", "Treatment", "Condition", "Preference", "DCES_Tot"))

d = d %>%

  mutate(Treatment = as.character(Treatment))

d5 = d %>%

  filter(ID!=12)

d3 = d %>%

  filter(ID!=10)

#bootstrapping dataset
n <- nrow(d)

Boot = d[sample(x = n, size = 75, replace = TRUE), ]

n <- nrow(d2)

Boot2 = d2[sample(x = n, size = 50, replace = TRUE), ]

# Preliminary visuals & baseline model

library(flexplot)

library(lme4)

Baseline = lmer(PHQ~1 + (1|ID), data=d)

Time = lmer(PHQ~1 + Time+(1|ID), data=d)

visualize(Time, plot="residuals")

compare.fits(PHQ~Time, data=d, Baseline, Time, clusters=10)

model.comparison(Baseline, Time)

visualize(Time, plot="model", sample=10)

estimates(Time)
# renaming the Time model for use in further aims

baseline = lmer(PHQ~1 + Time+(1|ID), data=d)

# aim one: are apps effective & does app condition/outside treatment affect decreases in sx?

# reduced model

Condition = lmer(PHQ~Condition + Time+ Time:Condition + (1|ID), data=d)

visualize(Condition, plot="residuals")

compare.fits(PHQ~Time|Condition, data=d, Condition, baseline, clusters=10)

model.comparison(baseline, Condition)

# full model (originally modeled without an interaction between condition x treatment)

Condition_Treatment = lmer(PHQ~Treatment + Condition + Time + Time:Condition + Time:Treatment+Treatment:Condition+(1|ID), data=d)

visualize(Condition_Treatment, plot="residuals")

compare.fits(PHQ~Time|Treatment+Condition, data=d, Condition_Treatment, Condition, clusters=10)
model.comparison(Condition, Condition_Treatment)

model.comparison(baseline, Condition_Treatment)

visualize(Condition_Treatment, plot = "model",
formula=PHQ~Time+ID|Treatment+Condition, sample=10)

estimates(Condition_Treatment)

#interaction between condition x treatment was favored

int = lmer(PHQ~Treatment + Condition + Time + Time:Condition + Time:Treatment + 
Condition:Treatment+(1|ID), data=d)

compare.fits(PHQ~Time|Treatment+Condition, data=d, Condition_Treatment, int,
clusters=10)

model.comparison(Condition_Treatment, int)

#aim two: are levels of user engagement related to effectiveness?

Minutes = lmer(PHQ~Condition+Time:Condition+Min + Time+ Min:Time +(1|ID),
data=d3)

visualize(Minutes, plot="residuals")

#running Condition model again without outlier
Condition = lmer(PHQ~Condition + Time+ Time:Condition + (1|ID), data=d3)

compare.fits(PHQ~Time|Min+Condition, data=d, Minutes, Condition, clusters=10, jitter=c(-.5,.5))

model.comparison(Condition, Minutes)

UES = lmer(PHQ~Condition+ Condition:Time+UES + Time+ UES:Time+ (1|ID),
data=d)

visualize(UES, plot="residuals")

visualize(UES, plot="model", formula=PHQ~Time+ID|UES, sample=10)

compare.fits(PHQ~Time|UES, data=d, UES, Condition, clusters=10, jitter=c(-.5,.5))

model.comparison(Condition, UES)

estimates(UES)

library(ggplot2)

ggplot(d, aes(Time, PHQ, group=ID, col=ID)) +

  geom_point() +

  geom_line(aes(lty=ID)) +
scale_y_continuous(

"PHQ",

sec.axis = sec_axis(~ .*.21, name = "UES")
)

#aim four: do user preferences/expectancies explain changes in effectiveness?

#for preference model, had to eliminate P12 to use compare fits d/t missing data

Preference = lmer(PHQ~Preference + Time+ Preference:Time+(1|ID), data=d)

visualize(Preference, plot="residuals")

compare.fits(PHQ~Time|Preference, data=d, Preference, baseline, clusters=10)

model.comparison(baseline, Preference)

Expectancy = lmer(PHQ~DCES_Tot+ Time+ DCES_Tot:Time+(1|ID), data=d)

visualize(Expectancy, plot="residuals")

compare.fits(PHQ~Time|DCES_Tot, data=d, Expectancy, baseline, clusters=10)

model.comparison(baseline, Expectancy)
#aim five: do changes in coping skills explain changes in effectiveness?

#fixed vs random slopes for mindfulness

FFMQ_random = lmer(PHQ~FFMQ + Time+ FFMQ:Time+(FFMQ|ID), data=d)

Fixed = lmer(PHQ~FFMQ+ Time+ Time:FFMQ+ (1|ID), data=d)

model.comparison(Fixed, FFMQ_random)

compare.fits(PHQ~Time|FFMQ, data=d, Fixed, FFMQ_random, clusters=10)

visualize(FFMQ_random, plot = "model", formula=PHQ~Time+ID|FFMQ, sample=10)

#model comparison favors fixed slopes

Mindfulness = lmer(PHQ~Condition+ Time:Condition+ FFMQ+ Time+ Time:FFMQ+ FFMQ:Condition+(1|ID), data=d)

visualize(Mindfulness, plot="residuals")

compare.fits(PHQ~Time|FFMQ+Condition, data=d, Mindfulness, Condition, clusters=10)

model.comparison(Mindfulness, Condition)
visualize(Mindfulness, plot = "model", formula=PHQ~Time+ID|FFMQ+Condition, sample=10)

estimates(Mindfulness)

CBT = lmer(PHQ~Condition + Condition:Time+CBTSQ + Time+
CBTSQ:Time+Condition:CBTSQ+(1|ID), data=d)

visualize(CBT, plot="residuals")

compare.fits(PHQ~Time|CBTSQ+Condition, data=d, CBT, Condition, clusters=10)

model.comparison(Condition, CBT)

visualize(CBT, plot = "model", formula=PHQ~Time+ID|CBTSQ+Condition, sample=10)

estimates(CBT)

CSE = lmer(PHQ~Condition+Condition:Time+CSE + Time+
CSE:Time+CSE:Condition+(1|ID), data=d)

visualize(CSE, plot="residuals")

compare.fits(PHQ~Time|CSE+Condition, data=d, CSE, Condition, clusters=10)

model.comparison(Condition, CSE)
visualize(CSE, plot = "model", formula=PHQ~Time+ID|CSE+Condition, sample=10)

estimates(CSE)