

Study Title: Cognitive Behavioral Immersion: A Randomized Control Trial of Peer-Based Coaching in the Metaverse [NCT Number: NCT06418997]

PI Name: Iony D. Ezawa, Ph.D.

Study Procedures

1. Background/Rationale

- a. Major depressive disorder is the single most common mental illness, with an estimated 17.3 million adults in the United States having had at least one major depressive episode in 2017 (i.e., 6.7% of the U.S. population) (WHO, 2017). Approximately 50% of those experiencing depression also experience comorbid anxiety (Hirschfield, 2001). Symptoms of depression and anxiety contribute to considerable detriments in wellbeing, health, and functioning, and rank among the leading public health burdens worldwide (Whiteford et al., 2013). Concerningly, the prevalence and severity of these disorders are increasing, with a 25% increase in their prevalence since 2019 (WHO, 2010). While there are effective treatments for each, only a tiny fraction of those in distress have access to treatment. There is a critical need for effective, affordable, and scalable interventions to manage depressive and anxiety disorders.

Cognitive behavioral therapy (CBT) is one of the most effective tools for the treatment of depression and anxiety. CBT is based on a model of how one's thoughts, feelings/emotions, behaviors, and physiology interact in one's environment. According to the model, although it can be difficult to directly change emotion, a person can do so indirectly by changing maladaptive cognitions, behaviors, or physiological states. In this approach, therapists teach clients a host of cognitive and behavioral tools to identify, evaluate, and correct inaccurate negative thoughts, and increase behaviors that build efficacy and agency and boost positive affect. CBT for depression and anxiety has been rigorously tested in clinical trials and has been shown to be both efficacious (better than its absence) and specific (superior to nonspecific controls) by outperforming waitlist and nonspecific control conditions and performing on par with antidepressant medications, the current standard of treatment (DeRubeis et al., 2005). CBT also has been shown to have a lasting effect not found for medications that persists beyond the end of treatment (Hollon et al., 2005). The beneficial effects of CBT are driven by patients' use of cognitive behavioral skills (Strunk et al., 2007), which has been shown to lead to cognitive and symptom change in depression and anxiety (Schmidt et al., 2019). Despite its many strengths, CBT-oriented clinical care is often very expensive and difficult to access. Further, because of the stigma associated with psychological conditions, many individuals do not seek help.

To address problems of accessibility, the field is increasingly utilizing lay therapists (also commonly referred to as coaches or peers) to provide cognitive

behavioral coaching. Investigations of lay therapists provide empirical support that individuals without any prior formal therapy training can be trained to deliver effective cognitive behavioral interventions (e.g., Singla et al., 2014). This is exciting because it offers a way to reach millions of affected individuals who might otherwise not receive any care.

There is also accumulating evidence indicating that technology can be leveraged to deliver affordable, accessible, and scalable treatments with outcomes comparable to traditional face-to-face psychotherapy treatments (Karyotaki et al., 2021). While some digital treatments have tried to eliminate the need for therapists altogether by creating self-guided computerized cognitive behavioral programs, findings from a recent meta-analysis suggest that the efficacy and engagement in these digital programs are improved when provided with coach-support via email or phone communication (Karyotaki et al., 2021). These interpersonal relationships can facilitate many of the nonspecific effects that account for the bulk of the outcomes in all psychotherapeutic modalities (i.e., common factors) (Cuijpers et al., 2012). Most forms of digital CBT come in the form of self-guided computerized programs that can be accessed via a flat screen device (e.g., computer or mobile phone). However, internet-connected 3-D virtual environments have also been developed that allow end-users to interact as avatar (i.e., ‘the metaverse’). There are many disparate 3-D networked environments, many of which are games, that make up the metaverse. These include environments such as Roblox, Fortnite, Minecraft, VRChat and Meta’s (formerly Facebook) Horizon Worlds. These environments are primarily accessed via flat screen devices, and some can also be experienced through immersive virtual reality (VR) technology.

Immersive VR extends the metaverse by enabling people to be immersed into a three dimensional (3-D) metaverse environment. Immersive VR is powerful because it facilitates a sense of presence – the subjective experience of being in a place or environment when one is physically situated in another (Witmer & Singer, 1998). Immersive VR utilizes a head-mounted display to update graphics in real time as a person moves his or her head to look around an environment. Controllers can be used to manipulate the 3-D position of the person such that the person has a spatial awareness of the virtual environment. Although flat screen devices such as computers or smart phones also allow a person to navigate a 3-D environment (like how one might play a traditional videogame), the experience is not as evocative. Research suggests that immersive virtual stimuli experienced in immersive VR, but not via a flat screen, can elicit affective reactions much like real-life stimuli, as evidenced by immersive VR exposure paradigms showing similar effect sizes to in vivo exposure (Wechsler et al., 2019).

Due to the immersive nature of VR, this modality has particular utility for psychological interventions that involve interpersonal interaction. The immersive VR modality for social interaction is unique as a technological communication modality insofar as it creates a great sense of presence (more

than text chat, voice chat, or video chat) while also maintaining a complete sense of anonymity (more than an in-person or video chat meeting). With this sense of anonymity, VR participants may feel safer to try new skills, be open with strangers, and take risks such as disclosing vulnerabilities that may be difficult to do in the absence of anonymity. In addition, being transported into different immersive worlds allows participants to not only target typical cognitive behavioral skills (i.e., changing thoughts, behaviors, and physiology), but they can also change their environment allowing the opportunity to change the context in which their thoughts, feelings, behaviors, and physiology occur to facilitate cognitive behavioral skills acquisition. While there is evidence supporting the use of VR in mental health treatment (almost exclusively in the treatment of PTSD and phobias), these prior VR interventions have not been networked, and therefore do not allow for the occurrence of social and non-specific effects such as individuals interacting with each other. The use of social VR interventions is relatively unexplored, yet is uniquely positioned to deliver an immersive experience that can create human connection in a way that no other technological medium can. VR has great therapeutic utility since it allows participants to experience an intervention in a safe, controlled environment that is available anytime and anywhere, while facilitating easy outcome assessment. Social VR may be a distinctively powerful medium through which to deliver engaging, affordable, accessible, and scalable mental health interventions.

Innerworld, Inc. (formerly named Very Real Help) developed its namesake application Innerworld (formerly named Help Club) during a Phase I SBIR project. Innerworld is a clinical research platform that provides Cognitive Behavioral Immersion (CBI)[™], a novel synchronous cognitive behavioral skills training program delivered in a group setting (to provide interpersonal support) by coaches in a virtual world. Although CBI incorporates skills taught in CBT, it differs markedly from CBT because CBI immerses a person in a virtual environment to change the context of their thoughts, feelings, behaviors, and physiology to facilitate cognitive behavioral skill acquisition. CBI was developed to combine the strengths of recent psychological and technological advances to potentially deliver efficacious, affordable, accessible, and scalable mental health support.

While there is initial evidence supporting the feasibility of CBI for depression and anxiety, a direct comparison of this modality against a control has yet to be conducted, and its absolute efficacy is unknown. In addition, it is unknown whether accessing CBI via immersive VR confers any advantage over and above accessing CBI via less immersive flat screen devices (like watching a regular television).

2. Purpose/Objectives/Aims/Research Questions

- a. This study is designed to address the main following research questions: (1) Is CBI efficacious? And (2) Does immersive VR confer any advantage over and

above accessing CBI via less immersive flat screen devices? In line with these questions, this study aims to accomplish the following objectives:

- i. Compare Cognitive Behavioral Immersion (CBI) accessed via immersive VR (CBI-VR) to a delayed access control (DAC) condition and CBI accessed via less immersive flat screen devices (CBI-flat screen) in a randomized control trial (RCT).
- ii. Explore baseline indices that predict general (prognostic) program response.
- iii. Explore baseline indices that predict differential (prescriptive) program response (moderators).
- iv. Explore and contrast mechanisms of clinical and functional change across conditions (mediators).

Hypotheses: We predict that participants randomized to CBI-VR will experience greater clinical and functional improvements than participants randomized to the other two conditions. We also predict that cognitive behavioral skill use, group alliance, and program usage will predict general program response, whereas level of immersion in CBI will moderate program response. Finally, we predict that individuals' experience of cognitive change and online social support will mediate program response.

3. Participants (sample)

- a. Inclusion criteria: In order to be eligible to participate in this study, an individual must meet all of the following criteria: (1) Score ≥ 10 on the PHQ-9, (2) 18 years of age or older, (3) Able and willing to give informed consent, (4) Has a Mac or PC with an internet connection that can run Innerworld, (5) Able to speak and read English.

Exclusion criteria: An individual who meets any of the following criteria will be excluded from participation in this study: (1) Individuals showcasing sufficiently elevated risk of suicide or harm at the baseline assessment. A document containing more information and a comprehensive standard operating procedure (SOP) for managing participant suicidality is included in this application under the title "SOP for Suicidality". If, at baseline, a participant is found to be at imminent risk for suicide or harm (#4 or #5 on the research Columbia Suicide Severity Rating Scale [C-SSRS] conducted during baseline), they will not be randomized, and the study team will refer them for appropriate assessment and care to ensure their well-being and safety. If a participant scores a 5 on the research C-SSRS at baseline, additional steps will be taken which are detailed in the SOP for Suicidality.

- b. We will not be including individuals under the age of 18 in this study. We are specifically interested in the efficacy of CBI for adults. Furthermore, many of the currently available metaverse mental health interventions are only legally available to persons above the age of 18.

- c. Since we are collecting a sample from the general population, participants from certain special subject populations may choose to enter the study if they meet the general inclusion criteria. These participants may include employees, students, pregnant individuals, military personnel, and/or indigenous individuals. Adults not competent to consent or minors are excluded since they cannot consent to the study and non-English-speaking populations are excluded due to study materials being English-language only.

4. Recruitment/Screening Process (sampling strategy)

- a. Recruitment will include online advertising (e.g., via our online website, social media, listservs, and email), physical distribution of flyers on and around the USC campus, and word of mouth.
- b. Online recruitment will be done through various methods. One method will be posting a brief description of the study with a link to study details and screener as a post or advertisement to social media platforms such as Reddit, Twitter, and Facebook. Another online method will be sending emails containing a brief description of the study with a link to study details/screener to colleagues of the principal investigator and relevant psychology listservs. Recruitment flyers will be posted in the discussion boards in the Seeley G. Mudd building. They will also be posted on walls and distributed by hand by research assistants to students and other individuals on and near USC campus. Another online method will be posting a brief description of the study with a link to study details and screener to a dedicated “VR RCT” section on our lab website. We will also recruit for this survey via word of mouth. Since the inclusion criteria for this study is broad, we aim to recruit broadly.
- c. Materials used for recruitment will consist of flyers, social media posts, and emails.
- d. Interested individuals will be directed to the study page of our lab website. On this website, they will complete an initial prescreening online questionnaire containing a brief description of the study, and questions assessing study eligibility (including the Patient Health Questionnaire). At the prescreen stage, those who do not meet study inclusion criteria will be provided an automatic message letting them know that they are not eligible for the study. They will also be provided with links to learn more about the Innerworld application and CBI. Those who do meet initial inclusion criteria on the pre-screening link will be emailed a link to schedule a brief intake session via video or phone with study personnel. We will not retain screening data.

5. Methods

- a. Participants currently experiencing symptoms of depression will be recruited via online advertisements, electronic communications, printed flyers, and word of mouth. Interested individuals will be directed to the study website, where they will complete an initial prescreening online questionnaire containing a brief description of the study, eligibility questions, and the Patient Health

Questionnaire (PHQ-9). Those who meet initial eligibility criteria on the pre-screening (i.e., aged 18 or older, PHQ-9 ≥ 10 , have a computer with an internet connection, and speak and read English) will be emailed a link to schedule a brief (~15 minute) intake session conducted via video or phone call with study personnel. Prior to attending their scheduled meeting, each subject will also be asked to complete a set of intake measures via a secure online REDCap link. During the intake, study personnel will explain the study, answer any questions, reassess participant eligibility, and obtain informed consent via the REDCap eConsent module from eligible participants. Participants also will be given a copy of the consent form and study personnel contact information in the case they have any additional questions about the study. Upon providing consent, participants will be randomized to one of three conditions: (1) CBI-VR, (2) CBI-flat screen, or (3) DAC. We plan to randomly assign equal numbers of participants in a 1:1:1 ratio to the three study conditions (without any stratification).

Participants randomized to the CBI-VR or CBI-flat screen conditions will be sent an email alerting them of their study condition and asked to select an available CBI group that they can attend (including start date/time of group they can attend on a weekly basis). Once they select a group start date and time, they will be emailed confirmation of their group start date and directions on how to access the Innerworld application (which is the application that will provide CBI). These conditions will consist of eight weekly, 60-minute cognitive behavioral skills training groups led by trained non-professional coaches in Innerworld. Prior to their first group session, participants will be asked to participate in a short tutorial hosted by Innerworld about how to navigate Innerworld on their internet-connected device and how to design their avatar. Innerworld will collect passive participant information (number of logins, time logged in, group attendance, type of device used), which will later be shared with the USC study personnel. We will use REDCap to send weekly emails with survey links for participants to complete during each week in their randomized condition. The emails will also include reminders of upcoming sessions. After participants have completed the eight-week group, they will be sent a link to complete post-condition measures. The only difference between these CBI-VR and CBI-flat screen conditions is the device from which participants will attend Innerworld meetings. Those in the CBI-VR condition will be asked to sign into Innerworld via a VR headset. To facilitate this request, participants randomized to the CBI-VR condition will be asked to provide a mailing address to which a VR headset can be sent, and a VR headset will be mailed to the participant prior to their group start date. If they have any issues setting up the device or logging into Innerworld, participants will be directed to reach out to study staff. If the devices get damaged, participants will also be directed to reach out to study staff to help fix or replace the device. The headset we will be using for this study is the Oculus Quest. Participants in the CBI-flat screen condition will be asked to join Innerworld meetings via a flat screen device (i.e., computer). Those in the CBI-

flat screen condition will be sent a VR headset as compensation after the eight week program. Participants will be informed of this during study consent.

Participants randomized to DAC will be sent an email alerting them of their study condition. This condition will last eight weeks. During this condition, participants will be asked not to attend a CBI group for an 8-week period and no invitation will be sent. However, we will not prohibit participants from pursuing other treatment (regardless of condition) and we will monitor any additional treatment received. Those in the DAC condition will be told they will be sent a VR headset as compensation after the 8 weeks. Participants will then have the option to participate in a non-study immersive CBI program after completing the DAC condition if they are still interested. REDCap will send automated emails with survey links for participants to complete during each of the eight weeks of the DAC condition (see measures section below for a list and schedule of assessments). On the ninth week, after completion of the DAC condition, participants will be sent a link to complete the post-condition measures (see measures section below for the list and schedule of assessment that will be distributed at this time point). After completion of the eight-week program and post-condition assessment, participants will be invited to keep attending Innerworld (or begin attending Innerworld if in the DAC condition). Participants in the CBI-VR condition will keep their VR headsets, and participants in the CBI-flat screen and DAC conditions will be provided with a VR headset. Participants will be asked to provide a mailing address to which a VR headset can be sent, and it will be sent to them immediately.

Following the eight week randomization, all participants from each condition will be asked to participate in six monthly follow-up surveys to monitor their continued progress and Innerworld usage (see measures section below for a list and schedule of assessments to be distributed at the monthly follow ups). Consequently, total study duration including the randomization and follow up will be eight months.

Intervention. CBI will be provided by Innerworld. Innerworld is a freely available application, and it provides CBI outside of the research context. Innerworld will not be involved in the research in any way beyond providing their standard content. Innerworld has a team of engineers to maintain the technological aspects of the program and a 24/7 technical support to help individuals who experience challenges with the platform. Innerworld trains lay coaches (adults with no prior professional training) to serve as CBI coaches. Each coach undergoes an intensive 20-hour training program led by Innerworld support staff to learn how to facilitate the eight-week cognitive behavioral training program. This training program was developed by investigators expert in CBT. See the table below for a condensed outline of the coach training program.

#	Training Content	Training Methods	Homework	Hours
---	------------------	------------------	----------	-------

1	Introductions; therapy vs peer support; CBI model; basic therapeutic skills	Group based discussion; roleplays	Shadow 2 CBI sessions; read CBI manual	4
2	Review HW; how to lead a meeting; overview of CBI skills	Didactic; demonstrations; role-plays; Q&A	Shadow 2 CBI sessions; practice skills	4
3	Review HW; how to teach skills; overview of additional CBI skills	Demonstrations; role-plays; Q&A	Shadow 2 CBI sessions; practice skills	4
4	Review HW; dealing with trolls; moderation tools; crisis situations; administer surveys	Live demonstrations; Q&A; role-plays	Shadow 2 CBI sessions; review crisis protocol	4
5	Review HW; practice leading a session; discuss next steps	Q&A; role-plays	Co-lead CBI sessions and get evaluated	4

Following completion of the training program, each coach must co-facilitate 5 full CBI sessions in the presence of a trainer. The coach will be provided feedback following each session. If the coach is meeting competency standards according to the trainer's assessment following the 5th meeting, the coach will be ready to begin leading their own meetings. If the coach is deemed not yet ready, the coach will need to continue this supervision and evaluation process until deemed ready.

In the eight-week CBI-VR and CBI-flat screen program, coaches will follow a detailed manual describing the content/practices to be followed for each session (see table below for a condensed outline of session topics). Participants will meet in groups in which they can learn lessons on various "thinking" (cognitive) and "doing" (behavioral) strategies. Cognitive strategies include but are not limited to identifying the connections between thoughts, feelings, physiology, and behaviors and identifying and examining negative automatic thoughts. Behavioral strategies include but are not limited to activity scheduling, problem solving, and assertion training. Each session begins with setting the agenda and reviewing homework from previous session. Each session ends with a session summary, homework assignment, and feedback about the session. Psych-ed for any tools or concepts involves didactics and prompting the group for examples to use in explanations. As is standard practice in conventional CBT, session agendas will follow a standard format, but coaches will have the flexibility to adjust content to meet participants' needs. Groups will consist of a coach and up to 10 participants. Participants will be allowed to attend other

sessions, but will be strongly encouraged to attend their assigned weekly session.

#	Topic & Session Focus	Homework
1	Introduction to CBI model: Icebreakers, practice with CBI conceptualization	CBI model
2	Behavioral Activation (BA) & Mood Tracking: Relation between behaviors & mood, practice mood tracking	Mood tracking
3	BA & Mood Tracking: Psych-ed on BA, activity scheduling, problem solving	Activity schedule
4	Identifying Automatic Thoughts (ATs): Psych-ed on ATs & relation to mood, group practice with 3-column thought record (TR)	3-column TR
5	Reframing Automatic Thoughts: Psych-ed on distortions & reframe questions, individual practice with 4-column TR, review as group	4-column TR
6	Motivation & Performance: Psych-ed on Yerkes-Dodson, introduction of SMART goals, individual practice creating goals, discussion on problem solving around goals	SMART goals
7	Assertion Training: Psych-ed on assertiveness curve tool, group role-plays	Assertion practice
8	Relapse Prevention Plan & Next Steps: Individual practice completing relapse prevention plan, feedback on course, check-outs to explore key takeaways	

Compensation. Participants will also be incentivized by receiving \$10 for each of the eight weekly survey sets submitted during the initial randomized condition, \$10 for the post-study survey, and \$10 for each of the six-monthly follow-up surveys submitted (for a total of \$150 if all surveys sets are submitted). Payment will be paid via electronic gift cards after each submitted survey set. All participants will be compensated with a VR headset (the timing of the headset distribution will depend on study condition; those randomized to CBI-VR will receive the headset prior to study participation whereas those in the other two conditions will receive the headset after the eight-week acute condition to avoid contaminating the results).

Confidentiality and data storage. Individuals who provide consent and are subsequently enrolled in the study will be assigned password-protected ID numbers. The research material obtained from the study subjects will include demographic data and self-report surveys. All survey information will be recorded via the secure online database, REDCap. Innerworld, Inc. will also collect passive participant information (e.g., number of program logins, time logged in, group attendance, type of device participant joined on). Innerworld, Inc.'s data will be stored under confidential usernames that participants create. These data will then be passed to USC study personnel, via a data sharing agreement through USC's Steven's Center, to integrate with data collected via REDCap. USC study personnel will ask participants directly what their usernames

are after their first group session. We will then be able to connect the data passed from Innerworld to USC via the usernames. All information will be collected, handled, and stored securely. After study completion, all data will be de-identified and stored electronically in a password-controlled database at USC or at Innerworld, Inc. Only those researchers at either USC or Innerworld, Inc. working on data analysis or reporting will have access to the data.

Data analysis. After we analyze the study data (see Data Analysis section below for full data analysis plan), we plan to disseminate our findings by writing research reports to be published in a peer-reviewed scientific journals. We also plan to present our findings at national scientific conferences. However, this will only be done after the data has been statistically analyzed and our results have been investigated and contextualized within the body of literature.

Suicidality prevention. It is possible that at baseline or during the course of the trial, participants could admit sufficiently severe suicidal ideation to require further action from investigators. We will be assessing suicidality via a self-report item on the PHQ-9 throughout the study (i.e., once during prescreen, once during intake, weekly during the 8 weeks of the condition, once during the post-study survey in the 9th week, and monthly during the 6-month follow up period). An automatic threshold will be established on REDCap for PHQ-9 assessments to manage reports of suicidality or self-harm. If a participant or potential participant meets the criteria for suicidality (as indicated by scoring a 2 [“More than half the days”] or a 3 [“Nearly every day”] on the suicide item of the PHQ-9), a series of actions will be initiated as described in the SOP for Suicidality.

6. Instrumentation

- a. **Questionnaires/Survey Measures (names and citations)**
 - i. Screening Measures (not collected for research data.)
 - A. Prescreen Measure. This is a measure that assesses study eligibility criteria and collects basic contact information. Screening data will not be retained.
 - ii. Outcomes.
 - A. Patient Health Questionnaire (PHQ-9). The PHQ-9 (Kroenke et al., 2001) is a 9-item self-report measure with strong psychometric properties that mirrors the nine criteria for ascertaining whether someone meets criteria for major depressive disorder (MDD) in the Diagnostic and Statistical Manual Fourth Edition (DSM-IV). Scores range from 0 to 27, with higher scores indicating great symptom severity.
 - B. Generalized Anxiety Disorder Questionnaire (GAD-7). The GAD-7 (Spitzer et al., 2006) is a 7-item self-report measure with strong psychometric properties based on the diagnostic criteria for Generalized Anxiety Disorder (GAD-7) from the DSM-IV. Scores

range from 0 to 21, with higher scores indicating great symptom severity.

- C. WHO Quality of Life-Brief (QOL). The WHO QOL (Whoqol Group, 1998) is a 26-item self-report cross-cultural measure of quality of life within the domains of physical health, psychological health, social relationships, and environment. During the follow-up period, we will transition to using only the first item of the measure assessing overall quality of life. Scores are transformed on a scale from 0 to 100, with higher scores indicating a higher quality of life. This measure has strong psychometric properties.

iii. Predictors

- A. Competencies of Cognitive Therapy Scale (CCTS). This 29-item self-report scale measures three areas of cognitive behavioral skill competency: behavioral activation, automatic thoughts, and core beliefs (Strunk et al., 2007). Items are measures on 1-to-7-point scales, with higher scores indicating greater skill use. This measure has strong psychometric properties.
- B. Group Session Rating Scale (GSRS). The GSRS (Quirk et al., 2013) is a 4-item measure that assesses group-therapy alliance and the quality of individual sessions and helper skill. This measure is administered post-session. Total scores can range from 4 to 40 with higher scores indicating greater levels of the group alliance. It has demonstrated acceptable psychometric properties.
- C. Program Engagement (ENG). Innerworld will collect passive user information regarding engagement with the platform (e.g., log ins, sessions attended, minutes logged in, group attendance, type of device).

iv. Moderators

- A. Demographics Questionnaire (DEMO). This is a basic demographic information survey that includes age, sex, gender identity, sexual orientation, race, ethnicity, educational attainment, employment/occupation, technology experience, past therapy experience, prior diagnosis, among others.
- B. Immersion Questionnaire (IMMQ). This questionnaire adapts a three-item measure of social presence from Bailenson et al. (2005) and a five-item measure of environmental presence from Fox et al. (2009) to assess the degree to which the participant feels immersed in the metaverse.

v. Mediators

- A. Cognitive Change Immediate Scale (CCI). The CCI (Schmidt et al., 2019) is a five-item measure administered immediately post-session designed to assess participants' experience of cognitive change and use of cognitive skills during sessions. Items are measured on a 7-point Likert scale, where lower scores indicate

less agreement. This measure has demonstrated acceptable psychometric properties.

- B. Online Social Support Scale (OSSS). The OSSS (Nick et al., 2018) is a 40-item scale with strong psychometric properties that assesses the level of online social support participants received from the virtual sessions. We plan to use a 30-item version of the OSSS, measuring constructs of esteem/ emotional support, social companionship, and informational support. Each item is scored from 1 to 5, with higher scores indicating a greater sense of online social support.

vi. Harm Management

- A. Columbia Suicide Severity Rating Scale (C-SSRS). The C-SSRS (Posner et al. 2007) is a validated, structured interview measure that seeks to assess four constructs found to be predictive of suicide and suicide attempts: severity of ideation (2-5 items), intensity of ideation (5 items), behavior (4 items), and lethality (1-2 items). We are utilizing the self-report version of this measure as well as the interview version designed for research settings.

Participants in all three conditions will receive the same set of surveys at the same time points apart from the two assessments (GSRS, CCS). These two session assessments will not be given to the DAC participants because they would have not yet completed a session. The assessment schedule is:

Measures		Prescreen	Intake	Wk 1	Wk 2	Wk 3	Wk 4	Wk 5	Wk 6	Wk 7	Wk 8	Post	Follow-up
Screen	Pre screen	X											
Outcomes	PHQ-9	X	X	X	X	X	X	X	X	X	X	X	X
	GAD-7		X	X	X	X	X	X	X	X	X	X	X
	QOL		X									X	X
Predictors	CCTS		X									X	
	GSRS			X	X	X	X	X	X	X	X		
	ENG			<i>X (continuous passive data collection)</i>									
Moderators	DEMO		X										X ^a
	IMMQ						X					X	
Mediators	CCI			X	X	X	X	X	X	X	X		
	OSSS		X				X					X	
Harm Management	C-SSRS	X ^b	X ^b	X ^b	X ^b	X ^b	X ^b	X ^b	X ^b	X ^b	X ^b	X ^b	X ^b

^aParticipants will only be asked two optional questions pertaining to changes in psychiatric medication or other psychological treatment during the follow-up period.

^bOnly participants that report sufficiently high incidence of thoughts of suicide or self-harm on the PHQ-9 will be shown the Self-Report C-SSRS. Only participants that score in the “imminent risk” range of the Self-Report C-SSRS (a 4 or 5) will then be asked to complete the research C-SSRS with a study team member.

b. Data Analysis

- i. Data analysis will primarily use R Statistical Software (R Core Team 2021). Theory-driven procedures (i.e., inferential statistics) will use two-tailed tests, $\alpha < 0.05$, and power $> 80\%$. When applicable, families of multiple significance tests will control for family-wise false discovery rate using resampling. Descriptive statistics, including histograms, will be computed for each study variable, along with 95% bootstrap confidence intervals. Appropriate data transformations will be made as needed. We will use nonparametric statistics or generalized models with a link function to accommodate non-normal distributions.

Because the primary aim concerns trajectories of symptoms over the 8-week intervention period, the main analytic approach will be a longitudinal random coefficients slope-as-outcome hierarchical linear model (HLM). HLM is required for exploring the longitudinal nature of the data because repeated measurement of participants violates ANOVA's assumption of independent errors. To control for correlated errors within participants, the model estimates an individual growth curve for each participant. Advantages of this method over classical approaches include: (a) repeated measurements increase statistical power; (b) the shape of change over time is described; (c) missing values and unequal time intervals between measurements for different participants can be handled; and (d) psychometric problems with pre-post change scores are avoided.

This study requires a 2-level HLM to represent multiple time points for each individual. Level 1, within participant, represents each participant's time-varying symptoms as an individual linear trajectory. Level 2 uses individual fixed characteristics (e.g., sex, condition arm) to predict individual growth parameters (i.e., participant's intercept and slope). In this regression framework, we can then add covariates to ask more nuanced questions, observe interaction effects, and increase power by reducing error variance. Identification of significant moderators of outcome informs potential subgroups of individuals who may respond particularly well to CBI-VR compared to CBI-flat screen or TAU. We expect to estimate distinct linear slopes during the 8-week period (from baseline

to post-condition) and during the 6-month follow-up period (from post-condition to 6-month follow-up). Therefore, we plan to specify time in a piecewise manner to allow for separate slope estimates for the active condition and follow-up periods. A strength of this method is the ability to assess “acute” and “long-term” condition effects, respectively, in the same regression model.

Under the HLM framework, we will examine whether CBI-VR produces better clinical and functional outcomes compared to TAU and CBI-flat screen in terms of differences in trajectories (i.e., slopes) and outcomes at each timepoint (i.e., least-squares means and pairwise contrasts). Separate regression models will be used for each of the primary outcome variables (depression, anxiety, and quality of life). To examine differences in slopes, the primary effect of interest in the HLM model will be the Condition (CBI-VR, CBI-flat screen, TAU) x Time (active condition, follow-up) interaction effect. Using a piecewise specification of time, the models will yield estimates for the Condition x Time interaction effect during the active condition and follow-up periods, corresponding to detecting acute and long-term differences in condition response (i.e., slopes). We will also examine absolute change in outcome variables and differences in outcome variables by condition at each timepoint through estimation of marginal means or least-squares means and pairwise contrasts at each timepoint. This analysis will inform timepoints during the active condition and follow-up periods that outcomes of interest start to significantly differ.

To explore baseline indices that predict general (prognostic) program response, we will enter proposed prognostic variables as predictors (i.e., covariates) in the HLM models using a stepwise approach. To mitigate concerns of multicollinearity, we will first perform correlation analyses between prognostic variables. As correlations among predictors of 0.7 or higher have demonstrated cause for concern, predictors with high correlation may be combined using dimensionality reduction. If needed, we may also embed a sequential search or penalized regression algorithm to facilitate variable selection of prognostic variables.

To explore baseline indices that predict differential (prescriptive) program response, hypothesized prescriptive variables will be entered as potential moderators of program response in the HLM models. Specifically, separate Prescriptive Variable x Time interaction effects will be assessed in each HLM model. Again, the piecewise specification of time will allow for the prescriptive variables to be tested and identified during the active condition and follow-up periods.

To identify mechanisms of change of clinical/functional change across conditions, we will test for mediation of proposed mechanisms in the longitudinal framework by specifying hypothesized mediators as time-varying random effects. Estimation and inference for this time-

varying mediation model allow mediation effects to vary as a function of time. This is suitable for detecting processes by which the intervention affects outcomes as proposed mediators will be collected at multiple timepoints.

A longitudinal power analysis was conducted using the 'powerlmm' package in R with $\alpha < 0.05$ two-tailed, power $> 80\%$. Power analyses were focused on the ability to detect a medium effect size (Cohen's d of 0.50) at post-condition in the context of a 2-level random coefficients HLM. Given that the primary contrast of interest would be that between CBI and the other study conditions, calculations focused on detecting significant differences between two active conditions. Where study parameters in this 2-level model were not yet known, we used standard values and also varied parameters to approximate an effective sample size. We specified 8 equally spaced timepoints, amount of baseline variance at the subject level at 0.50, and allowed the ratio of the random slope variance to the within-subject variance to vary from 0.02 to 0.05 in increments of 0.01. Based on typical rates of dropout in digital CBT treatment, we set a dropout rate of 25%. With anticipated missing data, the sample size estimate in each study arm varied from $n=65$ to $n=138$, yielding an average $n=102$. It is also recommended that power calculations should be based on the minimum detectable effect size (MDES), or smallest difference between conditions that one would be willing to fail to detect. We assessed this using the 'PowerUpR' package (v1.1.0). Using the parameters specified above, we estimated $MDES=0.26$, 95% confidence interval= $[0.08, 0.44]$ with $n=102$ in each study arm and $MDES=0.22$, 95% confidence interval= $[0.07, 0.37]$ with $n=138$ in each study arm. Considering our pilot data, we anticipate differences between conditions that will be considerably larger. Thus, a total sample size of $N=306$ ($n=102$ in each study arm) will be effective for detecting meaningful differences in outcomes between CBI and the other conditions.

7. References

- Bailenson, J.N., Swinth, K., Hoyt, C., Persky, S., Dimov, A., Blascovich, J. (2005) The independent and interactive effects of embodied agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments. *Presence Teleoper Virtual Environ*, 14(4):379–393. <https://doi.org/10.1162/105474605774785235>
- Beck Institute for Cognitive Behavior Therapy. (2018). Cognitive Therapy Scale rating manual. Retrieved from <https://www.beckinstitute.org/wp-content/uploads/2015/10/CTRS-current-10-2011-Cognitive-Therapy-Rating-Scale-2.pdf>.
- Cuijpers, P., Driessen, E., Hollon, S. D., van Oppen, P., Barth, J., & Andersson, G. (2012). The efficacy of non-directive supportive therapy for adult depression: A meta-

- analysis. *Clinical Psychology Review*, 32(4), 280-291. <http://doi:10.1016/j.cpr.2012.01.003>
- Depression and Other Common Mental Disorders: Global Health Estimates. Geneva: World Health Organization; 2017. Licence: CC BY-NC-SA 3.0 IGO.
- DeRubeis, R. J., Hollon, S. D., Amsterdam, J. D., Shelton, R. C., Young, P. R., Salomon, R. M., ... & Gallop, R. (2005). Cognitive therapy vs medications in the treatment of moderate to severe depression. *Archives of General Psychiatry*, 62(4), 409-416. doi:10.1001/archpsyc.62.4.409
- Fox, J., Bailenson, J., Binney, J. (2009) Virtual experiences, physical behaviors: the effect of presence on imitation of an eating avatar. *Presence Teleop Virt Environ*, 18:294–303. <https://doi.org/10.1162/pres.18.4.294>
- Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). Applied statistics for the behavioral sciences (Vol. 663). Houghton Mifflin College Division.
- Hirschfeld, R. M. (2001). The comorbidity of major depression and anxiety disorders: recognition and management in primary care. *The Primary Care Companion for CNS Disorders*, 3, 244–254.
- Hollon, S. D., DeRubeis, R. J., Shelton, R. C., Amsterdam, J. D., Salomon, R. M., O'Reardon, J. P., ... & Gallop, R. (2005). Prevention of relapse following cognitive therapy vs medications in moderate to severe depression. *Archives of General Psychiatry*, 62(4), 417-422. doi:10.1001/archpsyc.62.4.417
- Karyotaki, E., Efthimiou, O., Miguel, C., genannt Bermpohl, F. M., Furukawa, T. A., Cuijpers, P., ... & Forsell, Y. (2021). Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. *JAMA psychiatry*, 78(4), 361-371. doi:10.1001/jamapsychiatry.2020.4364
- Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16(9), 606-613.
- Nick EA, Cole DA, Cho SJ, Smith DK, Carter TG, Zerkowicz RL (2018). The Online Social Support Scale: Measure development and validation. *Psychol Assess*. ;30(9):1127-1143. doi:10.1037/pas0000558
- Posner, K. (2007). Columbia-Suicide Severity Rating Scale (C-SSRS) [Database record]. APA PsycTests.
- Quirk, K., Miller, S., Duncan, B., & Owen, J. (2013). Group Session Rating Scale: Preliminary psychometrics in substance abuse group interventions. *Counselling and Psychotherapy Research*, 13(3), 194-200. doi:10.1080/14733145.2012.744425
- Schmidt, I. D., Pfeifer, B. J., & Strunk, D. R. (2019). Putting the “cognitive” back in cognitive therapy: Sustained cognitive change as a mediator of in-session insights and depressive symptom improvement. *Journal of Consulting and Clinical Psychology*, 87(5), 446-456. doi:10.1037/ccp0000392
- Singla, D. R., Weobong, B., Nadkarni, A., Chowdhary, N., Shinde, S., Anand, A., Fairburn, C. G., Dimijdan, S.,

- Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine*, 166(10), 1092-1097.
- Strunk, D. R., DeRubeis, R. J., Chiu, A. W., & Alvarez, J. (2007). Patients' competence in and performance of cognitive therapy skills: relation to the reduction of relapse risk following treatment for depression. *Journal of Consulting and Clinical Psychology*, 75(4), 523-530. doi:10.1037/0022-006X.75.4.523
- Wechsler, T. F., Kümpers, F., & Mühlberger, A. (2019). Inferiority or even superiority of virtual reality exposure therapy in phobias?—A systematic review and quantitative meta-analysis on randomized controlled trials specifically comparing the efficacy of virtual reality exposure to gold standard in vivo exposure in agoraphobia, specific phobia, and social phobia. *Frontiers in Psychology*, 10. doi:10.3389/fpsyg.2019.01758
- Whiteford, H. A., Degenhardt, L., Rehm, J., Baxter, A. J., Ferrari, A. J., Erskine, H. E., ... & Vos, T. (2013). Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *The Lancet*, 382(9904), 1575-1586. [https://doi.org/10.1016/S0140-6736\(13\)61611-6](https://doi.org/10.1016/S0140-6736(13)61611-6)
- Whoqol Group. (1998). Development of the World Health Organization WHOQOL-BREF quality of life assessment. *Psychological medicine*, 28(3), 551-558.
- Witmer, B. G., & Singer, M. J. (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*, 7(3), 225–240. doi:10.1162/105474698565686
- World Health Organization. (2010). *Covid-19 Pandemic Triggers 25% Increase in Prevalence of Anxiety and Depression Worldwide*.