Randomized Controlled Trial of Pacifica, a CBT and Mindfulness-based App for Stress, Depression, and Anxiety Management with Health Monitoring

A Dissertation
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

AnnaMarie Huong Vu

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Adviser: Patricia Frazier

August 2018

Acknowledgements

First, I would like to thank Professor Pat Frazier for her steadfast guidance over the course of my PhD while performing this research among numerous other milestones as life happened. The profound commitment she possesses towards her students will be impressed upon me for a lifetime.

Second, I would like to thank the members of my committee: Professor Moin Syed for his empathic instruction and expansive knowledge base over the course of my studies – all the while nurturing my own interests; Alexander Rothman for his mentorship and advocacy as to the societal implications of research; and Kathleen Harder for her willingness to extend her expertise in usability, service design, and cognitive psychology, all the while helping this square peg find an intellectual home.

I would also like to thank my friends and family: Wolf, whose support for me over the past several years is unparalleled; Sarah Fraser, my long-distance cohort-mate and dearest battle buddy; my friends and colleagues from the Community of Scholars program including Teréz Iacovino for her empowering tenderness and ability to make us laugh at the absurd; my little-big sister Lisa, who always made science look cool; and Jessica Engle, a kindred collaborative spirit, for her life-giving work of showing up as a friend.

Finally, I would be remiss if I did not thank the many mentors who were unwavering in their confidence in me to complete this work: Drs. Lisa Possis, Kathleen Feil, Kate Zona, Martina Rodgers, Bob Seybold, and Carey Gleason.

Dedication

This thesis is dedicated to my parents, Trung T. Vu and Van H. Nguyen, and my ancestors Tin T. Pham, Thu Vu, Tran T. Nguyen, Phien K. Nguyen who imbued my generation with the values of education, hard work, and service. I would like to make special note of those family members who were unable to obtain a formal education but survived for when we someday could.

Abstract

mHealth smartphone apps have inundated the market. The promise of mHealth apps is that they increase access to psychotherapeutic content while also expanding options beyond face-to-face care. However, information on the quality and efficacy of commercial mHealth apps is sparse. Pacifica, a Cognitive Behavioral Therapy and Mindfulness-based mHealth app, is one of the most popular and publicly endorsed apps on the market. An initial week long pilot study was performed to assess feasibility and acceptability of the Pacifica app and Pacifica Lite, an active control version of the app, in a group of college students (N = 41). Participants generally used the app as intended and rated the aesthetics and functionality positively. The subsequent efficacy study extended the intervention period, added a waitlist control group, and enrolled a greater number of participants. To the best of our knowledge, this study was the first randomized controlled trial designed to examine the feasibility and efficacy of a commercial mHealth app using a smartphone active control app as a comparison. Participants (N=420) were randomly assigned to one of three groups: 1) The Pacifica app that includes the psychotherapy-based components, 2) the active control app "Pacifica Lite" without psychotherapy-based components, and 3) waitlist control. The intervention period was 14 days over the final weeks of a college semester. Participants completed pretest and posttest measures of mental health symptoms and general well-being, state-based affect, perceived stress, and mindfulness prior to and following the intervention period. Regression models with baseline scores as moderators indicated there was an effect of the Pacifica intervention

compared to the Waitlist control on negative affect and global mental health. The effect of intervention group on negative affect, global mental health, and perceived stress were all moderated by baseline scores, whereby participants who had higher baseline distress levels reported greater responses to the Pacifica intervention compared to being on the Waitlist. A measure of trait mindfulness showed significant between-group differences, with significantly higher scores in the Pacifica group compared to the waitlist control. There were no effects related to group at posttest for total or subscale scores for symptoms of depression, anxiety, positive state-affect, and mindfulness practice. Overall, the Pacifica app appeared feasible and usable over a high demand final exam period, was significantly more effective than Waitlist on some measures, but it was not significantly more effective than the active control on any measures. Pacifica and the majority of mHealth apps available are not designed as alternatives to face-to-face care but college counseling centers and healthcare systems may choose to include apps like Pacifica as additional resources.

Table of Contents

Page

List of Tables

List of Figures

Chapter 1: Introduction

Young adulthood is a distinct developmental period in which many face multiple new life stressors, exacerbation of continuing stressors and mental health problems, and the transition from having a caregiver navigate healthcare services to having sole responsibility for addressing one's own mental healthcare needs. This chapter will review the literature regarding stress and mental health problems in young adults (18-35 years) and college student populations, barriers to accessing in-person mental healthcare, features of mHealth delivered interventions, and finally, the effectiveness of mHealth smartphone applications for the improvement of mental health symptoms and well-being.

Prevalence of Mental Health Diagnoses among College Students

Mental health disorders are prevalent among college students. According to a recent multinational survey conducted by the World Health Organization (WHO) which contained of a large sample of college students (n = 1,572), in high-income countries (including the United States) 25% of college students met diagnostic criteria for a mental health disorder within the last 12 months (Auerbach et al., 2016). Most (83%) college students who met diagnostic criteria from all countries sampled reported onset of mental health diagnoses prior to college. In the recent National Survey of College Counseling Centers (Gallagher, 2015), the majority of college counseling centers (86%) in the US reported an increase in the number of incoming students that were treated with psychiatric medication (Gallagher, 2015). Thus, the college student population is well-situated to benefit from an expansion in the breadth of mental health services offered, given the prevalence of mental health diagnoses warranting ongoing treatment.

Stress among College Students

College students face stressors unique to their educational experience. In a review by Robotham (2008), the author highlighted issues related to academics, finances, and the transition to college as primary sources of stress. The American College Health Association (ACHA) has conducted annual surveys of students enrolled in post-secondary institutions nationally. In a recent survey (ACHA, 2017), the majority of students indicated above average stress over the past 12 months, with 44% reporting more than average stress and 12% reporting tremendous stress. Relatedly, 87% reported 'feeling overwhelmed by all you had to do' at some point in the last 12 months. Additionally, 48% of students indicated that academics had been traumatic or very difficult to handle in the last 12 months. In terms of mental health, 22% had been diagnosed or treated for anxiety and 18% had been diagnosed or treated for depression in the previous 12 months.

The deleterious effects of stress generally have been documented for well over a century. Stressful life events can have bearing on factors as far upstream as our genetic code (Hunter, 2012). The impacts of stress include poor physical health (e.g., headache, fatigue, disturbed sleep, muscle tension), mental health (e.g., anxiety, depression, irritability, anger, active psychosis), cognitive abilities (e.g., problem solving, memory, attention) and health behaviors (e.g., alcohol consumption, tobacco use, changes in appetite) (Schneiderman, Ironson, & Siegel, 2005). The effects of stress on college students are numerous. In the review by Robotham (2008), higher stress correlated with greater alcohol and nicotine consumption, decreased self-esteem, and decreased academic

performance. In a recent literature review of qualitative or mixed-method studies that examined sources of college student stress (Hurst, Baranik, & Daniel, 2012), relationship stressors (be it with family, romantic, peer, and faculty) were the most commonly reported. Interestingly, the authors noted that issues around diversity (e.g., being a racial minority, disability status, first-generation college student, sexual orientation) was a unique stressor theme that arose in qualitative and mixed-method studies.

Identifying the Need for Services

The process of identifying the need for treatment and following through to seek and receive adequate professional services can be arduous for students. In a large web survey of university students (*N*= 2,785), 37% to 84% of students who met screening criteria for depression or anxiety did not use mental health care services (pharmacotherapy or psychotherapy) in the past year (Eisenberg, Golberstein, & Gollust, 2007). This percentage varied from 37% for those with depression and anxiety to 84% of those with non-major depressive disorder and no anxiety. Common barriers for students to seek mental health services included low awareness of services, a lack of a perceived need for help, and doubts about whether treatment would work (Hunt & Eisenberg, 2009). Students of certain backgrounds, such as those who reported having low socioeconomic status, international students, and Asian American students markedly underutilized mental health services (Hunt & Eisenberg, 2009). Self-stigmatizing attitudes about having mental illness were also associated with lower help-seeking behavior. However, for those who do seek services, the picture is not that much brighter. According to a study that investigated the prevalence of mental health disorders among

college students across the world, 16% of those with mental health disorders received only “minimally adequate treatment” (Auerbach et al., 2016).

Despite barriers to seeking and receiving adequate mental health services on campus, the demand for services on campus is actually on the rise (Watkins, Hunt, & Eisenberg, 2011). This is evidenced by the National Survey of Counseling Center Directors in which the percentage of enrolled students seeking counseling rose from 9% in 2008 to 11% in 2014 (Gallagher, 2008; Gallagher, 2015). Additionally, in this survey the ratio of counselors to students was 1 to 2,081 in 2014, a decrease from 1 to 1,906 in 2008 (Gallagher, 2008; Gallagher, 2015). In an earlier survey (Gallagher, 2008), almost all (96%) counseling directors reported that their centers could not meet the demands of students with severe psychological problems which included staff shortages during periods of peak utilization times (64%) and the need for providers to terminate cases prematurely (36%). Other service provision concerns reported by counseling center directors surveyed included an increase in crisis counseling (67%), issues finding referral sources for students who require long-term care (67%) and waitlist problems (22%).

Alternatives to Face-to-Face Delivered Interventions

In the last decade, alternatives have been developed to extend in-person services to underserved populations. One approach to solving the provider shortage problem has been to leverage technological capabilities (e.g., the internet) to facilitate, bolster, or even substitute for in-person care. These interventions have been developed to intensify treatment programs, provide introductory skills to identify and cope with mental health symptoms, be used adjunctive to in-person therapy, or used as stand-alone whereby

contact with a licensed professional is not made during the course of engagement with the intervention.

Muñoz (2010) termed interventions that are not self-guided, automated, or reusable as *consumable interventions* "that once used, cannot be used again" (p. 2). Examining mental health services in this light, 30 minutes of outpatient therapy services with a client would be categorized as a consumable intervention—those 30 minutes cannot be used again to treat other clients in the future. Muñoz does not make the moral argument that all interventions should be without consumable components. Rather, he proposed five categories or levels of interventions with varying degrees of non-reusable design and applied these to internet-delivered interventions. The five levels of internet interventions in order of least to greatest consumable components are as follows:

1) Automated self-help interventions
2) Guided internet interventions
3) Adjuncts to existing health care interventions
4) Interventions that expand health care beyond current offerings (e.g., language spoken, geographic proximity)
5) Proactive interventions (targeted toward persons who need greater level of care but are not currently engaged in services at established healthcare facilities)

For example, a web-based CBT program only offered to clients already seen by a provider in an outpatient clinic to reinforce skills learned in person would be categorized as a Level 3 internet intervention, an "Adjunct to Existing Health care" (p. 6). Rather than

assuming that all internet interventions must attend to the mental health care access and resource problem in the same way, researchers could benefit from using these categories to clarify the context in which interventions are designed to be implemented. Reliance on an exclusively consumable intervention delivery model may narrow the potential to adequately provide mental health care and well-being support services on college campuses.

Effectiveness of Self-Guided Interventions for College Students

Self-guided interventions - level one in Muñoz's (2010) framework - for college students have been shown to be effective for a variety of outcomes including anxiety, depression, and stress. Conley and colleagues (2016) performed a meta-analysis of 48 technology delivered interventions to postsecondary students for both universal interventions for those "without any presenting problems" (p. 659) (n = 22 studies) and indicated interventions for those with "mild to moderate subclinical problems" (p. 659) (n = 26 studies). The most commonly used intervention strategy was cognitive behavioral, which was used in 50% of the studies included in the meta-analysis. The authors found significant effects for both universal (g = 0.19) and indicated (g = 0.37) interventions at post-intervention compared to control groups, which included waitlist, informational, or active controls, depending on the study. Davies and colleagues (2014) performed a meta-analysis of 17 computer and web-based interventions for depression, anxiety, and psychological well-being in university students. When comparing interventions to waitlist controls, there was a significant reduction for those in the intervention condition for measures of anxiety (standardized mean difference (SMD) = -0.56) and depression (SMD

= -0.43). However, when comparing interventions to active controls, there were not significant differences in measures of anxiety (SMD = -0.18) or depression (SMD = -0.28). Taken together, these meta-analyses support the position that self-delivered technology based interventions can improve anxiety and depression in college students but only in comparison to waitlist controls.

Effectiveness of Smartphone Application Delivered Interventions for Stress and Anxiety

The development of even more accessible intervention delivery methods, such as smartphone applications, is another way to reduce stigma about seeking mental health treatment and increase treatment access (Clough & Casey, 2015). In 2013, there were estimated to be more than 10,000 mental health applications available to the public for download (Ben-Zeev et al., 2013). At that time, only five mobile phone apps met inclusion criteria for a systematic review of mental health apps research (Donker et al., 2013a) and, of these, only two were available for download to the public at the time. The systematic review of smartphone mental health programs by Donker and colleagues was the first published in this field and pointed to the lack of scientific rigor in the research designs used in mHealth app testing.

Of the 5,646 abstracts identified by Donker and colleagues (2013a), only eight papers describing five apps ($N = 221$) met the authors' inclusion criteria. Of these five apps, one was developed for stress management and assessed stress as a primary outcome measure, three apps were for the treatment of depression, and one app was aimed at reducing substance use for individuals with borderline personality disorder. The one app

focused on stress management was a stand-alone self-guided app based on stress inoculation training (SIT) called Mobile Stress Management. Two studies by Villani and colleagues used Mobile Stress Management in an RCT study design and targeted female oncology nurses who watched 8 videos over 4 weeks or were assigned to an attention control group (Villani et al., 2013; Villani et al., 2012). Villani and colleagues (2013) compared intervention and control groups (n = 15 in each group) using two COPE subscales (active and denial coping) and a measure of state-trait anxiety. Donker and colleagues converted Villani et al.'s (2013) results from t-tests to effect sizes and found medium to large effect sizes: d = 0.42 for active coping, d = 0.84 for state anxiety, and d = 1.08 for denial coping.

There were major limitations among the studies included in Donker et al.'s (2013a) meta-analysis. Multiple papers did not report statistics (e.g., Villani et al., 2012), sample sizes, or provide enough information to assess risk of bias. Of the five apps included in the review, only three reported on usability, helpfulness, and satisfaction supplied by its users. Of those studies that assessed usability, primary issues reported were connectivity and app freezing problems. Interestingly, when adherence rates were reported, these app-based interventions outperformed the internet-based interventions reviewed in a different study (Donker et al., 2013b). It is unclear what could explain this difference, although the authors postulated it could be due to the greater convenience and availability of mobile phone devices. Overall, the studies included were rated to be of low quality due to having small numbers of participants, poor reporting, limited follow-up efforts, or poor fidelity to treatment guidelines. The authors concluded that "Research is

particularly weak in the domains of sleep disturbances, anxiety disorders, and smoking cessation and needs further investigation" (p. 9).

Fortunately, some progress has been made in mental health smartphone app study designs and reporting since that first systematic review in 2013. Gee and colleagues (2015) conducted a systematic review of 15 studies that delivered ecological momentary interventions (EMI) for stress and anxiety using a range of mobile technologies (i.e., mobile devices, smartphones, and hand-held computers). EMI refers to "the use of interventions in patients' natural environments" (Heron & Smyth, 2010, p. 1). All study designs were either RCTs that compared EMI to wait-list control, no intervention, attention control, an alternative intervention or an unspecified control group. Smartphones are able to take advantage of delivery and monitoring modalities unique to their technology including a combination of audio, video, virtual reality, ambulatory feedback, and physiological monitoring. EMIs were categorized using the following typology provided by Carter and colleagues (2007). *Simple* EMIs were primarily informative, *interactive* EMIs allowed users to record their current state and display information on request, and *integrative* EMIs unobtrusively collected, computed, and interpreted patterns of an individual's momentary input, and incorporated this information to tailor a subsequent intervention to the individual. The EMI's used from most to least common were integrative (8), simple (7), and interactive EMI (3). These EMI's were either entirely self-administered, minimally supported by a therapist, or therapist-administered.

Of the five studies that targeted stress management (Lemaire et al., 2011; Villani et al., 2013; Lappalainen et al., 2013) and reductions in stress and anxiety generally (Proudfoot et al., 2013; Reid et al., 2011), four resulted in a reduction in stress relative to control groups with the exception of an EMI self-monitoring intervention (Reid et al., 2011). Interestingly, the majority of those studies that used self-administered EMIs found reduced stress and anxiety relative to control groups (Lemaire et al., 2011; Lappalainen et al., 2013; Proudfoot et al., 2013; Villani et al., 2013) whereas the majority of therapist-administered EMIs did not significantly reduce stress and anxiety compared to control groups. These studies included a range of EMI types including integrative (Proudfoot et al., 2013; Lappalainen et al., 2013), interactive (Lemaire et al., 2011; Reid et al., 2011) and simple (Villani et al., 2013).

In the same article, Gee and colleagues (2015) conducted a meta-analysis of the seven studies that were targeted to treat Generalized Anxiety Disorder (Riva et al., 2007; Mosso et al., 2009; Pallavicini et al., 2009; Newman et al., 2014; Reid et al., 2011; Grassi et al., 2009; Proudfoot et al., 2013). Of these studies, six had a control group (waitlist or no-intervention) and one had an active comparison treatment group. Overall, there was an effect size of $d = 0.32$ (95% CI 0.12-0.53) for the mobile delivered EMI interventions, a small to moderate effect. Sample sizes for these studies were small and the studies were most likely underpowered; improved reporting of study design would also allow future systematic reviews and meta-analyses to account for potential study bias.

Recent reviews of mHealth apps for mental health support these earlier findings of symptom improvement with use. Firth and colleagues (2017) reported the results of a

meta-analysis of nine studies examining smartphone interventions to reduce the symptoms of anxiety in adults. The authors found significant effects of smartphone interventions of $g = 0.45$ compared to waitlist controls and $g = 0.19$ when compared to active controls. Neary and Schueller (2018) reported that, although there is still a dearth of evidence supporting the use of apps for improving mental health across diagnoses, there is a steadily increasing number of studies showing significant effects.

College age students have begun to be evaluated as a population for mHealth apps targeting stress reduction. Lee and Jung (2018) evaluated a mindfulness-based app with undergraduate students. The study design was a waitlist controlled RCT ($n = 163$ at post-intervention) with the intervention used five times a week for four weeks and the evaluation of anxiety and stress at pre and post intervention. The authors reported reduced trait anxiety ($p = .01$, $\eta_p^2 = .05$) and improved general health ($p = .001$, $\eta_p^2 = .07$), but no significant change in perceived stress ($p = .06$, $\eta_p^2 = .02$) or state anxiety ($p = .17$, $\eta_p^2 = .01$) between the intervention and control groups. Another recent study by Harrer and colleagues (2018) evaluated an internet and app based mindfulness intervention for college students with elevated stress (PSS-4 ≥ 8). The study design was a waitlist controlled RCT ($n = 150$) with the intervention consisting of 8 modules (each taking 30-90 min) to be completed at a rate of 1-2 per week over a total of 5-7 weeks. A diary app was supplied to track mood ratings, daily interventions, stressful events, and health behaviors. A designated online eCoach who monitored completion of the intervention modules sent motivational messages to users. The intervention significantly decreased scores on the PSS relative to the waitlist control group at post intervention ($d = 0.69$) and

at three month follow-up (*d* = 0.57), state anxiety (STAI) at post intervention (*d* = 0.76) and the three month follow-up (*d* = 0.56), and depression (CES-D) at post intervention (*d* = 0.63) and the three month follow-up (*d* = 0.56). These studies provide a glimpse into the potential of mindfulness based training via web and mHealth apps for college students to reduce stress.

Evaluation of Quality of mHealth Applications

The promise of mHealth smartphone apps is that the technology increases access and convenience to psychotherapeutic material without many of the hurdles required to obtain consistent face-to-face care (e.g., health insurance coverage or out of pocket pay, referral processes, provider availability, and travel). It would not be a leap to imagine that given the low barrier to entry, consumers may be tempted to try a mHealth app before seeking in-person psychotherapy. However, information on the value and efficacy of commercial mHealth apps is sparse due to the often low quality and inadequate evaluation of these products. The cornucopia of mHealth apps currently available and the hurried pace in which these are released makes it difficult for providers, developers, and consumers to effectively vet these apps. Further, research on the effectiveness of smartphone apps "in the wild" that were developed apart from research institutions are underrepresented in peer-reviewed academic journals. Even for those who are knowledgeable about mHealth app developments and psychotherapy literature, finding one that is not only compatible with current treatment needs, targeted populations, easy to use, empirically based, and unobtrusive can feel a lot like panning for gold.

Chan and colleagues (2015) combined various guidelines into an evaluation

process that theoretically could be used by both patients and providers when selecting smartphone apps. These evaluation criteria consist of three dimensions: usefulness, usability, and integration or infrastructure. The dimension of *usefulness* pertains to the validity and accuracy of the app (i.e., does it work as promised); the reliability or consistency of its functionality (e.g., frequency of bugs and other fixes required); effectiveness for the targeted population, disease, or disability; and the expected number of sessions the user would have to engage in to receive benefit. The *usability* dimension pertains mainly to how the user interacts with the application's features and their subjective experience of it. According to the authors, usability includes satisfaction and reward, the amount of training required to use and understand the app (which should be minimal), accessibility for different disabilities (hearing, reading, impaired vision), cultural accessibility (including aspects such as ethnicity and language), as well as how user features may alter usability (implications for digital literacy such as user's age and SES status). Lastly, *integration or infrastructure* has to do with security, workflow, data sharing, safety, privacy policy features, and a theory as to how the app interacts with the user's behavior to facilitate change.

Efforts that seek to increase the efficiency of quality evaluation while still attending to the requisite breadth continue to be developed. The American Psychiatric Association (2018) has released guidelines for Psychiatrists to rate apps that highlight Privacy and Safety; Evidence, including peer-reviewed research; Usability; and Data Sharing. A review by Coulon and colleagues (2016) examined smartphone apps for stress management based on three domains: evidence-based stress management, transparency,

and acceptability of the user interface. A recent review by Neary and Schueller (2018) provided an overview of app rating platforms and guidelines, including that of the American Psychiatric Association (2018), and introduced a new mental health app rating platform and nonprofit initiative, PsyberGuide, for use by the public. PsyberGuide integrates the Mobile App Rating Scale (MARS), expert reviews, transparency ratings, and their own rating system that includes an overview of published studies examining the effectiveness of the app. Other research groups such as Bakker and colleagues (2016) aimed to develop a checklist and overview of specific features with the hopes of reaching app developers to ultimately improve the quality of mHealth apps. The authors reviewed currently available mHealth apps and developed a list of 16 recommended features to improve the quality of apps. These features included CBT-based content, aimed at addressing anxiety and low mood, designed for nonclinical populations, offers user reporting of thoughts and behaviors, encourages non-technology-based activities, and provides experimental evidence of efficacy. Bakker and colleagues evaluated 25 apps for mental health that are currently available to the public using their list of recommended features. The scores ranged from three to 13 (M = 6.96, SD = 2.33) which indicates that most apps have approximately half of the recommended features. The primary aim of these studies was to develop systems to effectively communicate app quality information across stakeholders, including the general public, mental health care providers, and app developers.

Summary of Reviewed Research

College students and college-aged adults report higher stress than the rest of the adult population (APA, 2018). Mental health diagnoses are prevalent in both populations, though college students experience additional stressors specific to their educational pursuits. Self-guided interventions delivered through the internet or book manuals can improve mental health and well-being in college students (Conley et al., 2016). As nearly all college-age adults have smartphones (Pew Research, 2018) and self-guided interventions can be used to reduce stress and anxiety, research into mobile app delivery of these interventions is needed. Initial studies reported encouraging small to medium effect sizes (Gee et al., 2015). However, issues regarding the current state of mHealth apps remain, including a lack of collaboration between mental health experts, consumers, and developers to inform the creation and on-going revision of mhealth apps; need for agreed upon processes for translating treatments to the mobile app scale; shortage of standard metrics in determining app quality; and the sparse number of randomized controlled trials conducted independently in proportion to commercial apps available on the market (Neary & Schueller, 2018).

Rationale for the Proposed Study

In addition to comprehensive treatment programs like CBT, brief interventions such as Mindfulness-based activities are useful in addressing stress. The shift toward adapting these traditionally in-person delivered interventions to mobile health platforms is timely. Approximately 95% of US adults own a mobile phone device, and 77% own a smartphone (Pew Research Center, 2018), with some evidence that device ownership is equivalent between clinical and non-clinical samples (Campbell, Caine, Connelly, Doub,

& Bragg, 2015; Carras et al., 2014). Smartphones in particular have become the most popular mobile technology with 94% of adults age 18-29 owning one (Pew Research Center, 2018).

The few mental health smartphone apps that have been developed to translate evidence-based practices using rigorous study designs do not seem to be easily identified and used by mental health providers and end users (East & Harvard, 2015). How do we ensure these innovations in mental health care delivery are not neglected? Chambers and Azrin (2013) suggested that researchers should engage in both dissemination and implementation science—partnerships with community stakeholders and organizations in which all sides reap the benefits of the mHealth technology developed. Although support has been garnered for public-private partnerships in mHealth from the research community, Tomlinson and colleagues (2014) warned that these partnerships "cannot happen at the expense of good science and good public health" – improvement in user mental health outcomes above all else should stay its central focus (p. 1).

These two studies attempted to integrate the strengths of rigorous academic research methods, standards of clinical practice, and the rapid iterative and implementation potential of the mobile software industry by partnering with Pacifica Labs, Inc., a company that developed a commercially available mHealth smartphone application. By collaborating with Pacifica Labs, Inc., I sought to bridge the knowledge-practice gap by examining the efficacy of a popular smartphone application that offers tools to help users manage stress, anxiety, and depression. Specifically, a college student sample was used because this is a population that demonstrates increased need but with

the safety net of access to mental health resources in place on campus and experience greater stress in comparison to other adult age groups (American Psychological Association, 2017). I conducted a small pilot RCT and a subsequent larger RCT with undergraduate college students over the course of the final exam period to evaluate the efficacy of the Pacifica app as an intervention for reducing stress and anxiety.

Introduction to Pacifica

The Pacifica application was developed by two programmers in broad consultation with two clinical psychologists. Its flagship smartphone mobile app provides "daily tools for stress and anxiety based on Cognitive Behavioral Therapy & Mindfulness" (Pacifica Labs, Inc., 2015). Applying the categories provided by Muñoz (2010), Pacifica could be considered a completely automated self-help mobile app intervention for the management of stress and anxiety. With over 1.9 million users there has been great interest in using this technology for anxiety, stress, and mood management.

Largely informed by the CBT model, the Pacifica App features aim to provide psychoeducation about symptoms and the CBT model; monitoring of mood, sensations, and health behaviors (e.g., time spent with friends and family, hours of sleep, amount of exercise, water intake); identification of maladaptive thoughts and behavioral patterns; exposure and reappraisal activities; social support and normalization; and self-care activities to maintain mood and manage stress levels borrowed from Behavioral Activation therapy. The app also teaches skills that target the physical tension that is a common symptom of anxiety and stress. Mindful Based Stress Reduction (MBSR)

activities such as guided imagery for relaxation and breathing exercises also are provided within the Pacifica app. (For further details, please see Appendix A).

The quality of Pacifica as a mHealth app has been evaluated in the literature. Neary and Schueller (2018) used PsyberGuide to evaluate Pacifica and reported a credibility score of 2.85/5. Pacifica's highest scores were the following items: clinicians provided input during the intervention development process, it is aimed at targeting mood and stress, and it is regularly updated. Pacifica's lowest score was the item assessing research support due to Pacifica not having a discernible research base supporting the app's effectiveness (PsyberGuide, 2018). The user experience was rated 4.7/5, with acceptable transparency. Expert reviews by licensed professionals suggested it may be useful for stress management. Bakker and colleagues (2016) scored Pacifica as having nine of the 16 recommended features, which is above the mean of 6.96 across the all apps.

Chapter 2: Mobile Application Pilot Study

Among college students, those who report higher stress are less likely to engage in self-care behaviors (e.g., regular hygiene, nutritious diet, adequate sleep) (Hudd et al. 2000) which are common targets of Behavioral Activation therapy for depression and CBT for anxiety disorders. A suite of interventions that combines treatment elements that have demonstrated efficacy for anxiety, stress, and depression, such as CBT, MBSR, and Behavioral Activation, could be potentially helpful for college students. This first study was designed as a feasibility and acceptability pilot of Pacifica, a low-intensity mobile app delivered intervention for stress management, anxiety, and well-being based on CBT and Mindfulness-based activities. The aims for the pilot study were two pronged: to ensure 1) that data collection and on-boarding processes in collaboration with Pacifica Labs, Inc. were sufficient for running a larger study in the future and 2) that the Pacifica and Pacifica Lite apps were perceived as acceptable by college students.

Methods

Intervention

The pilot study compared two conditions: 1) Pacifica, the mobile application developed by Pacifica Labs, Inc. and released to the public on January 27, 2015 and 2) Pacifica Lite, the active control condition co-designed for the purpose of this study and the student population. Version 2.0 of Pacifica used for participants in the pilot contained the following mobile app and CBT/Mindfulness features: the addition of a stress-rating feature presented after mood rating; push notifications throughout the day at random or a timed schedule per user preference that prompted the user to rate mood and stress; the

opportunity to set daily health goals (e.g., sleep duration, taking medication, time with relationships, caffeine intake) and mark these as completed; identify cognitive distortions via free text; and guided breathing and visualization exercises (for example app screens see Appendix B). Pacifica Lite contained in-app mood and stress ratings and health activity self-monitoring. The publicly released version 2.0 did not have the stress-rating feature and contained a feature for users to interact with each other in a group message board. Given constraints in our ability to monitor the message board for safety issues and to offer greater privacy protection to study participants, the message board feature was disabled.

Rather than compare the full version of the Pacifica app to Waitlist, which has been criticized in the psychotherapy research literature as “a low bar” for interpreting efficacy (Kazdin, 2014, p. 4), we co-designed the Pacifica Lite app as an active control condition. By providing users randomized to this condition with a smartphone app containing the same aesthetics, self-monitoring features, and psychoeducation as the full Pacifica app, we hoped to reduce disparities in possible outcome expectancies between these two conditions in the pilot and larger study. In this way, we were better able to examine whether it was the specific psychotherapy-based components available only in the full Pacifica App that were effective in reducing stress, anxiety, and depressive symptoms (Lambert & Ogles, 2013).

Participants and Procedures

Participants were students at a large Midwestern university who received extra credit for their participation through one of their Psychology courses. Participants were

recruited in one wave near the end of fall semester 2015. All participants 18 or over with regular access to a smartphone and no prior experience with the Pacifica app were eligible to participate in this study. Participants were provided with contact information for the researchers for the study, the institutional review board, as well as mental health services available on campus.

All participants who expressed interest in participating in the pilot study were sent a link to the consent form and the pre-test survey, 41 of whom consented and completed the pretest survey securely online. The participants were then randomized using the Random Integer Generator from random.org to the full version of the Pacifica app (n = 21) or the active control, Pacifica Lite app (n = 20). Of these, one participant assigned to the Pacifica Lite condition did not download the app, and one did not log any activities in the app; in addition, two full intervention participants did not log any activities in the app, and one did not download the app by the cut-off date. Thus, of the original 41 participants that completed the pretest survey, 88% (n = 36) successfully downloaded the app and used a component at least once over the duration of one week and 94% of those (n = 34; n = 17 Pacifica Lite; n = 17 Pacifica) completed the posttest survey. Pilot study attrition can be found in Figure 1.

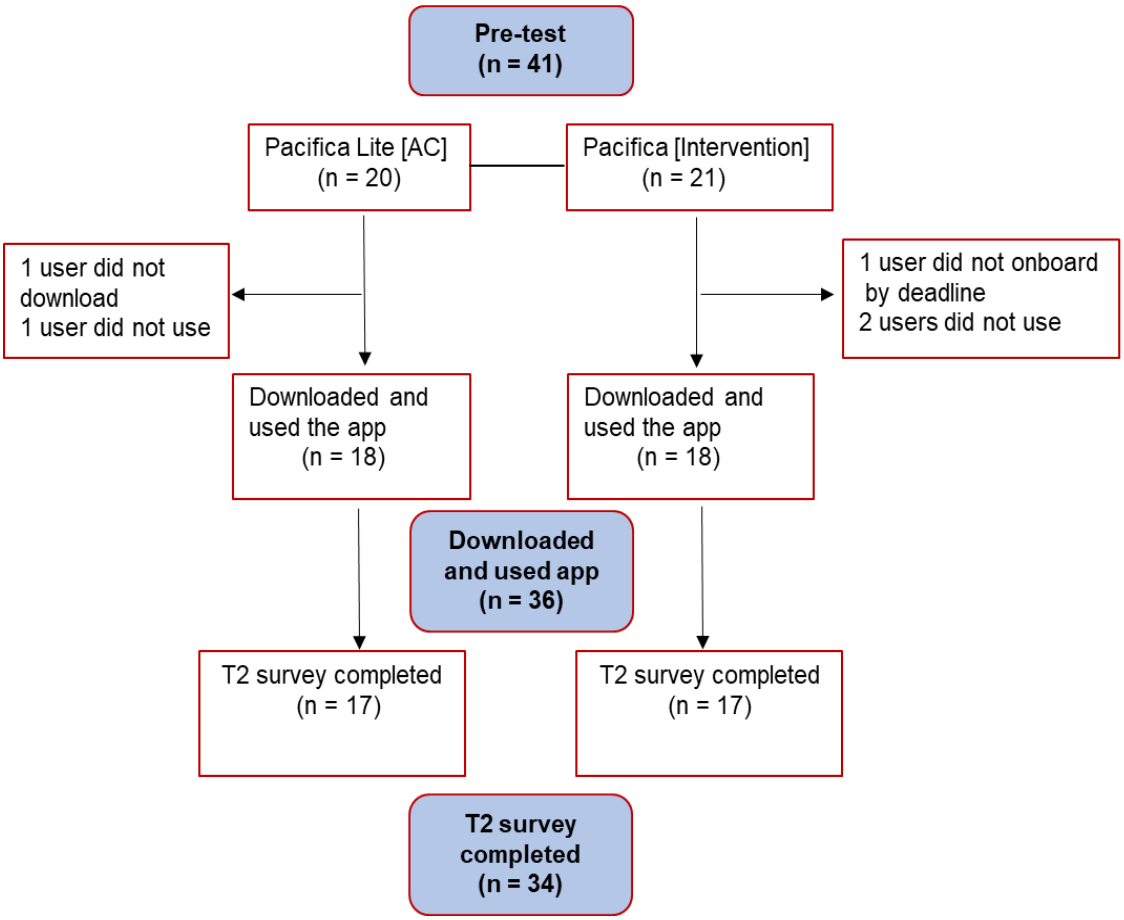

Figure 1. Pilot participant flowchart

Participants randomized to either condition of the app were given broad instructions to use the app daily for one week. Over the course of the week, participants could view a summary of their data entries through a data visualization feature within the app as well as through email via Pacifica Labs, Inc.

The following demographics and information regarding resources, access, and utilization of traditional face-to-face medical and mental health appointments were collected. Of the 36 participants who downloaded and used the app, the majority were between the ages of 18-21 years of age (72%), self-identified as female (72%), and European American and White (67%), domestic or not international students (81%), and

had a part-time job while enrolled in college (56%). The majority of participants were not Freshman with near equal distribution of Sophomores (25%), Juniors (25%), and Seniors (31%), with the remaining participants (19%) either Freshman, exchange students, or in 12 grade taking the class for college credit. Most participants reported that they had enough income to meet their basic needs (92%) and to meet their leisure needs (64%). Finally, the majority of responders reported that they had the opportunity to schedule medical and mental health appointments in a timely manner (67%) although most did not report that they had received mental health services in the preceding six months (61%). Using chi-square tests, there were no significant differences between the two conditions with respect to age, self-identified sex, race, year in school, international student status, employment, resources, access, and mental health care utilization (all $p > .30$).

Measures

The following measures were completed by participants prior to the usage of the app (Time 1) and at the end of one week of app usage (Time 2): depression, anxiety, and stress symptoms, global health, and state affect. Both qualitative and rating scale feedback of the assigned mobile app were completed at Time 2 only.

Depression, anxiety, and stress symptoms. Depression Anxiety and Stress Scales (DASS-21; Lovibond & Lovibond, 1995) is an abbreviated version of the original 42-item DASS. It is composed of three 7-item subscales. Each item is rated on a scale of 0 (*does not apply)* to 3 (*very much)* with regard to the past week. Alpha coefficients for scores on the DASS-21 subscales when used in college student samples in recent research have ranged from .81 to .93 (Frazier et al., 2009; Frazier et al., 2015; Hintz, Frazier, &

Meredith, 2014). Cronbach's alpha for the pilot sample across all subscales ranged from .79 to .92 across the two time points.

Global mental health. The Patient Reported Outcome Measurement Information System (PROMIS) Global Health Measure-Short Form 10 Subscale was developed by the National Institute of Health (Cella et al., 2010). It is a self-report measure of general physical and emotional health and well-being that has been widely used in primary care clinic screening and population health research. Because the Pacifica App targets mental wellness, the PROMIS Mental Health Subscale was the most pertinent aspect of the full scale for this study. The subscale consists of four items. Three items ask participants to rate their mood and cognition, quality of life, and social satisfaction on a scale of 5 (*excellent*) to 1 (*poor*) generally (e.g., "In general, how would you rate your mental health, including your mood and your ability to think?"). The fourth item "How often have you been bothered by emotional problems such as feeling anxious, depressed, or irritable in the past 7 days?" is rated on a scale of 5 (*never*) to 1 (*always*). In a U.S. adult majority White sample (N = 21,133), Cronbach's alpha was .92 (Hays et al., 2009). The Cronbach's alphas for the pilot sample for scores on the mental health subscale ranged from .77 to .78 across the two time points.

Positive and Negative State Affect. The Positive and Negative State-Short Form (PANAS SF-10) (Thompson, 2007) is a 10 item measure of two dimensions of state affect adapted from the original 20 item PANAS (Watson, Clark, & Tellegen, 1988). Participants indicated to what extent they felt a certain state "right now, that is, at the present moment" on a scale of 1 (*very slightly, or not at all)* to 5 (*extremely*). Feeling

states associated with positive affect included 'inspired' and 'excited' and feeling states associated with negative affect included 'hostile' and 'upset'. Internal consistency reliability in similar samples of college students have ranged from .61 to .83 for positive affect and .74 to .86 for negative affect (Moneta, Vulpe, & Rogaten, 2012; Ratelle, Simard, & Guay, 2012). The initial validation study (N = 444) by Thompson (2007) reported reliability of .76 for scores on the negative affect subscale and .75 for positive affect with test-retest reliabilities at 8 weeks of .84 for both subscales. For the pilot sample, the PANAS Negative subscale score reliability was .77 at T1 and .84 at T2 and the PANAS Positive subscale score reliability was .79 at T1 and .81 at T2.

Participant feedback. The Mobile App Rating Scale (MARS; Stoyanov et al., 2015) was one of few available in the research literature for participants to rate the quality of health mobile applications at the start of this study. The MARS prompts users to evaluate app quality on five dimensions: Engagement (five items), Functionality (four items), Aesthetics (three items), Information (seven items), and Subjective Quality (four items) and an optional fifth dimension (six items) that focuses on the likelihood that the mobile app increases change for the target population (e.g., awareness, knowledge, attitudes, intention to change, help-seeking for stress and well-being, and behavior change for college students). Although the anchors for each item vary considerably, all items are rated on a 5-point scale and, in general, 1 indicates inadequacy and 5 indicates excellence. The Cronbach’s alphas for the pilot sample ranged from .68 to .88 across the two time points for both the total scores and subscale scores.

Participants were also prompted to submit additional comments about the app. This information was collated to inform the roll out of the larger-scale study.

Results and Discussion

Analysis of the data collected during the pilot study was completed with the knowledge that the study was underpowered to detect significant differences between the two groups. An a priori power analysis using G*Power 3.17 (Faul, Erdfelder, Buchner, & Lang, 2009) indicated that 128 participants (N = 64 per condition) would be necessary to detect a medium effect in an ANCOVA with two groups and one covariate at a significance level of .05 and power of .80. Thus, the pilot study analyses were completed for exploratory purposes only.

The histograms and boxplots of all outcome measures were examined at T1 and T2 and Grubbs' test was performed to identify outliers. There were no extreme scores (Z > 4.0) using the Grubbs' test; therefore, all data were kept in the sample. Skewness and kurtosis were calculated for each scale and subscale at T1 and T2. The DASS Depression subscale at T1 and T2, the DASS Anxiety subscale at T1 and T2, the PANAS Positive subscale at T1 and T2, and the PANAS Negative subscale at T1 and T2 were found to be highly skewed (absolute value of the z-test of skew > 1.96) (Kim, 2013). Additionally, the DASS Depression subscale at T1 and T2 had heavy tails (absolute value of the z-test of kurtosis > 1.96) (Kim, 2013). As this was a pilot study, no transformations of the data to account for this non-normality were performed.

All analyses were completed using listwise deletion of individuals who did not complete survey two. Within-group Cohen's *d* effect sizes were calculated by subtracting

the mean of each group at T2 from the mean of that group at T1 and dividing by the standard deviation (SD) of that group at T1 corrected for the correlation between T1 and T2 (Morris & Deshon, 2002). The between-group Cohen's d effect sizes were calculated by subtracting the within-group effect sizes for the two groups.

DASS

To assess whether symptoms of depression, anxiety, and stress were different between groups at T2, four analyses of covariance (ANCOVA) were performed with the randomly assigned fixed effect of group condition as the independent variable, T2 DASS total scores or subscale scores as the dependent variables, and the T1 DASS-21 total or subscale scores as covariates. There were no significant between-group differences on the DASS-21 total or subscales (Table 1). Between group effect sizes ranged from small to medium with *d*s ranging from 0.14 to 0.33 (Table 2) favoring the Pacifica condition. For the DASS Total and Depression and Stress subscales, the Pacifica group showed a small to medium within group effect size decrease from T1 to T2 (d = 0.25, d = 0.29, d = 0.30, respectively) (Table 2). The Pacifica Lite group showed no consistent within group changes across the DASS total or subscales.

PANAS

ANCOVAs were completed for the PANAS Negative and Positive subscales with the randomly assigned fixed effect of group as the independent variable, T2 PANAS Positive or Negative subscale scores as the dependent variables, and the T1 PANAS Positive or Negative subscale score as covariates. The PANAS Negative subscale trended toward significance, $F(1,31) = 4.11$, $p = .051$, partial $\eta^2 = 0.12$. Participants in the

Pacifica condition reported lower negative state affect at time 2 than the participants in the Pacifica Lite condition with a large effect size of $d = 0.94$ (see Tables 1 and 2). However, for both conditions, negative state affect increased from T1 to T2; with a large effect size, $d = 1.07$, for Pacifica Lite. The PANAS Positive subscale was not significantly different between groups, $F(1,31) = 0.23$, $p = .64$. The Pacifica lite group displayed a medium to large effect size increase in PANAS Positive. The between-group effect size also was medium to large, favoring the Pacifica lite group (Table 2).

PROMIS

An ANCOVA was performed for the PROMIS Global Mental Health scale with the randomly assigned fixed effect of group condition as the independent variable, T2 Global Mental Health score as the dependent variable, and the T1 Global Mental Health score as covariate. There was not a significant main effect of group at T2 for PROMIS Global Mental Health, $F(1,31) = 2.60$, $p = .12$, indicating no between-group differences for this scale at T2. Both groups had an increase in PROMIS scores between T1 and T2 (Table 2) and the Pacifica group had a small to medium within-group effect size of $d = 0.36$ and a small to medium between-group effect favoring the Pacifica app.

Table 1. ANCOVA Analyses Assessing Between-Group Differences at Post-Intervention

	Pacifica Lite		Pacifica				
	T1: Mean (SD)	T2: Mean (SD)	T1: Mean (SD)	T2: Mean (SD)	*F*	*p*	partial η^2
DASS21 Total	0.78 (0.57)	0.78 (0.52)	0.77 (0.48)	0.65 (0.50)	1.37	.25	.04
DASS Depression	0.72 (0.56)	0.64 (0.56)	0.75 (0.59)	0.58 (0.68)	0.28	.60	.01
DASS Anxiety	0.62 (0.67)	0.67 (0.60)	0.46 (0.51)	0.43 (0.38)	1.52	.23	.05
DASS Stress	1.01 (0.70)	1.03 (0.67)	1.11 (0.66)	0.92 (0.63)	2.14	.15	.07
PANAS Negative	1.4 (0.46)	1.89 (0.73)	1.58 (0.51)	1.65 (0.75)	4.11	.05	.12
PANAS Positive	1.8 (0.62)	2.18 (0.82)	2.43 (0.82)	2.42 (0.86)	0.23	.64	.01
Global Mental Health	11.88 (3.14)	12.12 (3.35)	12.41 (2.98)	13.47 (2.70)	2.60	.12	.08

Summary. The initial pilot results were promising for the Pacifica application. The within group effect sizes for DASS total, and the depression and stress subscales, all indicated a small to moderate decrease in symptoms (d = -0.36 to d = -0.42*)*. Additionally, a moderate increase in the PROMIS Global Mental Health subscale (d = 0.55) was detected (Table 2). Taken together, these findings indicate an improvement in general mental health measures. It should be noted that there was a very small increase in

negative affect and no change in positive affect detected, indicating that not all measures demonstrated improvement from T1 to T2 (Table 2).

The initial pilot results were not promising for the Pacifica Lite application, which is expected given that this condition was designed to serve as an active control for the RCT. Within group effect sizes for symptoms of depression did suggest very small decreases in the Pacifica Lite group (d = -0.17), while symptoms of anxiety (d = .13) and stress (d = 0.06) remained relatively unchanged from T1 to T2 (Table 2). However, within group effect sizes ranged from medium to large in state ratings of both positive affect (d = 0.59) and negative affect (d = 1.12), with increases in both. This finding is not unexpected given that participants were asked to rate their mood and assign descriptive words to their mood states, possibly increasing awareness of both their positive and negative state affect overall. General ratings of mental health wellness remained relatively unchanged (d = 0.17) (Table 2).

Table 2. Within-Group and Between-Group Effect Sizes (Cohen's d)

Measure	Group	Within-group *d*	Between-group *d*
DASS21 Total	Pacifica Lite	0.00	-0.36
	Pacifica	-0.36	
DASS Depression	Pacifica Lite	-0.17	-0.24
	Pacifica	-0.41	
DASS Anxiety	Pacifica Lite	0.13	-0.20
	Pacifica	-0.07	
DASS Stress	Pacifica Lite	0.06	-0.48
	Pacifica	-0.42	
PANAS Negative	Pacifica Lite	1.12	-0.94
	Pacifica	0.18	
PANAS Positive	Pacifica Lite	0.59	-0.60
	Pacifica	-0.01	
Global Mental Health	Pacifica Lite	0.17	0.38
	Pacifica	0.55	

Note: Positive *d* indicates an increase in score between Time 1 and Time 2. For the first five measures a decrease in *d* indicates a reduction in symptoms and negative affective state. For the PROMIS-Global Mental Health subscale and PANAS-SF Positive subscale an increase in *d* indicates improved mental health and more positive affective state. Between group effect sizes indicate the differences in changes in symptom and affective ratings between the Pacifica and Pacifica Lite app participants.

Table 3. MARS Application Quality Scores

	Pacifica	Pacifica Lite			
Measure	Mean (SD)	Mean (SD)	*t*	*p*	*d*
Engagement	3.65 (0.61)	3.19 (0.65)	-2.12	.04	0.73
Functionality	4.44 (0.61)	4.47(0.46)	0.13	.90	-0.05
Aesthetics	4.27 (0.52)	4.12(0.72)	-0.73	.47	0.26
Subjective Quality	3.16 (0.76)	2.94(0.53)	-0.98	.33	0.35
App Specific	3.70 (0.52)	3.50 (0.91)	-0.80	.43	0.30

Feedback from participants on app quality as measured by the MARS was mostly positive. In particular, there were two subscales that demonstrated mean scores above the scale's midpoint (3) in the Pacifica group. The subscale of Functionality had a mean score of 4.44 (*SD* = 0.61) and the subscale of Aesthetics had a mean score of 4.27 (*SD* = .52) (Table 3). Within these subscales, items measuring performance, ease of use, navigation, layout, visual appeal, and graphics all had mean scores greater than 4. The Subjective Quality subscale was scored the lowest, though still above the scale's midpoint (*M* = 3.16, *SD* =0.76). In particular, the item 'Would you pay for this app?' had a high response rate of 'No' which strongly influenced this scale's sub score. There was also a significant difference between the Pacifica and Pacifica Lite group for Engagement with a moderate to large effect size (*d* = 0.73), with the Pacifica Full Intervention group endorsing higher engagement.

A total of seven out of 34 participants provided qualitative comments regarding their experience with the Pacifica or Pacifica Lite apps. Not surprisingly, participants randomly assigned to the Pacifica Lite app desired more information about stress, coping

skills, and mental health resources. Participants randomly assigned to the Pacifica app tended to have a positive experience and perceived some benefit over the one-week time period. These benefits included the timing of the intervention “I felt that during finals week [it] may have been a great time to use this app, because of increased level of stress generally felt by students.” Another participant randomized to the Pacifica app noted a distinct like for the soundscapes offered in the Relax feature and being able to track health activities “and I can see how my stress level and mood relate to those (activities).”

Chapter 3: Mobile Application Randomized Controlled Trial

The RCT study underwent several changes in comparison to the pilot. First, the sample size was increased. An a priori power analysis was completed using G*Power 3.1.9.2 (Faul, Erdfelder, Buchner, & Lang, 2009) which indicated that 380 participants (N = 127 per condition) would be necessary to detect a small to medium effect (f = 0.16) in an ANCOVA with three groups and one covariate at a significance level of .05 and power of .80. The effect size of 0.16 was derived from the mean effect size (d = 0.32) found by Gee and colleagues (2015) in their meta-analysis of mobile apps for stress and anxiety. Second, the RCT included a greater length of time of app usage (two weeks rather than one week). The third change was the addition of a waitlist control group and thus a comparison across three conditions: intervention, active control, and waitlist control. By adding the waitlist group to the study design, changes in outcome variables could be better accounted for by the Pacifica app specifically. This also allowed for an assessment of the possibility that the components of the Pacifica Lite app may have therapeutic benefits. Fourth, measures that detect the effects of the mindfulness-based components of the Pacifica app were added. Finally, baseline scores were assessed as potential moderators of the efficacy of the Pacifica app as seen in previous literature for depression and anxiety in which higher baseline levels of distress were associated with greater improvement (Bower et al. 2013; Hedman, Andersson, Lekander & Ljótsson, 2015).

The hypotheses for the RCT were as follows: 1) Participants randomized to the full Pacifica app condition will report greater reductions in the primary outcome measure of DASS-21 than those randomized to the Pacifica Lite or Waitlist conditions at post-intervention (T2). 2) Participants randomized to Pacifica Lite will report comparable or slightly greater reductions in the primary outcome measure of DASS-21 than those in the Waitlist condition. 3) There will be no difference between groups for the PROMIS Global Mental Health subscale given that three of four items assess well-being generally. 4) Participants in the Pacifica app condition (who have access to mindfulness-based activities) will report a significant increase in mindfulness traits and practice in comparison to participants in the waitlist and active control conditions. 5) Given that the app features do not target state-based affect changes, there will be minimal, if any, between group differences in state-based affect for the Pacifica App and Pacifica Lite app. 6) Participants in the waitlist condition may report lower positive affective states than those provided apps. 7) Finally, it is hypothesized that there will be no differences in perceived stress, given that there is minimal content in the app conditions that directly addresses perceived stress. However, it is still important to include a perceived stress measure in this study given its prominence in the stress and anxiety intervention literature. Within these hypotheses, those that hypothesize an effect of intervention group are also predicted to have the group effect moderated by T1 score as we expect participants with lower DASS and FFMQ, and higher PANAS Negative scores at T1 to receive greater therapeutic benefits from the intervention.

Methods

Intervention

This study was a randomized controlled trial of a low-intensity mobile-app delivered intervention for stress management, anxiety, and well-being. The RCT compared three conditions: 1) “Pacifica”, the full intervention application app (containing in-app monitoring in addition to all CBT and Mindfulness components), 2) “Pacifica Lite”, the active control application (containing in-app mood and stress ratings, in addition to health activity monitoring), and 3) Waitlist control.

Participants and Procedures

Participants were students at a large Midwestern university who received extra credit for their participation through one of their Psychology courses. The participants were recruited near the end of the Spring 2016 and Fall 2016 terms during the last two weeks of the semester. To obtain greater power to detect between-group differences, data from the two cohorts were combined. All participants 18 or older with regular access to a smartphone and no prior experience with the Pacifica app were eligible to participate in this study. Participants were provided with contact information for the researchers, the institutional review board, as well as mental health services available on campus. All participants who expressed interest in participating in the study were sent a link to the consent form and the pretest survey. In total, 420 participants consented and completed the pretest survey securely online. These participants were then randomized using the Random Integer Generator from random.org to the full version of the Pacifica app (n = 140), the Pacifica Lite version of the app (n = 138), or the waitlist control (n = 142)

(Figure 2). Participants were meant to be blinded to their assigned group, though despite study efforts, knowledge of assignment to the waitlist group in particular is probable given participants had classmates assigned to other conditions. The participants in the active control and intervention groups both received the app and were not informed which group they were in and likely remained blinded through the study as they were not previously exposed to the application or the elements included in the app. However, study investigators could not ensure that participants did not discuss or share the application with others, including participants enrolled in different conditions. Although approximately 16% of the participants who completed the pretest survey and were randomized into either app condition did not download the app, these participants were included in intent to treat analyses.

Data collected via internet surveys are vulnerable to inaccurate survey responses (Meade & Craig, 2012). To account for careless responding (i.e., responses that are poor indicators of participants' true levels of the constructs measured), two data cleaning methods recommended by Meade and Craig) were used to prepare the survey data for analyses. Four instructed response items were included across the Time 1 and Time 2 surveys. Participants who incorrectly responded to two or more instructed response items across the two surveys were removed from the study. In total, 23 participants (6% of the sample) were removed using this criterion. The second method was to ask respondents who completed the Time 2 survey if the researchers should keep their data given that the integrity of their responses are important to the study. An additional 35 participants (8%) were removed.

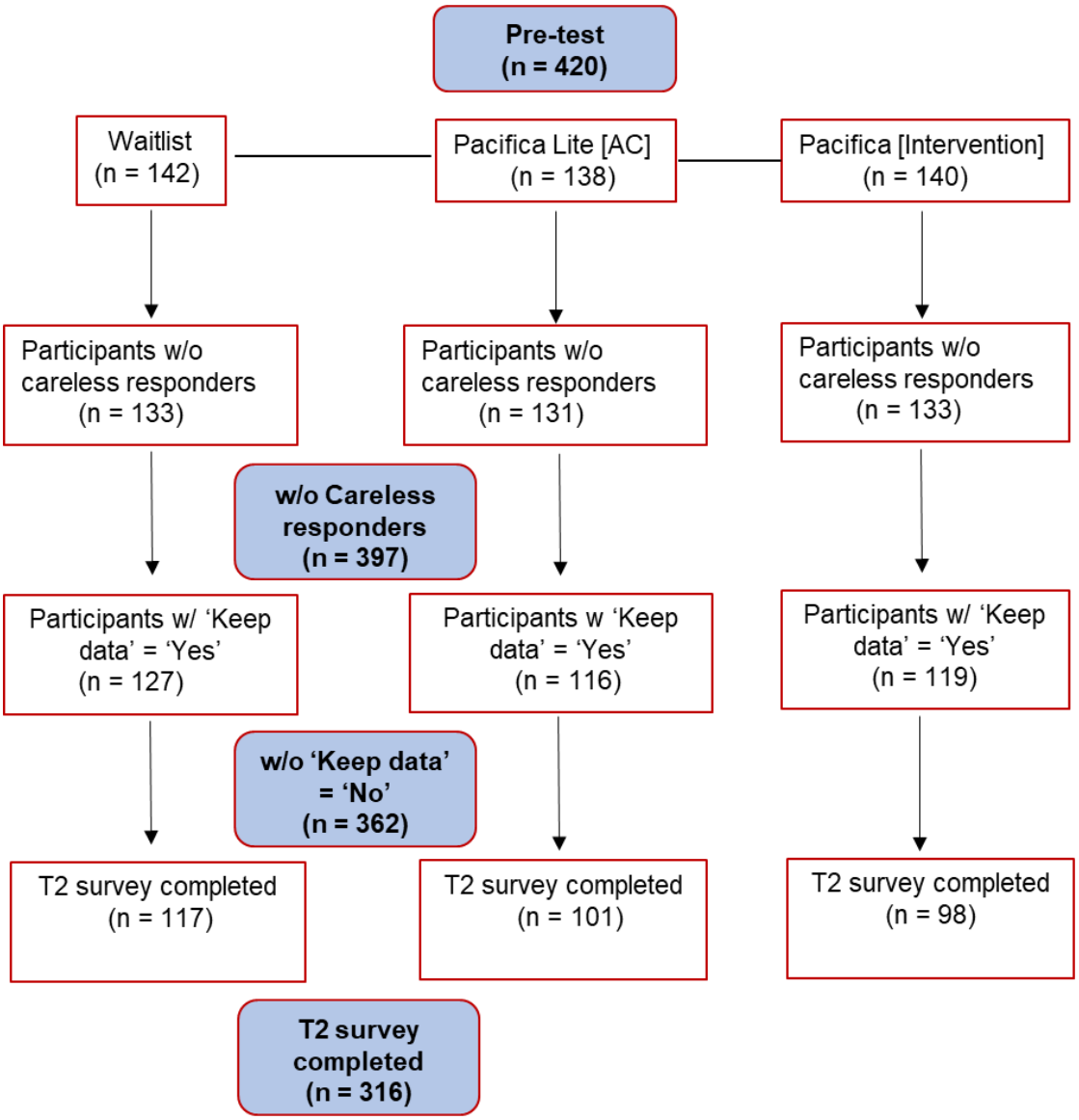

Figure 2. RCT Attrition Flowchart

Participants randomized to either condition of the app were instructed to use the app daily for two weeks; however, all features were available ad libitum. They could access a summary of their in-app ratings through a data visualization feature within the app as well as through receipt of an email via Pacifica Labs, Inc.

The following demographics and information regarding resources, access, and utilization of traditional face-to-face medical and mental health appointments were collected (Appendix B, Table 1). Of the 362 participants included in the analyses, the majority were between the ages of 18-21 years of age (74%), and self-identified as female (72%). Most of the participants identified as domestic (not international) students (87%). Most identified as European American or White (65%) with the next most frequent identification being Asian or Asian American (22%). Most participants (93%) reported that they had enough income to meet their basic needs and the majority (70%) reported they had enough income to meet their leisure needs (e.g., visiting friends, eating at restaurants, and travel). Slightly more than half of the participants held a part-time job while attending college (57%), 36% were not working, and 7% were working full time. The largest percentage of participants were college Juniors (28%), followed by Sophomores (27%), Seniors (25%), and Freshman (18%).

The majority of participants reported that they had the opportunity to schedule medical and mental health appointments in a timely manner (66%) and most (67%) did not seek mental health services in the six months prior to starting the study. About half of the participants did not currently or in the past practice mindfulness (56%) although some (14%) were currently practicing mindfulness, while the remaining participants had in the past (30%). Most participants had not previously received CBT (79%); however, a few (4%) were currently receiving CBT, with the remaining participants either receiving CBT in the past (10%) or were unsure if they had received CBT (7%).

Using chi-square tests, there were no significant differences between the three conditions with respect to age, self-identified sex, race, year in school, international student status, employment, resources, access to healthcare, or service utilization in the previous six months (all *ps* > .10).

Measures

The primary outcome measure was the DASS-21, and in particular the subscales for stress and anxiety. Secondary outcome measures included the Perceived Stress Scale (PSS), the Positive and Negative Affect Schedule-Short Form (PANAS-SF), PROMIS-Global Mental Health Subscale, the Five Facet Mindfulness Questionnaire (FFMQ), the Applied Mindfulness Process Scale (AMPS), and the Mobile App Rating Scale (MARS). Of these, the PROMIS Global Health was initially developed for online delivery (Hays et al., 2009) and the DASS21 has been evaluated for online delivery and found to have similar scores to the established norms via paper delivery (Zlomke, 2009). The PSS (Herrero & Meneses, 2006) and PANAS (Howell et al., 2010) have been validated in studies directly comparing participants randomly assigned to either pencil and paper or online delivery and there was no significant difference in scores between the two groups. The following measures were completed online by participants prior to the use of the app (T1) and at the end of two weeks of app usage (T2): depression, anxiety, and stress symptoms, perceived stress, global health, and state affect. Participants also completed a measure of dispositional mindfulness at both T1 and T2 and a measure of mindfulness applied to everyday life at T2 only. Additionally, participants evaluated the app using

both a quantitative and qualitative assessment at T2. For additional information for previously described measures, see "Methods" section for the pilot study in Chapter 2.

DASS-21. Cronbach's alpha for scores on the total scale and all subscales ranged from .82 to .93 across the two time points.

PROMIS. The Cronbach's alphas for scores on the mental health subscale ranged from .81 to .84 across the two time points.

PANAS SF-10. The Cronbach's alphas ranged from .73 to .83 across the two time points for scores on the PANAS total and PANAS positive and negative subscales.

Perceived Stress Scale. The Perceived Stress Scale (PSS) is a 10-item scale that measures the respondent's perception of stress over the previous month (Cohen, Kamarck, & Mermelstein, 1983). Item are rated on a scale from 0 (*Never*) to 4 (*Very Often*). Internal consistencies of greater than .80 have been reported with college student samples (Caldwell et al 2010; Cavanagh et al 2013). The Cronbach's alphas for PSS scores in this sample were .87 for time one and time two.

Five Facet Mindfulness Questionnaire. The Five Facet Mindfulness Questionnaire (FFMQ) is a 39-item scale that measures five facets (observing, describing, acting with awareness, non-judgment, and non-reactivity) related to mindfulness as a disposition (Baer, Smith, Hopkins, Krietemeyer, & Toney, 2006). Items are rated on a scale from 1 (*Never or very rarely true*) to 5 (*very often or always true*) and are instructed to be reported based on the rater's "own opinion of what is generally true for you." Internal consistency for the subscale scores ranged from .75 to .91 for undergraduate psychology students (Baer et al., 2006). In further studies with college

students, the internal consistencies for total scale scores and subscale scores ranged from .79 to .93 (Caldwell et al. 2010). The Cronbach's alpha of the FFMQ total scores was .89 at T1 and .90 at T2

Applied Mindfulness Process Scale. The Applied Mindfulness Process Scale (AMPS) is a 15-item scale that assesses three domains of applied mindfulness using 5-point response scale from 0 (*Never*) to 4 (*Almost Always*) (Li, Black, & Garland, 2015). The three domains include decentering, positive emotion regulation, and negative emotion regulation with each subscale consisting of 5-items. It is a self-report measure assessing the application of mindfulness practice during daily life to cope with adversity or stressors over the previous seven days. An internal consistency of .94 was reported in a sample that included university students, faculty, and staff participating in an online mindfulness training (Li et al., 2015). The Cronbach's alpha for the full scale scores was .93 and the subscale score reliabilities ranged from .79 to .90.

MARS. The Cronbach's alphas in this sample ranged from .69 to .88 for the total scores and subscale scores.

Results

Analyses

The histograms and boxplots of all outcome measures were examined at T1 and T2 and Grubbs' test was performed to identify outliers. One participant was identified as having an extreme score ($Z > 4.0$) on the DASS Anxiety subscale at T1. However, the data appeared valid so they were kept in the sample. Skewness and kurtosis were calculated for the scores of each scale and subscale at T1 and T2. All scales and subscales

scores were within the recommended range (Kim, 2013) for skew (< 2) and kurtosis (< 7).

All analyses were completed using listwise deletion of individuals who did not complete survey two. Listwise deletion was used because 13% of the participants did not complete the survey at T2, which is within the acceptable range to perform listwise deletion (5% to 20%) rather than multiple imputation (Schlomer, Bauman & Card, 2010). Little's missing completely at random test was performed on all T1 and T2 measures and was not significant, $X^2(97) = 91.4$, $p = .64$. Therefore, the null hypothesis, that the data were missing completely at random, was not rejected. Within-group Cohen's *d* effect sizes were calculated by subtracting the mean of each group at time 2 from the mean of that group at time 1 and dividing by the standard deviation (SD) of that group at time 1 corrected for the correlation between time 1 and time 2 (Morris & Deshon, 2002). The between-group Cohen's d effect sizes were calculated by subtracting the within-group effect sizes for the two groups.

To assess whether scores differed between groups at T2, participant scores at T2 were regressed on the intervention condition (dummy coded to compare Pacifica to Waitlist and Pacifica Lite to Waitlist), the T1 total score or subscale score of the measure, and the interaction of the group variables and T1 scores. The Johnson-Neyman method was used to probe the interaction of the group variables and T1 scores to calculate at which value of the T1 score the intervention significantly differed from the waitlist using the Omnibus Groups Regions of Significance (OGRS) package for SPSS (Hayes & Montoya, 2017). The Cohen's *d* effect size at the transition to significance, and on either

side of this transition, was converted from the model r^2 change at the T1 value by the standard conversion equation: $\frac{2*\sqrt{r}}{1-r^2}$.

Additionally, the percentage of participants in each group that improved or worsened was calculated using the reliable change index (RCI) on each measure. The RCI was calculated as $\sqrt{2 * SEM^2}$, where the SEM is the standard error of the measure and calculated as $SD_{T1} * \sqrt{1-\alpha}$, where SD_{T1} is the standard deviation of the measure at T1 and α is the reliability of the measure at T1 as calculated by Cronbach's alpha (Jacobson & Traux, 1991). The RCI is a measure to determine a boundary value beyond which the change scores are unlikely to occur by chance ($p < .05$) and therefore reflect actual change rather than measurement error.

DASS

There was no significant effect of group for DASS total or any subscales (Table 5). The interaction of T1 scores with intervention group was also not significant for DASS total or any of the subscales (Table 5). To further examine within-group changes from T1 to T2 (Table 4; Figure 3 and 4) and between group differences at T2, Cohen's *d* were calculated. The between-group effect sizes were largest for the DASS-21 total, DASS anxiety, and DASS stress subscales. The DASS-21 total score (d = -0.37 between the Pacifica and Waitlist groups) indicated a small to medium difference. The DASS stress subscale (d = -0.31 between the Pacifica and Waitlist groups) and DASS anxiety subscale (d = -0.36) both indicated a small to medium difference in decrease in stress between the full intervention and waitlist groups (Table 8). The DASS anxiety subscale also indicated a small to medium difference between the Pacifica group and the Pacifica

Lite group, indicating an effect of the intervention, not solely the use of the app for mood tracking, although this difference was not significant. The effect sizes for depression were all small.

Using the RCI, the largest percentage of participants improved (30%) and least worsened (14%) in the Pacifica group whereas the largest percentage of participants worsened (30%) and least improved (21%) in the Waitlist group for the DASS Total (Tables 9 and 10). A similar pattern occurred for the DASS Anxiety and Depression subscales where the largest percentage of participants improved (25% and 26%, respectively) and the smallest percentage worsened (9% and 8%, respectively) in the Pacifica group. The pattern of participants who significantly improved or worsened in the Pacifica Lite and Waitlist groups was less consistent. The Waitlist group had the least improvement (15% for DASS Anxiety and 20% DASS Depression). However, 24% of the participants in the Pacifica Lite group worsened in the DASS Anxiety whereas only 21% of the Waitlist participants worsened. The Pacifica and Pacifica Lite groups had similar percentages of participants who improved and worsened on DASS Stress (22% improved and 20% worsened for Pacifica; 25% Improved and 27% worsened for Pacifica Lite). The Waitlist group had a higher percentage of participants who worsened (31%) than improved (17%).

PANAS Positive and Negative Affect

There was a significant interaction effect between Pacifica and Waitlist groups for T1 score for PANAS negative (t = -3.86, p = .0001); however, there was not a significant interaction between T1 scores and the comparison between Pacifica Lite and Waitlist

(Table 5). Participants in the Pacifica group with T1 scores above 1.75 decreased significantly more in negative affect at T2 than those in the Waitlist group. Below this baseline value of negative affect, those in the Pacifica app group did not differ from the Waitlist group (*d*s= 0.09 to 0.14 below T1 = 1.75). The effect sizes were small to moderate within the significant region of T1 > 1.75 (d = 0.23 at T1 = 1.75 to d = 0.46 at T1= 3.72) which included 38% of the participants (see Table 7).

The overall within-group effect size for PANAS Negative for the Waitlist and Pacifica Lite groups was small to medium and indicated that PANAS Negative scores increased for both of these groups from T1 to T2, whereas the Pacifica group did not change. This resulted in small to medium between group effect sizes for the Pacifica group in relation the to Waitlist and Pacifica Lite groups (Table 8; Figures 5 and 6).

There were no main or interaction effects between groups for PANAS positive. All within and between group effect sizes for PANAS Positive were less than small (d < 0.15).

The largest percentage of participants improved (22%) in the Pacifica group and, within the Pacifica Lite and Waitlist groups, a similar percentage improved (14% and 15%, respectively) on the PANAS Negative. All intervention groups had a similar percentage that worsened on the PANAS Negative (Tables 9 and 10; Figures 7 and 8). For the PANAS Positive, a similar percentage of the participants in each group improved. There was no consistent trend within the groups for the percentage of participants who worsened in the PANAS Positive, though the Waitlist had the lowest percentage of those who did (18% compared to 23% for Pacifica and 27% for Pacifica Lite).

PROMIS Global Mental Health

There was a significant interaction effect between Pacifica and Waitlist for T1 score and intervention group for PROMIS Global Mental Health (t = -3.03, p = .003); however, there was not a significant interaction between T1 scores and the comparison between the Pacifica Lite and Waitlist groups (Table 5). Participants in the Pacifica group with T1 scores below 10.53 had significantly higher PROMIS Global Mental Health scores at T2 compared to those in the Waitlist group. The effect sizes were small within the significant region T1 < 10.53 (d = 0.24 at T1 = 9.6 to d = 0.20 at T1 = 10.53) which included 38% of the participants (Table 7). Above this baseline value of Global Mental Health, those in the Pacifica app group did not differ from the Waitlist group (d = 0.02 to 0.17 above T1 = 10.53). The within group effect was small (d = 0.25) for the Pacifica group as a whole, which indicated a small increase in global mental health. Both Pacifica and Pacifica Lite PROMIS Global Mental Health scores increased from T1 to T2, and a small between group effect was found between the Pacifica and Waitlist groups (Table 8). The largest percentage of participants improved (22%) in the Pacifica group whereas the largest percentage of participants worsened (13%) and least improved (11%) in the Waitlist group for PROMIS Global Mental Health (Tables 9 and 10).

PSS: Perceived Stress

There was a significant interaction effect between Pacifica and Waitlist for T1 score and intervention group for PSS (t = -2.43, p = .016); however, there was not a significant interaction between T1 scores and the Pacifica Lite vs. Waitlist comparison (Table 5). Participants in the Pacifica group with T1 scores above 2.29 had significantly

lower scores at T2 compared to those in the Waitlist group (Table 7). Below this baseline value of perceived stress, those in the Pacifica app group did not differ from the waitlist group at T2 (d = 0.08 to 0.15 below T1 = 2.29). The effect sizes were small (d = 0.20 at T1 = 2.29 to 0.23 at T1 = 3.20) at T1 scores within the significant region of T1 > 2.29 which included 31% of participants (Table 7). There was a small within group effect in the Pacifica group as a whole from T1 to T2, with a decrease following the intervention. This resulted in small between group effects for Pacifica to Waitlist and Pacifica Lite (Table 8). The largest percentage of participants improved (31%) and least worsened (13%) in the Pacifica group where the largest percentage of participants worsened (21%) and least improved (26%) in the Waitlist group for the PSS (Tables 9 and 10).

FFMQ: Dispositional Mindfulness

There was a significant interaction effect between Pacifica and Waitlist for T1 score and intervention group for FFMQ (t = -2.18, p = .03) (Table 5); however, the interaction effect T1 scores and Pacifica Lite vs. Waitlist did not reach significance (t = -1.75, p = .08) (Table 5). Participants in the Pacifica and Pacifica Lite groups with T1 scores below 3.22 had significantly greater scores at T2 compared to those in the Waitlist group (Table 7). Above this baseline value of trait based mindfulness, those in the Pacifica app group did not differ from the waitlist group at T2 (d = 0.00 to 0.10 above T1 = 3.22). Below this value, the effect sizes were small (d = 0.17 at T1 = 3.22 to 0.28 at T1 = 2.72) at T1 scores within the significant region of T1 < 3.22, which included 64% of participants (Table 7).

The FFMQ total score increased from time 1 to time 2 for both the Pacifica and Pacifica Lite groups as a whole whereas the Waitlist group decreased (d = 0.38 between Pacifica Lite and Waitlist; $d = 0.41$ between Pacifica and Waitlist) indicating a small to medium between-group effect size for both the Pacifica and Pacifica Lite groups compared to the Waitlist group (Table 8). A similar percentage of participants in the Pacifica and Pacifica Lite groups increased their FFMQ Total Score (29% and 26%, respectively) and decreased their FFMQ score (13% for both groups) and only 12% of the Waitlist group increased whereas 19% decreased (Tables 9 and 10).

AMPS: Applied Mindfulness

Four ANOVAs were conducted to assess between-group differences at T2 on the AMPS total and subscale scores. There was a significant overall effect of group condition only for the AMPS Negative Emotion Subscale, $F(2,317) = 3.24$, $p = .04$, partial $\eta^2 = .02$ (Table 6). Post-hoc analysis revealed that participants in the Pacifica condition reported significantly higher scores at T2 than those in the waitlist condition with a small to medium effect size ($p = .01$, $d = 0.36$) (Table 11; Figure 9). There were no significant group effects on the AMPS total, Decentering, or Positive Emotion Subscales.

Mobile App Rating Scale

Feedback from the users on the application quality as measured by the MARS was mostly positive (Table 12). Because the general population would not have access to the active control app co-designed specifically for this study, we focus here on the ratings provided by those participants randomized to the Pacifica app, given that small privacy changes and an in-app stress rating were the only functional differences between the

study version and the app commercially available. In particular, three subscales had mean scores above the scale's midpoint (3). These included the Functionality, Aesthetics, and App Specific Elements subscales. Within these subscales, items measuring performance, ease of use, navigation, layout, and graphics all had mean scores greater than 4. The Subjective Quality subscale was scored the lowest, although it was still above the scale's midpoint. In particular, the item 'Would you pay for this app?' had a high response rate of 'No' which strongly influenced this subscale's score.

The qualitative analyses only include participants who were assigned to the Pacifica group and completed survey two (n = 98). To gather qualitative feedback, these questions were asked: "If you found the application HELPFUL, please explain why or which aspects were HELPFUL to you", "What were your FAVORITE parts of the application or other Pacifica provided content?", and "If you found the application UNHELPFUL, please explain why or which aspects were UNHELPFUL to you". Coding responses was straightforward as a large portion of the responses simply named app features. Most (68%) of the participants found the app helpful. The helpful components specified in the free response section starting with the most frequently reported were health activity tracking (24%), being able to log-in and see their data over time (22%), the app making participants aware of their stress and feelings (21%), reminders (13%), mood tracking (12%), and the meditation exercises (10%). One participant commented, "It helped remind me of the little things in life that we could easily incorporate into my daily life that would increase my productivity." A smaller proportion of the group (35%) reported that they found the app unhelpful; while some participants specified components

that were unhelpful, there did not seem to be a consistent trend among these app elements. The majority (87%) gave feedback about their favorite part of the app with the greatest number commenting on the mood rating (32%), the stress ratings (26%), and health tracking (25%) and relaxation activities (25%). One participant stated their favorite aspects as: “I really enjoyed the daily tracking of health and mood. It made me take breaks from homework and other things to focus on my health.”

Table 4. Group Means at Pre- and Post-Intervention for DASS Total and Subscales, PSS, PANAS, PROMIS Global Mental Health and FFMQ Total

	Pacifica		Pacifica Lite		Waitlist	
Measure	Time 1 (n = 119)	Time 2 (n = 98)	Time 1 (n = 116)	Time 2 (n = 101)	Time 1 (n = 127)	Time 2 (n = 117)
DASS Total	0.89 (0.49)	0.81 (0.41)	0.79 (0.48)	0.82 (0.52)	0.86 (0.60)	0.88 (0.58)
DASS Depression	0.81 (0.61)	0.69 (0.51)	0.77 (0.64)	0.72 (0.61)	0.82 (0.74)	0.79 (0.73
DASS Anxiety	0.74 (0.58)	0.66 (0.48)	0.56 (0.49(	0.65 (0.56)	0.68 (0.66)	0.72 (0.62)
DASS Stress	1.12 (0.52)	1.08 (0.48)	1.05 (0.57)	1.10 (0.58)	1.09 (0.65)	1.16 (0.64)
PSS	2.02 (0.59)	1.90 (0.52)	1.95 (0.67)	1.96 (0.63)	1.93 (0.69)	1.91 (0.69)
PANAS Positive	2.48 (0.75)	2.49 (0.72)	2.40 (0.84)	2.45 (0.79)	2.50 (0.78)	2.56 (0.80)
PANAS Negative	1.71 (0.71)	1.68 (0.60)	1.62 (0.65)	1.85 (0.90)	1.71 (0.84)	1.84 (0.91)
Global Mental Health	11.52 (3.08)	12.26 (2.80)	12.03 (3.57)	12.28 (3.32)	12.19 (3.42)	12.20 (3.62)
FFMQ Total	3.08 (0.43)	3.16 (0.43)	3.06 (0.42)	3.13 (0.42)	3.12 (0.48)	3.09 (0.51)

Note: All models included the intervention group as the independent variable and the T1 score as a moderator variable. Values included in the table are for the main effect of intervention group and the interaction effect of the moderator with intervention group. * indicates a significant ($p < 0.05$) between group difference.

Table 5. Regression Analyses Assessing Between-Group Differences in Measures Post-Intervention with Pre-Intervention Score as a Moderator

	Pacifica v Waitlist			Pacifica Lite v Waitlist			Pacifica*T1			Pacifica Lite*T1		
Measure	coeff	*t*	*p*	coeff	*t*	*p*	coeff	*t*	*p*	coeff	*t*	*p*
DASS Total	0.02	0.26	0.80	0.03	0.33	0.74	-0.16	-1.73	0.09	-0.11	-1.14	0.26
DASS Depression	-0.05	-0.54	0.59	0.03	0.36	0.72	-0.11	-1.12	0.26	-0.09	-0.92	0.36
DASS Anxiety	0.03	0.37	0.71	0.04	0.41	0.68	-0.16	-1.80	0.07	-0.11	-1.32	0.19
DASS Stress	0.07	0.46	0.65	0.11	0.82	0.41	-0.18	-1.52	0.13	-0.18	-1.68	0.09
PSS	0.36	1.86	0.06	0.26	1.43	0.15	-0.23	-2.43	0.016*	-0.13	-1.48	0.14
PANAS Positive	-0.32	-0.99	0.32	-0.13	-0.44	0.66	0.10	0.81	0.42	0.02	0.19	0.85
PANAS Negative	0.66	2.83	0.005*	-0.04	-0.16	0.88	-0.48	-3.86	0.0001*	0.04	0.29	0.77
Global Mental Health	4.12	3.30	0.001*	2.24	1.91	0.06	-0.31	-3.03	0.003*	-0.17	-1.76	0.08
FFMQ Total	0.70	2.58	0.01*	0.56	2.09	0.04*	-0.19	-2.18	0.03*	-0.15	-1.75	0.08

Note: All models included the intervention group as the independent variable and the T1 score as a moderator variable. Values included in the table are for the main effect of intervention group and the interaction effect of the moderator with intervention group.

* indicates a significant ($p < 0.05$) between group difference.

Table 6. ANOVA Analyses Assessing Between-Group Differences in AMPS Measures Post-intervention

	Pacifica	Pacifica Lite	Waitlist				Post-Hoc (*p*)		
Measure	Time 2 (n = 98)	Time 2 (n = 101)	Time 2 (n = 117)	*F*	*p*	partial η2	Full*Lite	Full*Waitlist	Lite*Waitlist
AMPS Total	2.23 (0.66)	2.09 (0.71)	2.03 (0.81)	1.49	.23	0.01	.18	.028*	.41
AMPS Decentering	2.14 (0.68)	2.02 (0.62)	2.03 (0.76)	0.72	.49	0.01			
AMPS PosEmotion	2.45 (0.78)	2.29 (0.78)	2.31 (0.86)	1.31	.27	0.01			
AMPS NegEmotion	2.31 (0.64)	2.19 (0.74)	2.04 (0.82)	3.24	.04*	0.02	.38	.012*	.11

Table 7. Johnson-Neyman Boundary Scores, Significance, Effect Size, and Percentile of Participants for Measures with Significant Effect Sizes

T1 Score	*F*	*p*	*d*	Percentile
PSS				
1.40	0.54	0.58	0.08	22nd
1.70	0.43	0.65	0.07	37th
2.00	1.48	0.23	0.14	59th
2.29	3.02	0.05	0.20	67th
2.75	3.99	0.02	0.23	88th
3.20	4.02	0.02	0.23	98th
PANAS Negative				
1.00	1.11	0.33	0.14	25th
1.51	0.50	0.61	0.09	53rd
1.75	3.02	0.05	0.23	62nd
2.02	6.83	0.00	0.36	79th
2.70	11.00	0.00	0.46	88th
3.72	11.18	0.00	0.46	98th
PROMIS Global MH				
9.60	4.25	0.02	0.24	26th
10.53	3.03	0.05	0.20	38th
12.00	0.79	0.45	0.10	55th
13.60	0.03	0.97	0.02	63rd
15.20	0.75	0.47	0.10	85th
18.40	2.25	0.11	0.17	98th
FFMQ Total				
2.72	7.92	0.00	0.28	19th
2.98	6.69	0.00	0.26	41st
3.22	3.02	0.05	0.17	64th
3.38	1.02	0.36	0.10	75th
3.65	0.02	0.98	0.00	90th
4.04	0.30	0.74	0.05	98th

Table 8. Within-Group and Between-Group Effect Sizes (Cohen's d) from Pre-Intervention to Post-Intervention.

Measure	Group	Within Group d	Group Comparison	Between Group d
DASS21 Total	Pacifica	-0.25	Pacifica-Waitlist	-0.37
	Lite	-0.02	Pacifica-Lite	-0.23
	Waitlist	0.12	Lite-Waitlist	-0.15
DASS Depression	Pacifica	-0.27	Pacifica-Waitlist	-0.21
	Lite	-0.14	Pacifica-Lite	-0.12
	Waitlist	-0.06	Lite-Waitlist	-0.08
DASS Anxiety	Pacifica	-0.19	Pacifica-Waitlist	-0.36
	Lite	0.16	Pacifica-Lite	-0.35
	Waitlist	0.17	Lite-Waitlist	-0.01
DASS Stress	Pacifica	-0.12	Pacifica-Waitlist	-0.31
	Lite	-0.01	Pacifica-Lite	-0.11
	Waitlist	0.19	Lite-Waitlist	-0.20
PSS	Pacifica	-0.27	Pacifica-Waitlist	-0.24
	Lite	-0.04	Pacifica-Lite	-0.22
	Waitlist	-0.02	Lite-Waitlist	-0.02
PANAS-SF Positive	Pacifica	0.04	Pacifica-Waitlist	-0.07
	Lite	0.06	Pacifica-Lite	-0.02
	Waitlist	0.11	Lite-Waitlist	-0.04
PANAS-SF Negative	Pacifica	-0.01	Pacifica-Waitlist	-0.23
	Lite	0.30	Pacifica-Lite	-0.30
	Waitlist	0.22	Lite-Waitlist	0.07
Global Mental Health	Pacifica	0.25	Pacifica-Waitlist	0.22
	Lite	0.13	Pacifica-Lite	0.11
	Waitlist	0.03	Lite-Waitlist	0.11
FFMQ Total	Pacifica	0.29	Pacifica-Waitlist	0.41
	Lite	0.26	Pacifica-Lite	0.03
	Waitlist	-0.12	Lite-Waitlist	0.38

Note: Positive d indicates an increase in score between T1 and T2. For the DASS total and subscales, PANAS-SF Negative, and PSS a decrease in d indicates a reduction in symptoms and negative affective state. For the PROMIS-Global Mental Health subscale and PANAS-SF Positive subscale an increase in d indicates improved mental health and more positive affective state. For the FFMQ scales an increase in d indicates an increase

in mindfulness disposition. Between group effect sizes indicate the differences in changes in symptom and affective ratings between the Pacifica, Pacifica Lite, and Waitlist app participants.

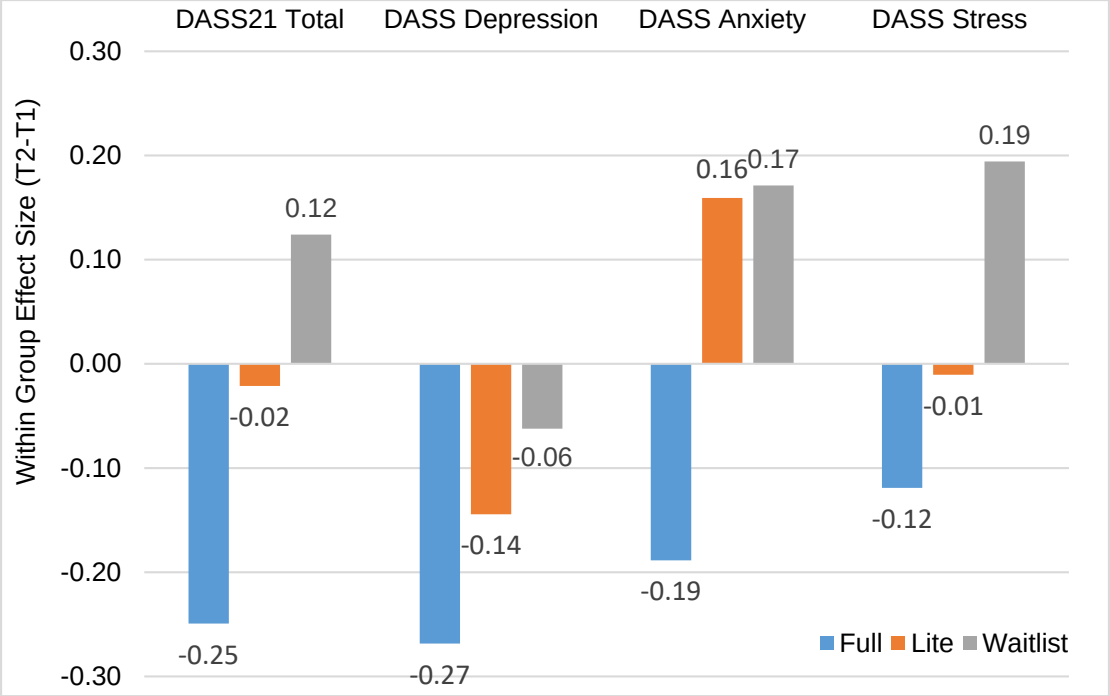

Figure 3. Within-Group Effect Size for DASS Total and Subscales across RCT Intervention Groups

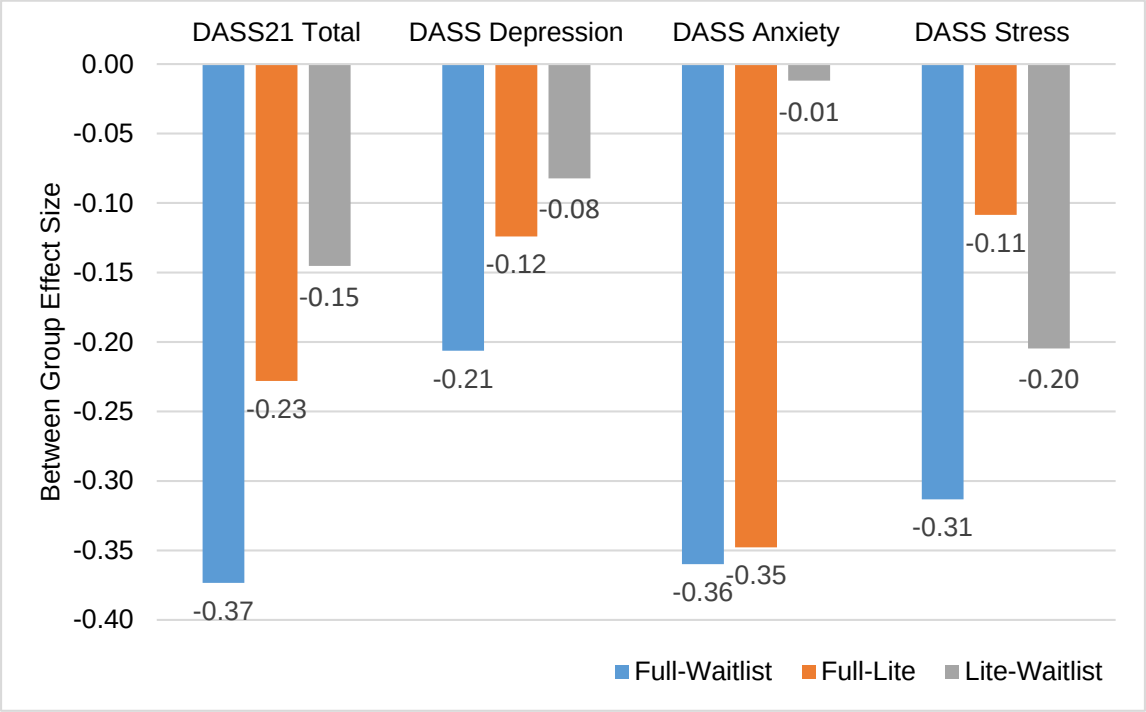

Figure 4. Between-Group Effect Size for DASS Total and Subscales for the RCT Intervention Groups

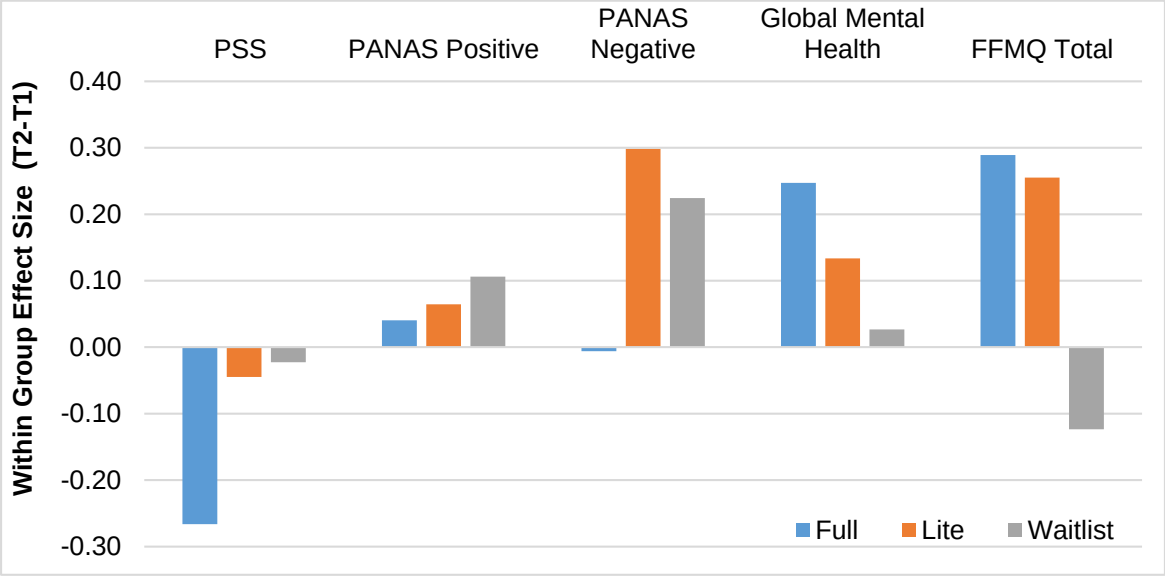

Figure 5. Within-Group Effect Size for Across the RCT Intervention Groups

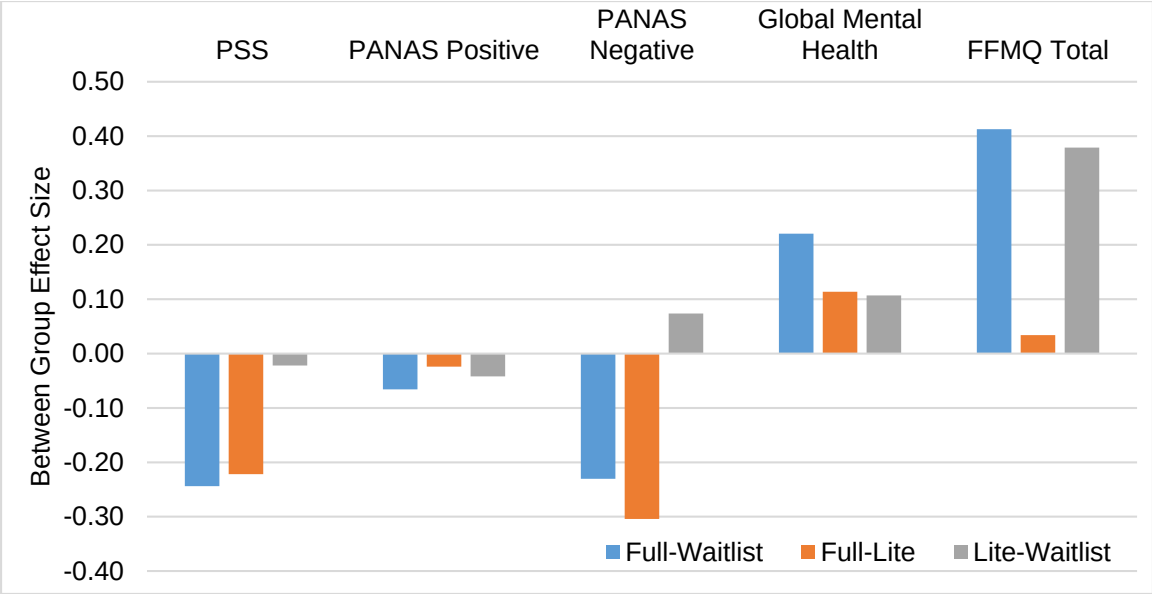

Figure 6. Between-Group Effect Size for Secondary Measures Across the RCT Intervention Groups

Table 9. Percentage of Participants Who Improved for Each Measure Between T1 and T2 Within Each Group

% Improved	Pacifica	Pacifica Lite	Waitlist
DASS Total	30%	27%	21%
DASS Anxiety	24%	20%	15%
DASS Depression	26%	26%	20%
DASS Stress	22%	25%	17%
PSS	31%	24%	26%
FFMQ Total	29%	26%	12%
Global Mental Health	22%	18%	11%
PANAS Negative	22%	14%	15%
PANAS Positive	26%	32%	26%

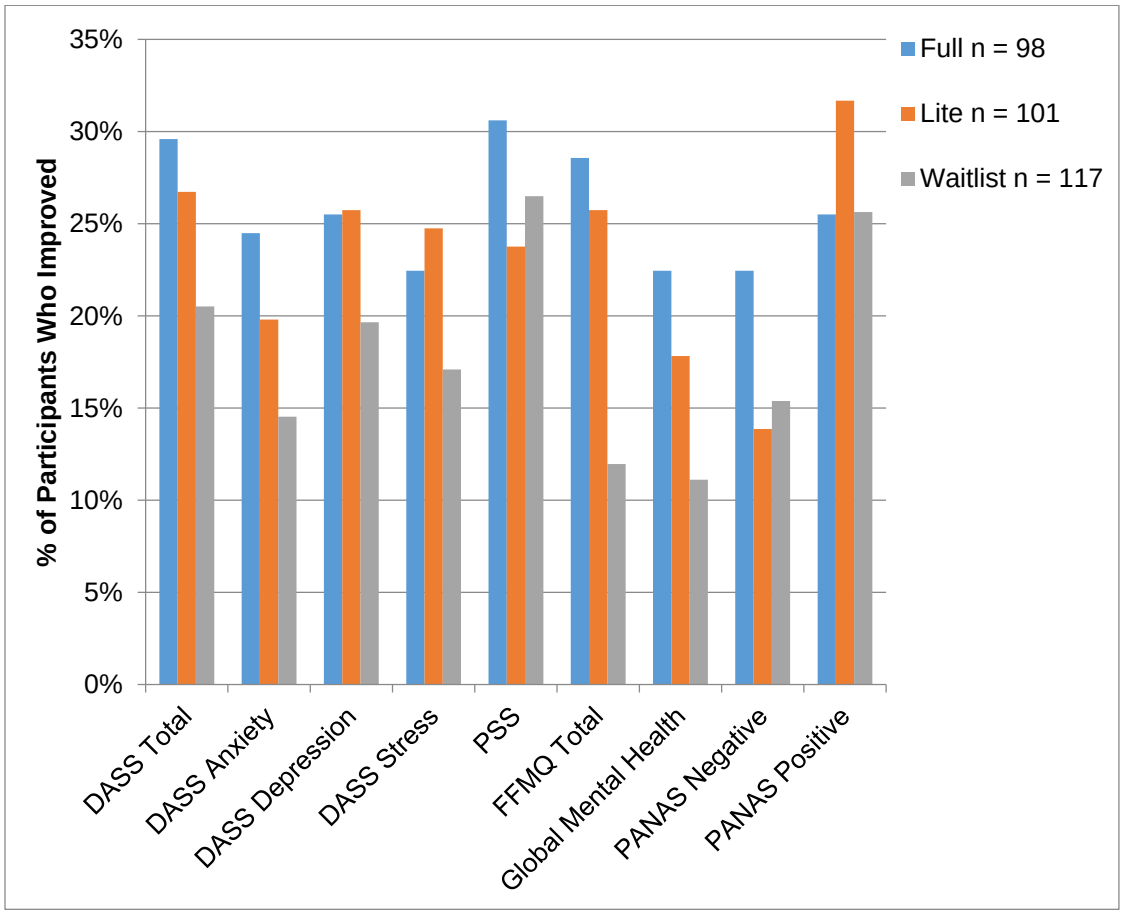

Figure 7. Percentage of Participants in the RCT Who Improved Beyond the RCI in Each Measure

Table 10. Percentage of Participants Who Worsened for Each Measure Between T1 and T2 Within Each Group

% Worsened	Pacifica	Pacifica Lite	Waitlist
DASS Total	14%	23%	30%
DASS Anxiety	9%	24%	21%
DASS Depression	8%	16%	18%
DASS Stress	20%	27%	31%
PSS	13%	21%	21%
FFMQ Total	13%	13%	19%
Global Mental Health	11%	10%	13%
PANAS Negative	27%	27%	26%
PANAS Positive	23%	27%	18%

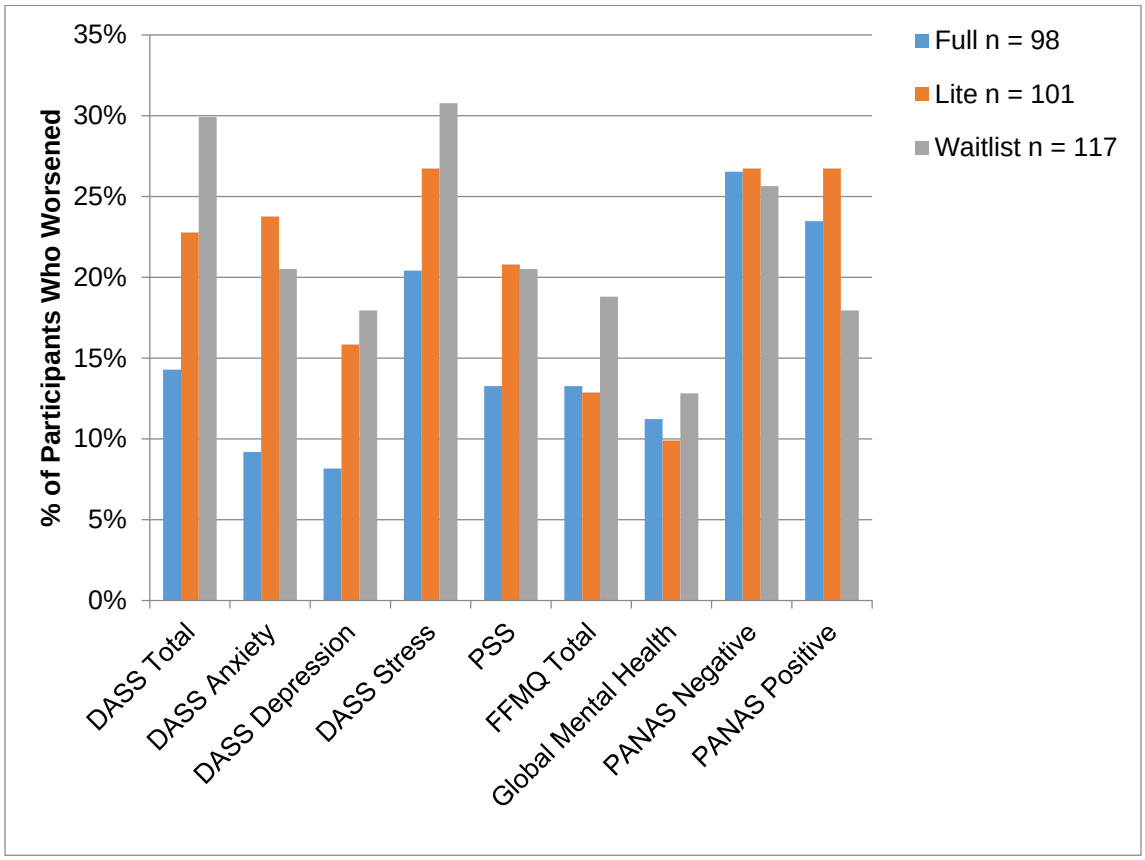

Figure 8. Percentage of Participants in the RCT Who Worsened Beyond the RCI in Each Measure

Table 11. Between Group Effect Sizes (Cohen's d) at Post-Intervention for AMPS

Measure	Group Comparison	*d*
AMPS Total	Pacifica-Waitlist	0.27
	Pacifica-Lite	0.21
	Lite-Waitlist	0.08
AMPS Positive Emotion	Pacifica-Waitlist	0.17
	Pacifica-Lite	0.21
	Lite-Waitlist	-0.03
AMPS Negative Emotion	Pacifica-Waitlist	0.36
	Pacifica-Lite	0.17
	Lite-Waitlist	0.19
AMPS Decentering	Pacifica-Waitlist	0.15
	Pacifica-Lite	0.18
	Lite-Waitlist	-0.02

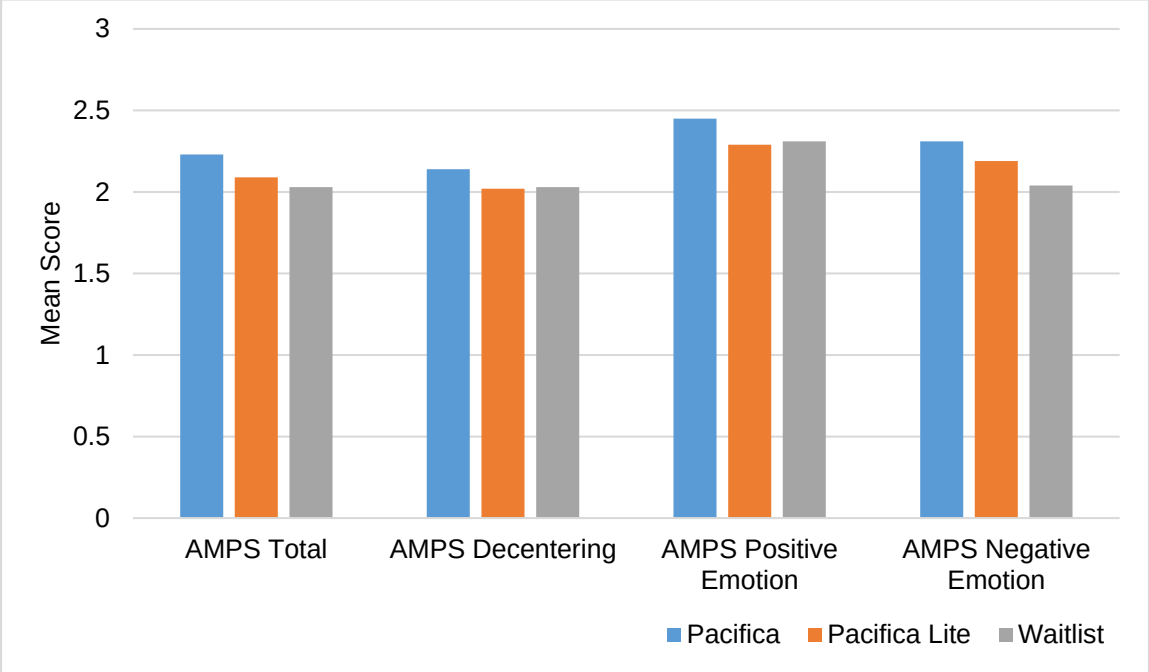

Figure 9. Mean AMPS Total and Subscale Scores Across Groups for the RCT

Table 12. MARS RCT App Quality Scores

Measure	Pacifica Mean (SD) (n = 98)
Engagement	3.54 (0.52)
Functionality	4.23 (0.59)
Aesthetics	4.18 (0.56)
Subjective Quality	3.20 (0.68)
App Specific	4.58 (0.83)

Linear Regression of DASS-Total Change with App Activity Use

An exploratory post hoc analysis of the relationship between app feature usage and change in DASS total scores was performed. The aim of this analysis with the Pacifica Lite group was to examine to what extent the number of non-psychotherapeutic activities completed, in isolation from the psychotherapy-based components offered in Pacifica, predicted DASS Total scores at T2 when controlling for T1 scores. The aim of this analysis with the Pacifica group was to examine to what extent the number of non-psychotherapy and psychotherapy-based activities completed when made available together predict DASS Total scores at T2 when controlling for T1 scores. All participants that completed the survey at T2, indicated to keep their data, and were not careless responders as defined above were included in this analysis. Cohorts one and two were combined (n = 99 participants in the Pacifica group; n = 102 participants in Pacifica Lite group).

The range and mean of app usage activities for the Pacifica and Pacifica Lite groups is included in Table 13. Most participants in Pacifica (89%) and Pacifica Lite (84%) completed Mood and Stress entries and some psychoeducation elements over the course of the two week study (Table 13 and Figure 10). In the Pacifica Full group, the

relax activities were used an average of 2.1 times over the course of the study and most (61%) users used these activities at least once. Less than half of the participants in the Pacifica group used the Thought Entry (48%) or Goals Entry (18%). The mood rating, stress rating, and psychoeducation use were not significantly skewed. The number of relaxation, goal, and thought entries were highly skewed (skew > 2). Using Grubbs' test, there were no outliers (z-score > 4) for the number of mood entries, stress entries, relaxation entries, or goal entries. There was one outlier for the number of thoughts entries; however, it appeared to be a valid data point and was kept in the analysis.

An analysis of the relationship between change in outcome measures from T1 to T2 and the usage of the app components was performed using linear regression. The value of the DASS total at T1 was used as a covariate, the value of the DASS Total at T2 was the dependent variable, and the usage of the app component was the independent variable. There were no significant relations between use of specific app elements and the DASS total score in either group (Appendix C).

Table 13. Descriptive Statistics for Activity Usage

Pacifica	Range	Mean Number of Entries (SD)	% of Participants Who Used Component At Least Once
Mood Entry	[0:38]	12.8 (7.8)	89%
Stress Entry	[0:38]	13.1 (7.8)	89%
Relax Entry	[0:18]	2.1 (3.4)	61%
Psychoed Entry	[0:10]	5.7 (2.9)	89%
Thought Entry	[0:17]	1.6 (2.9)	48%
Goal Entry	[0:14]	0.78 (2.2)	18%
Pacifica Lite			
Mood Entry	[0:44]	14.9 (9.8)	84%
Stress Entry	[0:46]	15.0 (9.9)	84%
Psychoed Entry	[0:8]	2.4 (1.8)	84%

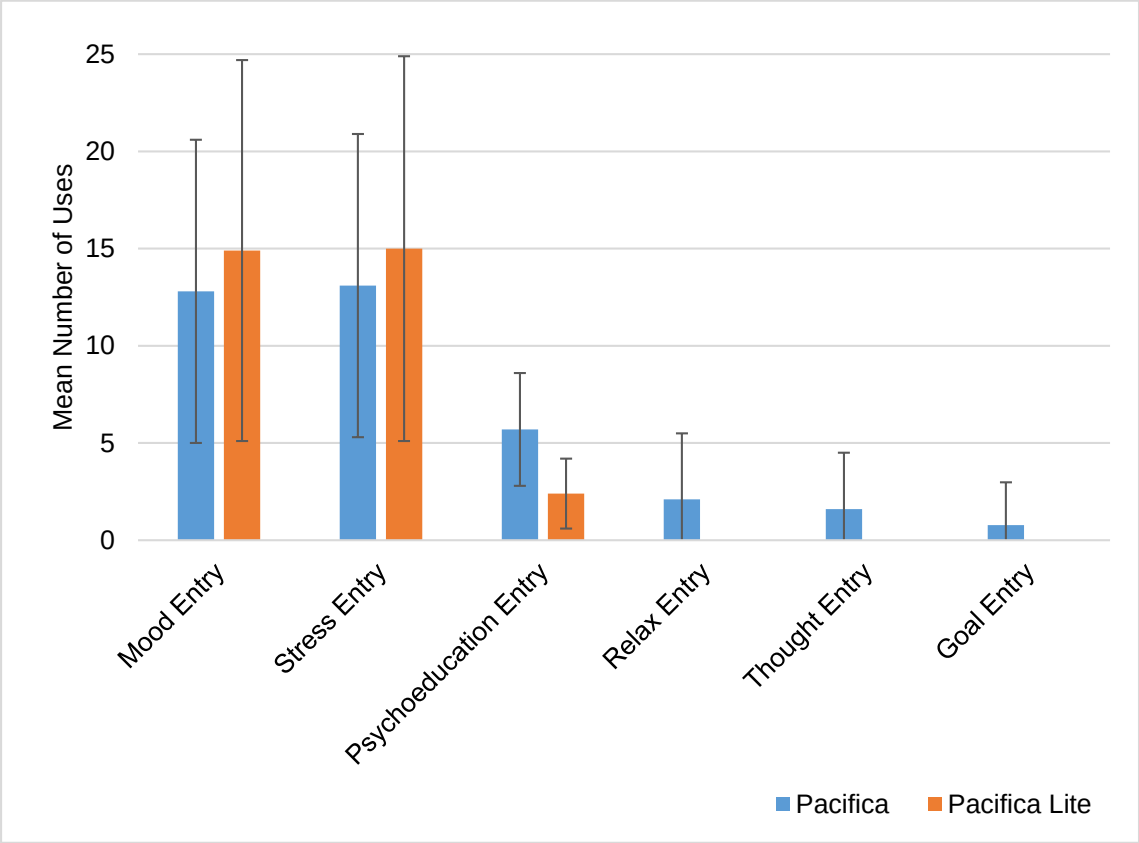

Figure 10. Mean Number of App Activity Uses Across Participants for Pacifica and Pacifica Lite Apps

Chapter 4: Discussion

The purpose of this study was to investigate the efficacy of a commercially available mindfulness and CBT-based mobile app aimed at improving stress and well-being in a college student sample during a demanding two-week period of the semester (i.e., during final exams). One of the primary criticisms of the mHealth literature in evaluating the efficacy of apps are the vast number of underpowered studies with poor quality designs. As a small to medium effect is difficult to detect reliably with smaller numbers of participants, the size of this study is a significant contribution to the mHealth literature. Furthermore, the experimental designs of mHealth studies often compare the intervention to a waitlist control or active controls of a different modality or content than the intervention condition. The experimental design of the current study went beyond the traditional RCT by co-developing and including an active control app in addition to a waitlist control group, resulting in a parallel group RCT design with over 300 participants. In this way, the current study is better suited to describe contributions unique to the purported therapeutic components of the Pacifica app rather than simply any app delivered intervention. Additionally, the effects of baseline scores on group outcomes were examined. An exploratory investigation of how app use was associated with changes in measures of well-being was undertaken. Finally, because the quality of a mHealth delivered intervention can greatly influence frequency of use (Bakker, Kazantzis, Rickwood, & Richard, 2016), participants rated usability and other quality metrics of the Pacifica mobile app. This section will summarize the study's key findings, limitations, and promising future directions.

Efficacy of Pacifica

Intervention Effects on DASS

The results of this study do not support the hypothesis that the intervention would lead to a significant difference in DASS total scores for the Pacifica group compared to the Waitlist control group. There were no significant effects of group on DASS total scores or subscale scores and the effect of the interventions did not vary based on participants' baseline scores. In other words, the DASS scores were not altered by two weeks of app use during a demanding academic period. However, there were small to medium effect sizes between the Pacifica and waitlist groups (d = 0.37) on the DASS total, and small to medium effect sizes between the Pacifica and waitlist groups for both the Stress and Anxiety subscales, which aligns with previous work showing small to medium effects in reduction of stress and anxiety with other apps (Gee, Griffiths, & Gulliver, 2015; Firth, et al., 2017; Neary & Schueller, 2018).

Intervention Effects on Other Measures

There were significant effects of the Pacifica app intervention on PSS, PANAS Negative, and PROMIS-Global Mental Health when including the interaction between the participants' baseline scores and intervention group. For the PSS, participants with baseline scores above 2.29 had larger reductions in the Pacifica group compared to the Waitlist group (d = .20); however, there was no difference between Pacifica and Pacifica Lite at any levels of baseline scores. For the PANAS Negative, participants with baseline scores greater than 1.75 had greater reductions at T2 in the Pacifica group compared to the Waitlist (d = .23-.46). For the PROMIS-Global Mental Health, participants with

baseline scores lower than 10.53 had greater increases at T2 in the Pacifica group compared to the Waitlist (*d* = .20-.24). Combined, these results suggest that participants with higher baseline symptomology, including increased stress and negative affect, and lower global mental health, improve more from the Pacifica intervention than those with lower levels of baseline symptoms (when Pacific is compared to the Waitlist control group). However, on these measures, there were no differences between Pacifica and Pacifica Lite or between Pacifica Lite and the Waitlist control group. These results are similar to previous studies that examined the moderating effect of baseline clinical scores on the effects of interventions (Bower et al., 2013; Hedman, Andersson, Lekander & Ljótsson). Significant improvements were found for those with higher levels of anxiety or depressive symptoms at baseline when comparing low intensity interventions vs. control.

There were significant effects between groups on measures of mindfulness. The FFMQ measures trait based mindfulness along five facets including the self-reported aptitude to describe and observe experiences, without judgment or reactivity, and act with awareness. Participants with lower baseline scores, below 3.22, had greater increases in the Pacifica group compared to the Waitlist (*d* = .17-.28). The small effect between the Pacifica group and the Waitlist control group on the FFMQ total score suggests that a combination of tracking mood, stress, the ability to review data over time via a within-app line chart, and mindfulness-based micro-interventions may increase dispositional mindfulness. van Emmerik and colleagues (2017) performed an 8 week RCT in adults using a mindfulness app intervention versus a waitlist control and found a large (*d* = 0.77) effect between groups on the FFMQ total score at the post-intervention time point.

This effect size was larger than that found in the current study when comparing the full app to the waitlist control; however, the duration of the van Emmerik study was longer than the current study. In the current study, there was also a small to medium effect of the Pacifica app compared to Waitlist control on applied mindfulness as measured by the AMPS, although there was no significant difference between Pacifica and Pacifica Lite.

App Quality to Maintain Usage/Interest

The user experience and the effectiveness of the app in promoting well-being are vital requirements for both initial engagement of the user and stimulating continuing use for a mHealth app's intended purpose (Bakker, Kazantzis, Rickwood, & Richard, 2016). If the user is not satisfied with the functionality, aesthetics, and interactions with the app, they will not use it regardless of content fidelity to in-person treatment. The participants rated the Pacifica app well in terms of its functionality and aesthetics using the MARS. The Pacifica app obtained MARS mean scores of 4.23 for functionality and 4.18 for aesthetics out of 5, which translates as between good and excellent. In a recent survey of mindfulness apps (Mani, Kavanagh, Hides & Stoyanov, 2015) the mean score for functionality was 3.03 and aesthetics was 3.91, illustrating that Pacifica is above these average for mindfulness based apps. The majority of participants also considered the Pacifica app to be helpful (68%) and most frequently cited the health activity and stress tracking elements to be particularly beneficial. However, most participants reported that they would not pay for the Pacifica app, an aspect that may be difficult for the long term viability of the product for this version. The majority of the participants used the mood and stress ratings, relaxation activities, and psychoeducation components; however, fewer

than 50% of participants who had access to the thought and goal entry elements used these features.

Limitations

A primary limitation of this study was its duration. A previously conducted unpublished secondary analysis of commercial users noted a relatively short time span of engagement with a prior version of the Pacifica app using the mood-rating feature under no time limitations (Vu & Frazier, 2016). For this RCT study, the intervention duration period was shorter than recent mHealth studies with college students, which have lasted between four to eight weeks (Harrer et al., 2018; Lee & Jung, 2018). The two week intervention period may not be extensive enough for participants to report noticeable changes in anxiety, stress, and depression symptoms measured by the DASS, even when using the micro-interventions daily. A longer term examination of app use and behavior would be needed to more accurately understand how college students would use and benefit from the Pacifica app in relation to school-related stress and anxiety as well as other life stressors. Similarly, a compensated follow up survey period with high response rate would have been ideal to track usage and changes in outcome measures for participants over time.

The timing of the intervention roll-out could have influenced the results either by amplifying or reducing the effect of the intervention. The final exam period may have constricted the potential for the intervention to lead to greater changes in stress and well-being given that it is a demanding two weeks for students overall. Conversely, during the course of the regular semester, the Pacifica app may not be as effective without the added

increase in situational stress due to exams. Additional studies would be needed to observe fluctuations in stress during the course of a standard semester, with assessments at multiple time points to evaluate how app use interacts with the amount of stress a student is dealing with at any given time.

While a positive user experience is the minimal requirement for initial app use and engagement, fidelity to curative components of in-person psychotherapy cannot be overlooked if the promise of the mHealth products is improved well-being. Because many of the psychotherapeutic micro-interventions of the Pacifica app were broadly reinterpreted and combined several source materials from different psychotherapy sources, fidelity could not be accurately measured. Likewise, the mHealth field has yet to develop an agreed upon process of translating evidence-based treatments to the scale and format of app-delivered micro-interventions.

The frequency of app use was not specified in the participant instructions beyond using the app a minimum of daily, with the option to use it more if the participant felt so inclined. These instructions were deliberate as there were no study hypotheses regarding a pre-defined frequency of app use. Similarly, many of the app components would not be appropriate to use more than daily (e.g., health activities such as timing of medications and sleep). A better understanding of optimal selection and dosage of micro-interventions for specific problems would elucidate how adjunctive or complementary services (e.g., email reminders, in-app push notifications, or in-person therapists) could play a role in the Pacifica app recommending these activities to users.

Another limitation of this study was its sample; most participants were White, English speaking, and undergraduate students currently enrolled in a psychology course, many of whom may have acquired basic knowledge about psychotherapy through psychology related coursework. The Pacifica app was designed instead for the general adult population, many of whom may be naïve to psychotherapy interventions. In addition, examples provided by the CBT and mindfulness-based components of the Pacifica app, as well as the psychoeducation and health-tracking features, did not use situations and activities commonly experienced by students.

Implications

The findings of this study support the use of a mobile app delivered suite of CBT and mindfulness-based micro-interventions by college students with particular levels of distress during a brief and demanding final exam period. Students with higher baseline stress scores (PSS), greater negative affect (PANAS Negative), and lower general mental health scores (PROMIS Global Mental Health) improved more (relative to being on a waitlist but not relative to using the Pacifica lite app) than those with lower baseline scores. Regarding trait-based mindfulness, participants that had lower initial mindfulness scores increased more than those with higher initial mindfulness scores when comparing the Pacifica and Waitlist conditions, suggesting that the app may be used by students starting at lower mindfulness levels to increase trait-based mindfulness. Somewhat unexpectedly, higher mindfulness scores were also reported for participants using Pacifica Lite, the condition without access to the explicit mindfulness components of the app. This study's findings invite the prospect of increasing mindfulness-based traits with

access to brief self-monitoring interventions, without explicit mindfulness skills over a short time span. Given the relative ease of access, use, positive feedback, and automated nature of the app, Pacifica and mHealth apps like it may be a feasible option for students to access psychotherapeutic content and engage in self-monitoring of wellness activities.

To the author's knowledge, this is the first large RCT of a mHealth app with an active control app conducted to date. The pilot and RCT studies tested two slightly varied intervention versions of Pacifica that were commercially available over the study duration; it should be noted that these versions were neither developed as a stand-alone treatment nor to complement in-person psychotherapy for specific mental health disorders. Given the need for mental health services and the shortage of credentialed providers, a persuasive argument could be made that any readily available wellness and psychotherapeutic content could be beneficial as a primary prevention intervention. However, academic health centers and healthcare systems should nonetheless be conservative when recommending Pacifica beyond its intended design as an additional wellness support product.

Future Directions

The vast majority of individuals in the United States have smartphones (Pew Research, 2018) and, with the increasing demand for mental health services, commercial mHealth apps are an abundant and accessible resource for introducing people to psychotherapy content. Nonetheless, the commercial mHealth field is greatly lacking in rigor that adequately assesses the quality and efficacy of these app-delivered interventions for mental health consumers, providers, and healthcare systems alike.

Broadly, this study aimed to add to the literature on the efficacy of apps in delivering micro-interventions to bolster mental health and well-being using a popular pre-existing and commercially available smartphone app.

Pacifica's psychotherapeutic components, and mindfulness-based interventions specifically, appeared to loosely package elements of third-wave therapies including Mindfulness Based Stress Reduction, Acceptance and Commitment Therapy, and even Positive Psychology. To accurately study the effects of intervention dose and type, the translation process from face-to-face psychotherapy source materials to app-delivered micro-interventions could be made clearer by developers and mental health advisors in the commercial app field. Similarly, a set of best-practice guidelines of appropriately scaling psychotherapy interventions to mHealth apps would be a benefit to the field as a whole. Without this groundwork, it is difficult to account for the ways in which the degree of fidelity of app-delivered interventions to empirically tested treatments contributes to mHealth intervention outcomes.

The intensity and chronicity of mindfulness practice needed to induce change in mindful characteristics is largely untested and unknown in the mindfulness intervention literature despite acknowledgment that mindfulness can be experienced in discrete and brief moments as well as in prolonged meditation (Davidson & Kaszniak, 2015). Dismantling studies designed to examine the timing and delivery of self-monitoring and mindfulness components via apps could provide a fascinating step forward for the mindfulness literature.

Further work that examines the effect of the use of Pacifica on measures of stress specifically experienced by this participant sample (e.g., academic stress in college students) is needed to uncover other types and aspects of stress that are more likely to be responsive to the app. In addition, a future study with a research design that contains several follow up periods at various intervals would illuminate the time course of effects.

As innovations in mHealth continue to develop, we cannot lose sight of efforts to protect, adequately inform, and involve those populations that have few, if any, treatment alternatives. At this time, commercially available mHealth apps can forego the lengthy process of FDA medical device approval if their promises fall just short of improving formally recognized mental health diagnoses (e.g., “Depression” vs. “Major Depressive Disorder). These slight differences in terminology can be overlooked by even the most savvy and health literate of mental health consumers. Through the development of testing and implementation guidelines that place consumer rights at the center, content area experts, developers, and under-resourced healthcare systems can more judiciously offer specific mHealth technologies for persons who would best stand to benefit from these products. Efforts to build and maintain infrastructures that support interprofessional collaboration within academia as well as between academia, industry, and consumers (general and clinical populations) in *both development and testing* of mHealth apps would undoubtedly improve the quality of innovating effective mHealth products.

References

American College Health Association. (2018) *American College Health Association-National College Health Assessment II: Reference Group Undergraduates Executive Summary Fall 2017.* Hanover, MD: American College Health Association.

American Psychiatric Association. (2018). *App Evaluation Model*. Retrieved from https://www.psychiatry.org/psychiatrists/practice/mental-health-apps/app-evaluation-model.

American Psychological Association. (2017). *Stress in America: The State of Our Nation.* Stress in America Survey.

American Psychological Association. (2018). *Stress in America: Uncertainty About Health Care.* Stress in America Survey.

Auerbach, R. P., Alonso, J., Axinn, W. G., Cuijpers, P., Ebert, D. D., Green, J. G., ... & Nock, M. K. (2016). Mental disorders among college students in the World Health Organization world mental health surveys. *Psychological Medicine*, *46*(14), 2955-2970.

Baer, R. A., Smith, G. T., Hopkins, J., Krietemeyer, J., & Toney, L. (2006). Using self-report assessment methods to explore facets of mindfulness. *Assessment*, *13*(1), 27-45.

Bakker, D., Kazantzis, N., Rickwood, D., & Rickard, N. (2016). Mental health smartphone apps: review and evidence-based recommendations for future developments. *JMIR Mental Health*, *3*(1).

Barak, A., Hen, L., Boniel-Nissim, M., & Shapira, N. A. (2008). A comprehensive review and a meta-analysis of the effectiveness of internet-based psychotherapeutic interventions. *Journal of Technology in Human Services*, *26*(2-4), 109-160.

Ben-Zeev, D., Schueller, S. M., Begale, M., Duffecy, J., Kane, J. M., & Mohr, D. C. (2014). Strategies for mHealth research: Lessons from 3 mobile intervention

studies. *Administration and Policy in Mental Health and Mental Health Services Research*, *42*(2), 157-167.

Bower, P., Kontopantelis, E., Sutton, A., Kendrick, T., Richards, D. A., Gilbody, S., ... & Meyer, B. (2013). Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data. *BMJ*, *346*.

Caldwell, K., Harrison, M., Adams, M., Quin, R. H., & Greeson, J. (2010). Developing mindfulness in college students through movement-based courses: effects on self-regulatory self-efficacy, mood, stress, and sleep quality. *Journal of American College Health*, *58*(5), 433-442.

Campbell, B., Caine, K., Connelly, K., Doub, T., & Bragg, A. (2015). Cell phone ownership and use among mental health outpatients in the USA. *Personal and Ubiquitous Computing*, *19*(2), 367-378.

Carras, M. C., Mojtabai, R., Furr-Holden, C. D., Eaton, W., & Cullen, B. A. (2014). Use of mobile phones, computers and internet among clients of an inner-city community psychiatric clinic. *Journal of Psychiatric Practice*, *20*(2), 94-103.

Carter, M. C., Burley, V. J., Nykjaer, C., & Cade, J. E. (2013). Adherence to a smartphone application for weight loss compared to website and paper diary: pilot randomized controlled trial. *Journal of Medical Internet Research*, *15*(4).

Cavanagh, K., Strauss, C., Cicconi, F., Griffiths, N., Wyper, A., & Jones, F. (2013). A randomised controlled trial of a brief online mindfulness-based intervention. *Behaviour Research and Therapy*, *51*(9), 573-578.

Cella, D., Riley, W., Stone, A., Rothrock, N., Reeve, B., Yount, S., ... & Cook, K. (2010). The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *Journal of Clinical Epidemiology*, *63*(11), 1179-1194.

Chambers, D. A., & Azrin, S. T. (2013). Research and services partnerships: partnership: a fundamental component of dissemination and implementation research. *Psychiatric Services*, 64(6), 509-511.

Clough, B. A., & Casey, L. M. (2015). The smart therapist: A look to the future of smartphones and mHealth technologies in psychotherapy. *Professional Psychology: Research and Practice*, *46*(3), 147–153.

Chan, S., Torous, J., Hinton, L., & Yellowlees, P. (2015). Towards a Framework for Evaluating Mobile Mental Health Apps. *Telemedicine and e-Health*, *21*(12), 1038-1041.

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, *24*(4), 385-396.

Conley, C. S., Durlak, J. A., Shapiro, J. B., Kirsch, A. C., & Zahniser, E. (2016). A meta-analysis of the impact of universal and indicated preventive technology-delivered interventions for higher education students. *Prevention Science*, *17*(6), 659-678.

Coulon, S. M., Monroe, C. M., & West, D. S. (2016). A systematic, multi-domain review of mobile smartphone apps for evidence-based stress management. *American Journal of Preventive Medicine*, *51*(1), 95-105.

Davies, E. B., Morriss, R., & Glazebrook, C. (2014). Computer-delivered and web-based interventions to improve depression, anxiety, and psychological well-being of university students: a systematic review and meta-analysis. *Journal of Medical Internet Research*. *16*(5), 18-39.

Donker, T., Petrie, K., Proudfoot, J., Clarke, J., Birch, M. R., & Christensen, H. (2013a). Smartphones for smarter delivery of mental health programs: a systematic review. *Journal of Medical Internet Research*, *15*(11).

Donker, T., Bennett, K., Bennett, A., Mackinnon, A., van Straten, A., Cuijpers, P., ... & Griffiths, K. M. (2013b). Internet-delivered interpersonal psychotherapy versus internet-delivered cognitive behavioral therapy for adults with depressive symptoms: randomized controlled noninferiority trial. *Journal of Medical Internet Research*, *15*(5).

East, M. L., & Havard, B. C. (2015). Mental Health Mobile Apps: From Infusion to Diffusion in the Mental Health Social System. *JMIR Mental Health*, *2*(1), e10.

Eisenberg, D., Golberstein, E., & Gollust, S. E. (2007). Help-seeking and access to mental health care in a university student population. *Medical Care*, *45*, 594-601.

Eysenbach, G., & Group, C. E. (2011). CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. *Journal of Medical Internet Research*, *13*(4).

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149-1160.

Frazier, P., Anders, S., Perera, S., Tomich, P., Tennen, H., Park, C., & Tashiro, T. (2009). Traumatic events among undergraduate students: Prevalence and associated symptoms. *Journal of Counseling Psychology*, *56*(3), 450–460.

Frazier, P., Meredith, L., Greer, C., Paulsen, J. A., Howard, K., Dietz, L. R., & Qin, K. (2014). Randomized controlled trial evaluating the effectiveness of a web-based stress management program among community college students. *Anxiety, Stress, & Coping*, *28*(5), 1-11.

Firth, J., Torous, J., Nicholas, J., Carney, R., Rosenbaum, S., & Sarris, J. (2017). Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. *Journal of Affective Disorders*, *218*, 15-22.

Gallagher, R. (2008). *National Survey of Counseling Center Directors – 2008.* The International Association of Counseling Services, Inc. Retrieved from http://d-scholarship.pitt.edu/28169/1/2008_National_Survey_of_Counseling_Center_Directors.pdf.

Gallagher, R.P. (2015). *National Survey of counseling centers – 2014*. The International Association of Counseling Services, Inc. Retrieved from http://d-scholarship.pitt.edu/28178/1/survey_2014.pdf

Gee, B. L., Griffiths, K. M., & Gulliver, A. (2015). Effectiveness of mobile technologies delivering Ecological Momentary Interventions for stress and anxiety: a systematic review. *Journal of the American Medical Informatics Association*, *23*(1), 221-229.

Grassi, A., Gaggioli, A., & Riva, G. (2009). The green valley: the use of mobile narratives for reducing stress in commuters. *Cyberpsychology & Behavior*, *12*(2), 155-161.

Harrer, M., Adam, S. H., Fleischmann, R. J., Baumeister, H., Auerbach, R., Bruffaerts, R. ,... & Ebert, D. D. (2018). Effectiveness of an internet-and app-based intervention for college students with elevated stress: randomized controlled trial. *Journal of Medical Internet Research*, *20*(4).

Hayes, A. F., & Montoya, A. K. (2017). A tutorial on testing, visualizing, and probing an interaction involving a multicategorical variable in linear regression analysis. *Communication Methods and Measures*, *11*(1), 1-30.

Hays, R. D., Bjorner, J. B., Revicki, D. A., Spritzer, K. L., & Cella, D. (2009). Development of physical and mental health summary scores from the patient-reported outcomes measurement information system (PROMIS) global items. *Quality of Life Research*, *18*(7), 873-880.

Hedman, E., Andersson, E., Lekander, M., & Ljótsson, B. (2015). Predictors in Internet-delivered cognitive behavior therapy and behavioral stress management for severe health anxiety. *Behaviour Research and Therapy*, *64*, 49-55.

Heron, K. E., & Smyth, J. M. (2010). Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. *British Journal of Health Psychology*, *15*(1), 1-39.

Herrero, J., & Meneses, J. (2006). Short Web-based versions of the perceived stress (PSS) and Center for Epidemiological Studies-Depression (CESD) Scales: a comparison to pencil and paper responses among Internet users. *Computers in Human Behavior*, *22*(5), 830-846.

Hintz, S., Frazier, P. A., & Meredith, L. (2015). Evaluating an online stress management intervention for college students. *Journal of Counseling Psychology*, *62*(2), 137.

Howell, R. T., Rodzon, K. S., Kurai, M., & Sanchez, A. H. (2010). A validation of well-being and happiness surveys for administration via the Internet. *Behavior Research Methods*, *42*(3), 775-784.

Hudd, S. S., Dumlao, J., Erdmann-Sager, D., Murray, D., Phan, E., Soukas, N., & Yokozuka, N. (2000). Stress at college: Effects on health habits, health status and self-esteem. College Student Journal, 34(2).

Hunt, J., & Eisenberg, D. (2010). Mental health problems and help-seeking behavior among college students. *Journal of Adolescent Health*, *46*(1), 3-10.

Hurst, C. S., Baranik, L. E., & Daniel, F. (2013). College student stressors: A review of the qualitative research. *Stress and Health*, *29*(4), 275-285.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*(1), 12.

Kazdin, A. E. (2014). Evidence-based psychotherapies I: qualifiers and limitations in what we know. *South African Journal of Psychology*, *44*(4), 381-403.

Kim, H. Y. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry & Endodontics*, *38*(1), 52-54.

Lambert, J.J., & Ogles, B.M. (2013). The efficacy and effectiveness of psychotherapy. In M.J. Lambert (Ed.), *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change* (6th ed., 169-218). New York: Wiley.

Lappalainen, P., Kaipainen, K., Lappalainen, R., Hoffrén, H., Myllymäki, T., Kinnunen, M. L., ... & Korhonen, I. (2013). Feasibility of a personal health technology-based psychological intervention for men with stress and mood problems: randomized controlled pilot trial. *JMIR Research Protocols*, *2*(1).

Lee, R. A., & Jung, M. E. (2018). Evaluation of an mHealth app (destressify) on university students' mental health: Pilot trial. *JMIR Mental Health*, *5*(1).

Lemaire, J. B., Wallace, J. E., Lewin, A. M., de Grood, J., & Schaefer, J. P. (2011). The effect of a biofeedback-based stress management tool on physician stress: a randomized controlled clinical trial. *Open Medicine*, *5*(4), e154.

Li, M. J., Black, D. S., & Garland, E. L. (2016). The Applied Mindfulness Process Scale (AMPS): A process measure for evaluating mindfulness-based interventions. *Personality and Individual Differences*, *93*, 6-15.

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*,*33*(3), 335-343.

Mani, M., Kavanagh, D. J., Hides, L., & Stoyanov, S. R. (2015). Review and evaluation of mindfulness-based iPhone apps. *JMIR mHealth and uHealth*, *3*(3).

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*(3), 437.

Moneta, G. B., Vulpe, A., & Rogaten, J. (2012). Can positive affect "undo" negative affect? A longitudinal study of affect in studying. *Personality and Individual Differences*, *53*(4), 448-452.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, *7*(1), 105.

Mosso, J. L., Gorini, A., De La Cerda, G., Obrador, T., Almazan, A., Mosso, D., ... & Riva, G. (2009). Virtual reality on mobile phones to reduce anxiety in outpatient surgery. *Studies in Health Technology and Informatics*, *142*(25), 195-200.

Muñoz, R. F. (2010). Using evidence-based internet interventions to reduce health disparities worldwide. *Journal of Medical Internet Research*, *12*(5).

Neary, M., & Schueller, S. M. (2018). State of the field of mental health apps. *Cognitive and Behavioral Practice*.

Newman, M. G., Przeworski, A., Consoli, A. J., & Taylor, C. B. (2014). A randomized controlled trial of ecological momentary intervention plus brief group therapy for generalized anxiety disorder. *Psychotherapy*, *51*(2), 198.

Pacifica Labs Inc. (2015). *Pacifica.* Retrieved from http://www.thinkpacifica.com/

Pacifica Labs Inc. (2018). *Pacifica.* Retrieved from http://www.thinkpacifica.com/

Pallavicini, F., Algeri, D., Repetto, C., Gorini, A., & Riva, G. (2009). Biofeedback, virtual reality and mobile phones in the treatment of Generalized Anxiety Disorder (GAD): A phase-2 controlled clinical trial. *Journal of CyberTherapy and Rehabilitation*, *2*(4), 315-327.

Pew Research Center (2018, February 5) *Mobile technology fact sheet.* Retrieved from http://www.pewinternet.org/fact-sheet/mobile/

Proudfoot, J., Clarke, J., Birch, M. R., Whitton, A. E., Parker, G., Manicavasagar, V., ... & Hadzi-Pavlovic, D. (2013). Impact of a mobile phone and web program on symptom and functional outcomes for people with mild-to-moderate depression, anxiety and stress: a randomised controlled trial. *BMC Psychiatry*, *13*(1), 312.

Psyberguide. (2018). *Pacifica*. Retrieved from https://psyberguide.org/apps/pacifica/.

Ratelle, C. F., Simard, K., & Guay, F. (2013). University students' subjective well-being: The role of autonomy support from parents, friends, and the romantic partner. *Journal of Happiness Studies*, *14*(3), 893-910.

Reid, S. C., Kauer, S. D., Hearps, S. J., Crooke, A. H., Khor, A. S., Sanci, L. A., & Patton, G. C. (2011). A mobile phone application for the assessment and management of youth mental health problems in primary care: a randomised controlled trial. *BMC Family Practice*, *12*(1), 131.

Richardson, K. M., & Rothstein, H. R. (2008). Effects of occupational stress management intervention programs: a meta-analysis. *Journal of Occupational Health Psychology*, 13(1), 69.

Riva, G., Grassi, A., Villani, D., Gaggioli, A., & Preziosa, A. (2007). Managing exam stress using UMTS phones: the advantage of portable audio/video support. *Studies in Health Technology and Informatics*, *125*, 406-408.

Robotham, D. (2008). Stress among higher education students: towards a research agenda. *Higher Education*, 56(6), 735-746.

Schlomer, G. L., Bauman, S., & Card, N. A. (2010). Best practices for missing data management in counseling psychology. *Journal of Counseling Psychology*, *57*(1), 1.

Schneiderman, N., Ironson, G., & Siegel, S. D. (April 2005). Stress and health: Psychological, behavioral, and biological determinants. *Annual Review of Clinical Psychology*, *1*, 607–628.

Spek, V., Cuijpers, P., Nyklícek, I., Riper, H., Keyzer, J., & Pop, V. (2007). Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. *Psychological Medicine*, *37*(03), 319-328.

Stoyanov, S. R., Hides, L., Kavanagh, D. J., Zelenko, O., Tjondronegoro, D., & Mani, M. (2015). Mobile app rating scale: a new tool for assessing the quality of health mobile apps. *JMIR mHealth and uHealth*, *3*(1).

Thompson, E. R. (2007). Development and validation of an internationally reliable short-form of the positive and negative affect schedule (PANAS). *Journal of Cross-cultural Psychology*, *38*(2), 227-242.

Tomlinson, M., Rotheram-Borus, M.J., Swartz, L., & Tsai, A.C. (2013). Scaling Up mHealth: Where Is the Evidence? *PLoS Medicine* 10(2): e1001382.

van Emmerik, A. A., Berings, F., & Lancee, J. (2018). Efficacy of a mindfulness-based mobile application: a randomized waiting-list controlled trial. Mindfulness, 9(1), 187-198.

Villani, D., Grassi, A., Cognetta, C., Cipresso, P., Toniolo, D., & Riva, G. (2012). The effects of a mobile stress management protocol on nurses working with cancer patients: a preliminary controlled study. *Studies in Health Technology and Informatics*, 173,524-528.

Villani, D., Grassi, A., Cognetta, C., Toniolo, D., Cipresso, P., & Riva, G. (2013). Self-help stress management training through mobile phones: An experience with oncology nurses. *Psychological Services*, *10*(3), 315.

Vu, A.H. & Frazier, P. (2016, April). *Access and need related group differences in self-monitoring using a CBT and mindfulness based smartphone application for anxiety and stress management.* Poster session presented at the meeting of the International Society for Research on Internet Interventions, Seattle, WA.

Watkins, D. C., Hunt, J. B., & Eisenberg, D. (2012). Increased demand for mental health services on college campuses: Perspectives from administrators. *Qualitative Social Work*, *11*(3), 319-337.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063.

Zlomke, K. R. (2009). Psychometric properties of internet administered versions of Penn state worry questionnaire (PSWQ) and depression, anxiety, and stress scale (DASS). *Computers in Human Behavior*, *25*(4), 841-843.

Appendix A

Methods following CONSORT guidelines (Eysenbach et al., 2011)

Trial design

a. Description of trial design (such as parallel, factorial) including allocation ratio

b. Important changes to methods after trial commencement (such as eligibility criteria), with reasons. Also, bug fixes, downtimes, content changes.

The trial design was parallel with three groups assigned in an approximately 1:1:1 ratio. There were no changes during the trial. There was no significant reported downtime of the app during the trial period. The following changes occurred initiated by Pacifica Labs, Inc. for its flagship Pacifica mobile app between the two cohorts run in Spring 2016 and Fall 2016.

- Improved onboarding experience (during first cohort onboarding, there were issues on Pacifica's end for onboarding people to download the full app which were corrected w/in one day, if not the same day).
- Ability to change the background theme.
- Retroactive mood and stress ratings
- Home Screen
    - ACTION button on mood ratings home feed (i.e. suggested activities).
    - Pop-up that suggests activities based on current mood and energy level within the app.
- Added ability to set individual reminders for health activities.
- Added 8 more meditations:
    - STRESSFUL SITUATIONS - "Social Situations"
    - STRESSFUL SITUATIONS - "Flying"
    - STRESSFUL SITUATIONS - "Public Speaking"
    - STRESSFUL SITUATIONS - "Public Transit"
    - STRESSFUL SITUATIONS - "Difficult Experience"
    - CALM DOWN - "Calm: Mind"
    - CALM DOWN - "Calm: Breathe"
    - MINDFULNESS - "Mindful: Walk"
- Added "Gratitude" and "Positivity" journals to 'Thoughts' activities.
- Progress/Longitudinal Tracking
    - Introduced "SKILLS" view with levels for the 5 main components of Pacifica.

Participants

a. Eligibility criteria for participants

All participants 18 or older with regular access to a smartphone and no prior experience with the Pacifica app were eligible to participate in this study. Participants received information about this research study via email as well as in-class announcements per the discretion of the Psychology instructor. Study personnel (PI and research assistants) corresponded with participants to complete surveys and deliver a link to download a version of the mobile application, as well as contact information to address questions related to the study, technical issues, on-campus resources (e.g., mental health),

and the institutional review board. Reminder emails were also sent to participants who had not completed the pre and/or post surveys within the allotted time frame. Participants were required to use their university email address to download the app and to complete the online surveys in order to detect/prevent multiple identities. These email addresses were kept separate from the randomized study identification number generated for each student as well as their data. The email addresses were used to accurately allocate extra credit points for completion of the study.

b. Information given during recruitment.

Participants were provided information regarding the study including potential risks and benefits as well as the voluntary nature of their participation. Participants were provided with contact information for the researchers for the study, the institutional review board, as well as mental health services available on campus. All participants who expressed interest in participating in the pilot study were sent a link to the consent form and the pre-test survey.

c. Settings and locations where the data were collected

Outcomes are based on self-report using standard assessment tools. Participants were blinded to their assigned condition and the components in the full vs. active control.

Interventions

The interventions for each group with sufficient details to allow replication, including how and when they were actually administered

a. Mention names, credential, affiliations of the developers, sponsors, and owners

The two primary developers, Dale Beerman and Chris Goettel, have a combined 20 years of experience in adapting consumer-oriented online educational materials. The developers studied the use of CBT for anxiety disorders by examining commonly available self-help books and therapist guides, as well as the online training program "Therapist Training on Cognitive Behavior Therapy for Anxiety Disorders" by the Center for Telepsychology based in Madison, Wisconsin. The site states that its content was developed by leaders in the area of CBT for anxiety-related disorders including Michelle Craske, Ph.D., Raphael Rose, Ph.D., and Kenneth Kobak, Ph.D. The initial selection of the functionality and features of the app was derived from this information and cross-checked in consultation with Christine Moberg, Ph.D., a clinical psychologist at the Palo Alto Veteran Affairs Healthcare System, who serves as advisor and content consultant for Pacifica Labs, Inc. The final iteration of CBT-related features was completed in consultation with an additional clinical psychologist, Ross Nelson, Psy.D.

The authors are independent from the developers of the mobile app and no monetary funding was provided by the developers. The developers provided the mobile app to the participants and the data from the participants to the authors. The developers had no influence on the reporting of the results.

b. Describe the history/development process of the application and previous formative evaluations

After the CBT-related features of the app were established, the Pacifica app completed two rounds of informal pilot testing that focused primarily on usability and debugging of the software prior to public release. The first pilot was administered using family and friends. The second pilot was a public beta with 85 adult U.S. users who self-enrolled via online request for the application. These users were instructed to submit feedback about the ease of use of the self-guided workflow, clarity of app content, and software errors. Pacifica Labs, Inc. subsequently fixed these issues and made minor edits to app instructions and content, which lead to the release of Pacifica version 1.0, the first version publicly available on January 27th, 2015.

c. ***Revisions and updating***.

Version 2.0 of Pacifica Full used for participants in the RCT contained the following mobile app and CBT/ Mindfulness features: push notifications throughout the day at random or a timed schedule per user preference that prompted the user to rate mood and stress; identifying cognitive distortions via free text; the opportunity to set and log daily health goals (e.g., sleep duration, taking medication, time with relationships, caffeine intake); and, finally, guided breathing and visualization exercises. The publicly released version of 2.0 also contained a feature for users to interact with each other in a group message board. Given constraints in our ability to monitor the message board for safety issues that could arise and to offer greater protection to study participants, the message board feature was disabled in both Pacifica Full and Pacifica Lite and version 2.0 to protect the confidentiality of students enrolled in the pilot study. The Pacifica Lite version was adapted from the Full version specifically for this study.

d. Ensure replicability by providing screenshots/screen-capture video

See screenshots in Appendix A.

e. Access

Participants were able to access the application on their phone for the duration of the study for free. They received extra credit for participating in the study.

f. Describe mode of delivery, features/functionalities/components of the intervention and comparator, and the theoretical framework [6] used to design them

Pacifica Full and Pacifica Lite both contained mood ratings, stress ratings, and health behavior monitoring. Mood and stress ratings consist of both quantitative and qualitative data. Quantitative self-reported mood was measured within the app using a single item daily mood measure presented on a dial in which ticks represented a Likert scale of 1 (Awful) to 7 (Great). Qualitative self-reported mood data were measured as word selections underneath the mood dial presented in boxes as well as a free response box. For example, a rating of 1 (Awful) to 3 (Not Good) presents default word selection options such as “panicked,” “stressed,” “angry,” and “sad.” A rating of 4 (Okay) presents default word selection options such as “calm,” “loved” “grateful” in addition to “stressed” and “anxious.” A rating of 5 (Good) to 7 (Great) presents the emotions “relaxed,” “loved,” “grateful,” “excited,” and “happy.” It should be noted that the rating

dial begins with the highest mood rating and decreases as the user navigates the dial clockwise which may introduce instrument bias for higher mood ratings overall (Appendix A, Figure 3). Users are able to review their mood ratings up to 30 days prior to their current day (Appendix A, Figure 4).

A variety of stress management and self-care behaviors can be self-monitored within Pacifica (Appendix A, Figure 5). These include logging usage of potentially harmful behaviors such as daily consumption of caffeine, alcohol, and marijuana (all of which can exacerbate state anxiety) to maintenance behaviors like hours of sleep per night, hygiene (e.g. showering), perceptions of one's eating habits (from poor to perfect on a scale of 1-5), and adherence to prescription medication regimens. Helpful behaviors such as the amount of water consumed per day, exercise activities, time spent outdoors or engaged in a meaningful hobby, or seeing friends or a partner can also be monitored. Each user is able to set a daily goal from the activities the individual chooses to monitor, for which the user may later record a raw score from the scale of measurement provided within the Pacifica App. For example, with caffeine usage, users input the number of cups they have had on that day (without specifying the type of drink or number of milligrams of caffeine) from 0 to 5 or more. For sleep, users can input the number of hours slept the previous night from 2 hours or less to 10 or more. For exercise, hobbies, and other social activities, users can enter the approximate number of minutes from 0 to 90 or more in 15-minute increments. A user is able to specify a set daily goal prior to entering any raw self-report data. The app stores this goal automatically and users are able to change these goals manually as they make progress or shift their focus to other activities. When a user has not reached their set goal, the rating is presented in the color red. Conversely, when a user's self-care goal is met, the rating is presented in the color green. There are no external incentives provided when users meet their goals. Although certain activities lend themselves to beginning or end of day measurement, the current Pacifica workflow does not prompt users to input data before bed or once they wake up (in the case of sleep).

Pacifica Full contained other features and activities with CBT and Mindfulness components:

Goals (Exposures and reappraisal)

Users of Pacifica are encouraged to set goals they are currently having difficulty achieving due to avoidance behaviors (Appendix A, Figure 6). Pacifica prompts users to enter this goal and rate "how difficult it will be" using their in-app version of the Subjective Units of Distress Scale (SUDS) which instead of 0 to 100 is rated on a dial from 1 to 10, clockwise. After you rate the anticipated difficulty, the user selects the "submit a challenge" icon. Users may also identify steps toward this goal to be completed first rather than completing the end goal directly. The goals screen presents a list of goals that have not been completed. The user is presented with text to "tap [the goal] when complete." Afterwards, the user is prompted to complete a new SUDS rating as to how difficult it was to complete, which guides the user to reappraise the anxiety provoking behavior. Completed goals (exposures) are maintained on a separate list on the goals page that can be revealed when selected.

Relax (Mindfulness meditation, progressive muscle relaxation, and breathing)

The relax area provides breathing, guided visualization, muscle relaxation, and pleasing sounds called "soundscapes" (Appendix A, Figure 7). The guided breathing exercises may last from 5 to 25 seconds, depending on the user's preference and can be accompanied by a multitude of sounds (e.g., ocean waves, summer night, and white noise) as well as no sound at all. Progressive muscle relaxation guides the user to focus on tensing and relaxing specific parts of the body in 10 cycles at the user's preferred pace (5 to 25 seconds per cycle). It is important to note that relaxation and meditation techniques in this area require audio for the most part, and so if used in public, would most likely require the user to wear ear buds which may not be acceptable in some settings and roles (e.g., while at work or on public transportation).

Thoughts (Identifying mistakes in thinking, cognitive restructuring)

The thoughts section provides many examples of mistakes in thinking coined by Pacifica as "thinking traps" (e.g., black and white thinking, fortune telling) so that users may readily identify these. Pacifica provides the prompt "Think of a recent situation that led to a strong change in emotion (anxiety, sadness, anger, etc.)" to prompt users to understand the connection between their feelings and thoughts (Appendix A, Figure 8). Prior versions of the app required users to audio record their unhelpful thoughts. In doing so, users could play back their audio and were prompted to identify segments of the playback audio in which they made unhelpful or overly negative and extreme comments by tapping on the screen. Users were then prompted to correct these unhelpful thoughts by recording a new audio file. A recent release of the app allows text input of the thought record instead and users are able to highlight text of distorted thinking and restructure these in the following screen.

Brief description of Pacifica App work-flow

The Pacifica App at the time of the data collection was primarily self-guided. Users could work through any component of the app open to them (based on pay status) in whichever order they preferred. However, mood ratings are on the home screen and they are typically the first activity accessed. For all activities, users are presented with a question mark icon on the upper right of each screen that provides brief psychoeducation flashcards around the activity they are about to begin. This material remains available to users throughout the course of owning Pacifica.

g. ***Describe use parameters***

Participants in both the Pacifica Full and Pacifica Lite group were instructed to use the app daily over the two week period, however all features available could be used ad libitum.

h. ***Clarify the level of human involvement***

Investigators corresponded with participants to troubleshoot issues with the app and to send the Time1 and Time2 Survey links. Apart from the initial onboarding pages

that describe how to use the app according to the conditions randomly assigned, there will be no other prompts, feedback, or human involvement throughout the duration of the study with the exception of participants contacting the research team (via email) if they are experiencing a worsening of symptoms (e.g., suicide ideation, acute panic and anxiety) or wish to leave the study for other reasons. Contact for university mental health services and crisis hotline will be provided in the consent form, on the onboarding site, within the app, and at the bottom of weekly emails.

i. ***Report any prompts/reminders used***

Upon signing up for the Pacifica app, users are required to select their preferred frequency of push notifications before moving forward. This push notification reads "How are you feeling?" from 1 to 3 times per day. Participants had the opportunity to change this option in the settings menu. However, participants also received weekly automated emails that provided visual feedback of the components they used and any self-ratings complete. Participants were able to review this feedback within the app whenever they wished throughout the course of the study. Participants in both Pacifica and Pacifica Lite conditions were guided to use components of the app two ways, via the onboarding website and the weekly emails: A fixed daily schedule designated by the user (morning or evening) as well as before and during experiences of acute distress. Users in both conditions were provided with accurate psychoeducation on how components of the Pacifica Lite and Pacifica full versions may be helpful to manage experiences of stress.

j. ***Describe any co-interventions***

No specific co-interventions were provided by the investigators. As detailed in the demographic description, some individuals already regularly performed mindfulness activities or were undergoing CBT to their understanding, though there were no differences between groups for this and few participants overall were undergoing CBT.

Outcomes

a. Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed. If outcomes were obtained through online questionnaires, describe if they were validated for online use.

Online questionnaires were used at the beginning and end of the study. These were entered in Google Docs by the participants. Data was also collected on Mood, Stress, Health Activities, etc. within the application. Primary outcome measures were DASS21 total, and the subscales of Depression, Stress, and Anxiety. The secondary outcome measures were the PSS, PANAS-SF-10, PROMIS Global Health Measure, FFMQ and AMPS. Of these, the following have been validated for online use: DASS-21 (Zlomke, 2009), PSS (Herrero & Meneses, 2006), PANAS-SF-10 (Howell et al., 2010), and PROMIS Global Health Measure (Hays et al., 2009).

b. Describe whether and how "use" (including intensity of use/dosage) was measured/monitored

Use was measured for each health activity, stress, and mood ratings. Dose was not pre-determined; participants were asked to use the app daily at a minimum and more if they were inclined.

c. Describe whether, how, and when qualitative feedback was obtained from participants

Qualitative feedback was obtained through open ended questions in the survey at T2.

Sample size

a. How sample size was determined. Describe whether and how expected attrition was taken into account when calculating the sample size.

Please see Chapter 3 Methods section.

Randomisation

a Method used to generate the random allocation sequence

The participants were randomized into groups using the Random Integer Generator [0,1,2;Waitlist, Pacifica Lite, Pacifica] from random.org.

Implementation

a. Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions

Randomization and assignment were completed by the lead author.

Blinding

Specify who was blinded, and who wasn't. Informed consent procedures (4a-ii) can create biases and certain expectations - discuss e.g., whether participants knew which intervention was the "intervention of interest" and which one was the "comparator".

Participants were blinded to their randomly assigned group. However, those in the waitlist control group likely knew that they were part of this group. The participants in the active control and intervention groups both received the app and were not informed which group they were in and likely remained blinded through the study as they were not previously exposed to the application or the elements included in the application. Qualitatively, multiple participants in the active control group stated that it would have been useful to have techniques to reduce stress included in the app, not just tracking of stress and mood; this would suggest that they were not aware of these other elements of the app that were available to the Pacifica app intervention group. It should be acknowledged that blinding of participants to treatment group assignment in online health-based trials is more difficult to implement. Study investigators cannot ensure that participants will not discuss or share the application with others, including participants enrolled in different conditions.

Description of the similarity of interventions

The full intervention "Pacifica" app contained in-app monitoring in addition to all CBT and Mindfulness-based components. The active control "Pacifica Lite" app contained

only in-app mood, stress, and health activity monitoring. The waitlist control did not receive an app.

Statistical methods

Statistical methods used to compare groups for primary and secondary outcomes

Imputation techniques to deal with attrition / missing values

Analyses of covariance (ANCOVA) were performed with the randomly assigned group condition and cohort, as well as their interaction as fixed effects, the scale score at T2 as the dependent variables, and the scale score at T1 as the covariate. Listwise deletion was used.

Table 14. Demographics

	Waitlist	Pacifica Lite	Pacifica	Total
Total	127	116	119	362
Age				
18-21	90	93	84	267
22-25	24	19	23	66
26-29	7	2	4	13
30-33	4	1	7	12
34-37			1	1
38-41		1		1
Gender				
Female	84	88	90	262
Male	41	28	27	96
Transgender/Non-Conforming			1	1
Race				
African American or Black	5	3	3	11
European American or White	82	79	74	235
Middle Eastern or Arab American	2	1	1	4
Asian or Asian American	26	22	30	78
Hispanic or Latino(a)	5	3	4	12
American Indian or Alaskan Native	0	1	0	1
Multiracial or Other	7	7	7	21
School Year				
Freshman	16	23	26	65
Sophomore	35	29	34	98
Junior	41	35	27	103
Senior	33	29	29	91
Other	2		3	5

International Student				
Yes	18	13	16	47
No	108	103	103	314
Employment				
Part-Time	72	64	70	206
Full-Time	7	3	6	16
Not Working	46	47	39	132
Basic Needs Met				
Yes	117	111	110	338
No	9	4	9	22
Leisure Needs Met				
Yes	96	81	76	253
No	31	35	43	109
Schedule Med MH				
Yes	90	74	74	238
No	37	41	45	123
MH Services Last 6 Months				
Yes	41	39	38	118
No	86	77	81	244
CBT				
Yes, currently	5	3	5	13
Yes, in the past	16	12	7	35
No	95	96	95	286
Unsure	11	5	9	25
Mindfulness				
Yes, Currently	18	15	17	50
Yes, in the past	34	38	35	107
No	74	63	67	204

Appendix B

Figure 11. Mood score reporting. Main screen for self-report of mood score in the Pacifica App with descriptive words that can be selected and included qualitatively with the mood rating. (Image from Pacifica Inc., 2015)

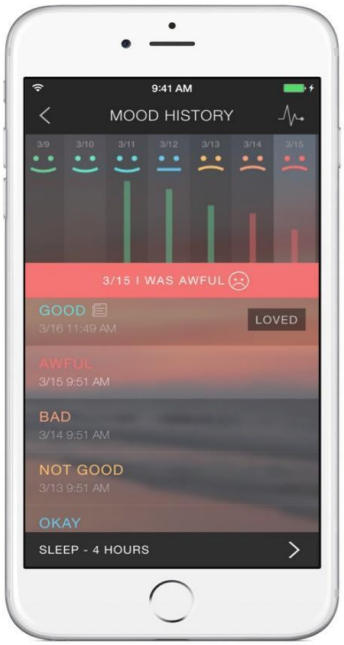

Figure 12. Mood rating timeline. Users can view their mood scores from a pre-defined period of time, up to 30 days prior to the current day. (Image from Pacifica Inc., 2015)

Figure 13. Health behavior monitoring. Users can self-report their activities for the day, including hours of sleep the night before, the number of coffee cups consumed, and the amount of exercise they performed that day. (Image from Pacifica Inc., 2015)

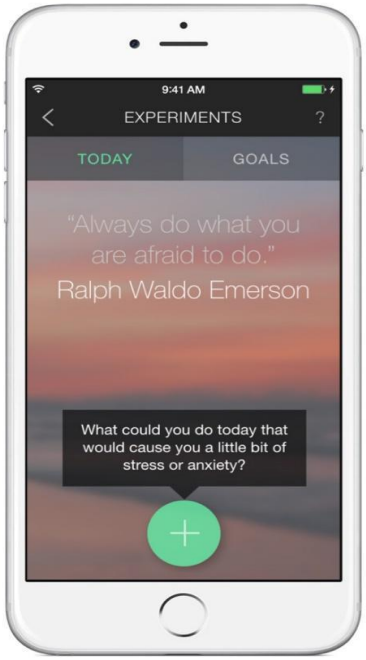

Figure 14. Goals (exposure) exercises. (Image from Pacifica Inc., 2015)

Figure 15. Relaxation exercises. The Pacifica App contains relaxation exercises including muscle relaxation, deep breathing, and meditation. (Image from Pacifica Inc., 2015)

Figure 16. Thought exercises. (Image from Pacifica Inc., 2015)

Appendix C

Table 15. Linear Regression of Pacifica App Activity Usage to DASS Total at T2

Mood Entry	*p*	Standardized β	*t*
Pacifica	.2	-0.09	-1.30
Pacifica Lite	.42	.062	.80
Stress Entry			
Pacifica	.19	-0.09	-1.33
Pacifica Lite	.43	.06	.79
Relax Entry			
Pacifica	.62	-.034	-.50
Psychoeducation Entry			
Pacifica	.77	-0.020	-.29
Pacifica Lite	.85	.015	.19
Thought Entry			
Pacifica	.703	0.03	0.38
Goal Entry			
Pacifica	.426	-.054	-.80