UNIVERSITY OF CALIFORNIA, SAN DIEGO UCSD

BERKELEY • DAVIS • IRVINE • LOS ANGELES • RIVERSIDE • SAN DIEGO • SAN FRANCISCO SANTA BARBARA • SANTA CRUZ

DEPARTMENT OF PSYCHIATRY
UCSD SCHOOL OF MEDICINE
858-822-4251
cdepp@health.ucsd.edu

February 17, 2026

This following document is linked with the following ClinicalTrials.Gov information

NCT Number NCT05899348

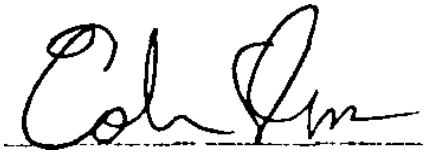
Unique Protocol Id 1R61MH129379-01A1

Title: iTEST: Introspective Accuracy as a Novel Target for Functioning in Psychotic Disorders

Note that for documents that are approved by the UC San Diego Human Protections Program the title of the study was listed as "A **Clinical Trial of iTEST: A Blended Intervention Targeting Introspective Accuracy"**

The last modified document date for the Study Protocol was 5/30/2023 and for the Informed Consent Document was 5/30/2023.

Best wishes,



Colin A Depp, Ph.D
Professor of Psychiatry, UCSD

UC San Diego

Office of IRB Administration (OIA)
Biomedical Protocol for Interventional Research
Version Date 12.08.21

Clinical Trial of iTEST: A Blended Intervention Targeting Introspective Accuracy

Protocol Number: 806997

National Clinical Trial (NCT) Identified Number: Pending

Principal Investigator: Colin Depp

Funded by: NIH (Pending) R61MH129379-01A1

Version Number: v.2

5/30/23

Summary of Changes from Previous Version:

Affected Section(s)	Summary of Revisions Made	Rationale
Multiple	UCSD HRPP requested changes	

Table of Contents

1 PROTOCOL SUMMARY

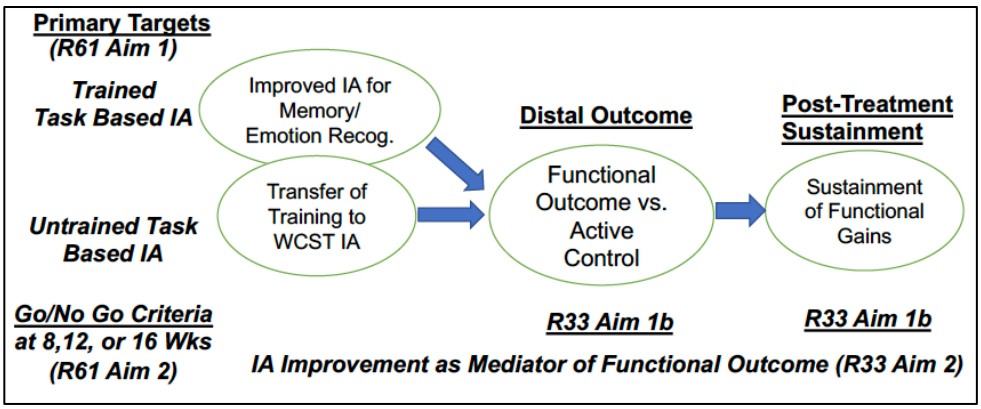
Leave blank. Text should be included under the relevant subheadings below.

1.1 SYNOPSIS

Title:	**Clinical Trial of iTEST: A Blended Intervention Targeting Introspective Accuracy**
Study Description:	
Objectives:	***Aim 1*. Target Engagement**: Evaluate in an open trial (n=60) if iTEST is acceptable and improves Introspective accuracy (IA); ***Aim 2*. Optimal Dosing:** Identify the briefest duration of iTEST (8, 12, or 16 weeks) that achieves criteria.
Endpoints:	*Primary Endpoint:* ≥75% adherence, and Cohen's d=0.5 increase in trained IA on two tasks and in an untrained IA task. *Secondary Endpoints:* Change in functioning based on a Cohen's d=0.5 increase in the Specific Level of Functioning scale score
Study Population:	People with psychotic disorders
Phase:	Ib
Description of Sites/Facilities Enrolling Participants:	UCSD; University of Texas at Dallas (Subcontracted site to UCSD)
Description of Study Intervention:	Behavioral intervention that integrates individual coaching with automated software that operates on mobile devices
Study Duration:	2 years
Participant Duration:	16 weeks

1.2 SCHEMA

This study is phase Ib (referred to an R61) of an NIH phased award (pending) wherein the goal is to determine target engagement in an open trial - the protocol here only reflects the R61 phase but he diagram below addresses the entire two phase project If the R61 reaches its go/no go criteria, the R33 is a randomized trial that would be a different protocol and is not detailed in this document. That R33 protocol would be separate and will be submitted at a later time pending the outcome of the R61 (protocol detailed in this document).



1.3 SCHEDULE OF ACTIVITIES (SOA)

All activities listed will be administered on both sites.

Procedures	Screening Day -7 to -1	Baseline Day 0	Study Visit 2 Day 56 +/-14 day	Study Visit 3 Day 84 +/- 14 day	Final Study Visit 4 Day 112 +/- 14 day
Informed consent	X				
SCID-5	X				
WRAT-4	X				
Demographics		X			
Medical history		X			
BACS		X			
VRFCAT		X			
MINI		X			
PSQI		X			
WHODAS 2.0		X			
PHQ-9		x			
PANSS		X	X	X	X
Calgary Depression Scale (CDSS)		X	X	X	X
Treatment Rationale Scale		X	X	X	X
CSSR-S		X	X	X	X
IA on METER and VLMT		X	X	X	X
Metacognitive WCST		X	X	X	X
Informant-Rated SLOF		X			X
GPS			X	X	X
Compensatory Cognitive Strategies Scale (Informant)		X			X
CAINS Motivation		X	X	X	X
Defeatist Performance Beliefs		X	X	X	X
Client Satisfaction Questionnaire (CSQ)		X	X	X	X
Negative Effects			X	X	X
Trials and Coaching Sessions Completed			X	X	X
HVLT			X	X	X
Metacog ER-40			X	X	X
Brief Regulation of Motivation Scale			X	X	X
Karolinska Sleepiness Scale			X	X	X
GAD-7			X	X	X

2 INTRODUCTION

Leave blank. Text should be included under the relevant subheadings below.

2.1 STUDY RATIONALE

Psychotic disorders are among the most disabling illnesses in the world[1], yet trials of cognitive training and related rehabilitative interventions have struggled to demonstrate durable impact on functional outcomes[2,3]. In part, this is due to a common target focus on cognitive capacity or first-order ability, which predicts no more than 20% of variation in functional outcome[4]. Introspective accuracy (IA), i.e., second-order judgements of one's abilities as compared to objective evidence, accounts for substantial variability in functional outcome independent from capacity[5,6,7]. As such, IA is a potential intervention target for functioning, but no interventions have directly and experimentally targeted IA in psychotic disorders using any format. This Experimental Therapeutics project tests a new blended digital health intervention called Improving Thinking through Everyday Self-Assessment Training (iTEST) that targets IA to ultimately improve functioning.

2.2 BACKGROUND

Schizophrenia and schizoaffective disorder, referred to hereafter as *psychotic disorders*, affect 1.5% of the population and cause the most substantial disability, greatest associated healthcare and indirect costs per person affected of any psychiatric disorder[1,11,12]. Over the past 20 years, a large body of clinical trials have evaluated interventions with cognitive capacity as a focal target for reducing this disability, including large scale NIMH-led efforts such as MATRCS[13], TURNS[14], and drug and behavioral trials[15]. To date, therapeutic impact on everyday functioning has been modest at best. A 2019 meta-analysis of 24 studies of drill-and-practice cognitive training found improved cognition but no impact on function[3]. Cognitive remediation interventions have much larger effects on functioning, but are intensive (35+ hours training), and the impact on function may not be durable[9,16-17]. Metacognitive components are important for transfer to change in function[18], but are not a focus of experimention[19], and so a key need is for experimentally tested efficient treatments targeting mechanisms underlying metacognition impairments for durable functional improvement[20,21].

A growing body of literature spanning cognitive science, education, and more recently, psychiatry, has associated inaccurate judgements of performance, or poor Introspective Accuracy (IA), with poor functioning. We define IA as the discrepancy between objective data and subjective estimation of task performance, ability, or behavior[5,22]. Our published work and that of others has identified the following: 1) Poor IA is Predictive of Poor Functioning: Poor IA, as measured by greater discrepancies between actual and judged performance, accounts for up to 42% of the variance in disability in social and nonsocial functional domains[23-26].; 2) IA is Independently Associated with Functioning Beyond Capacity: The contribution of IA to disability is independent from that associated with actual cognitive and social cognitive performance, and the relationship between cognition and functioning is minimal (~3% of variance)[27,28] once IA is accounted for; 3) IA Impairments are Worse in Psychotic Disorders vs. Other Conditions: In contrast to bipolar disorder, people with psychotic disorders produce judgements that are not correlated with actual performance despite contemporaneous feedback[29], indicating substantial disruption in self-assessment processes; 4) IA Impairments are Conserved across Domains and

Measurement Modalities: Discrepancies between self-assessments and informant reports of functioning are associated with task-based IA deficits[5,23]. In each of these studies, participant self-assessment was poorly linked to objective data even with performance feedback (e.g. task feedback[30]; objective data regarding vocational outcomes[28,31])

In contrast to psychiatric research, interventions targeting performance judgments are an active domain of cognitive science and education research in healthy adults. In a recent study, an experimental IA training intervention in healthy adults improved IA on a trained task as well as transfer of training to IA for a novel task[10]. In that intervention, adults were trained by repeated immediate feedback on the accuracy of their performance judgments on a signal detection paradigm, over the course of 8 weeks. The control condition received feedback regarding their performance but not on the accuracy of their self-ratings (i.e., IA). This intervention was self-directed and computerized, in aggregate accounting for only 8 hours of exposure. The active condition improved IA, and that improvement transferred to a novel task, whereas no such effects were evident in the control condition. As such, a targeted computerized drill-and-practice intervention could improve IA with generalization to IA on other tasks. Related experimental work in the educational literature indicates that improvement of task-based judgments of performance has a strong impact on learning[50,51], which has recently been extended to functional pursuits like employment searches[52,53] (again in healthy adults). To be impactful in a clinical population, a key consideration is distal transfer to functioning[54]. In experimental studies, poor IA is related to reduced use of compensatory strategies like reminders[55], as well as increasing risk for failure experiences when real-world tasks are attempted[56]. Thus, without training in IA, use of compensatory strategies may be low and chances of failure may increase. Given the potential pitfalls in distal transfer for translational interventions[57], iTEST's task-based training for cognitive IA is directly coupled with improving global judgements of actual behavior measured by EMA (i.e., IA for functional activity). Further, iTEST coaching centers on connecting IA training to use of compensatory strategies (e.g., calendaring, reminders) for achieving personalized goals. Thus, iTEST would be the first intervention of any kind to target IA in psychotic disorders, and would also be the first intervention to target IA using a blended mobile intervention format for any condition.

Although computerized IA training could occur in a lab-based setting, there are several advantages to a remote "virtual first" development strategy[59]. The need for and practicality of clinic-based training has diminished with the emergence of telehealth, and mobile interventions could increase access and reduce transportation costs[60]. Mobile interventions that require daily/several times per day interventions are acceptable, feasible, and effective in psychotic disorders, as our group[61-63] and others[64-67] have shown. Given the shift to telehealth, interventions may benefit from initial deployment in a remote format (rather than post-development adaptation). In addition, employing mobile devices as a platform for iTEST, a potentially transformative approach is that we can also link judgements of task performance to contemporaneous information (time and place, and social context) of real-time behavior, directly transferring IA on cognitive tasks to improved accuracy for judgements of actual engagement in functional activity.

iTEST's design incorporates recommended elements (e.g., human involvement, personalization) of cognitive remediation from a consensus panel[8]. That panel included experts in cognitive remediation and created a published consensus document called a white paper that detailed what the recommended ingredients of cognitive remediation for psychotic disorders should include, focusing on increasing adherence and potential impact. Yet, iTEST does not involve broad capacity-building activities. As such, the intensity of iTEST amounts to ~50% of the average contact time of cognitive remediation, which, if effective in improving function, would yield a favorable acceptability and scalability profile. The Task-based components of iTEST build upon our validated Mobile Variable Difficulty List Memory Test [VLMT])[9] and Mobile Electronic Test of Emotional Recognition (METER)[10], each of which includes IA judgements of performance and together target IA for two different cognitive domains (verbal learning and facial emotion recognition).

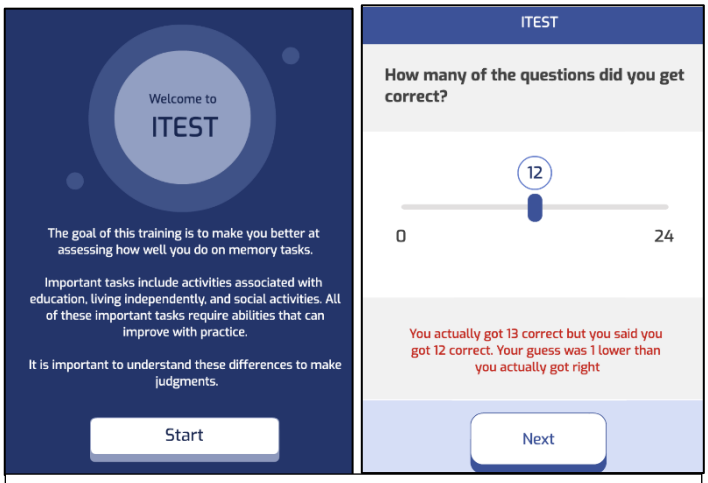


Figure 1. Selected Screen Shots of iTEST (Alpha Version)

During each VLMT and Meter administration, participants are presented with 3 trials where lists of words or faces are presented. For the VLMT, immediately following each exposure to the list, participants were shown target and recognition foil words one-by-one and asked to indicate whether or not the word appeared on the list. For the METER task, participants are asked to guess the emotion in a picture of a human face from among anger, sadness, happiness, no emotion, or fear. After completion of the task, scoring includes the percentage of correct identifications of target words or faces per trial. In embedded validity tests, after each administration, participants self-reported if they were distracted or interrupted.

These tasks are an ideal foundation for iTEST since they are proven to be acceptable, correlated with in-lab performance on tests of the same construct in people with psychotic disorders[9,10], and associated with functional outcome[11]. In addition, there were no practice effects or changes in IA associated with repeated testing either task, so administration of the tasks alone does not alter task performance or IA. The iTEST intervention contains brief (~15 min) Task-based IA training delivered 6 days per week on the individual's mobile device to improve accuracy of task-level and global judgments of performance. Training uses gamified adaptation of both the VLMT and METER to improve IA in two separate, but critical functional domains – memory and social cognition, thus broadening potential impact. Training is graduated, such that initial stages focus on engaging attention on the parameters of the task, then formation of global judgements based on trial-level feedback (rather than self-generated information) for both memory and social cognitive domains. In so doing, increased and continuous attention to external feedback (vs. self-generated information) is shaped. Individual Coaching is coupled with the task-based IA remote training and includes psychoeducation in how IA problems impact function and training in use of compensatory strategies (e.g., reminders, calendaring, asking others how they are feeling) to apply to everyday cognitive and social cognitive demands that are linked to individualized personal goals. Two rounds of alpha-testing of iTEST have been completed.

Since iTEST is the first intervention to target IA in psychotic disorders, and the first blended intervention to target IA in any condition, the scope of this research is a pilot trial. The goal of the R61 phase and this protocol is to evaluate whether iTEST can substantially improve IA and is associated with acceptable adherence.

2.3 RISK/BENEFIT ASSESSMENT

Leave blank. Text should be included under the relevant subheadings below.

2.3.1 KNOWN POTENTIAL RISKS

Assessment: Participation in the evaluation and responding to questions on the mobile device requires time and attention. Some participants may become bored or fatigued completing the evaluation. Some participants may also be come distressed during the psychiatric evaluation, which requires them to discuss their current symptoms, as well as history of their symptoms.

***Intervention*:** Although the blended computerized interventions to be tested in this study are not expected to produce adverse reactions, participants may be become bored or fatigued with the intervention procedures. There is a theoretical possibility that improvements in introspective accuracy may engender increases in depression alongside greater awareness of limitations.

Disclosure or Identification of Suicidal Ideation or Intent Harm Self: During structured interviews concerning suicide (via the Columbia Suicide Severity Scale[12] or CSSR-S) or extemporaneous discussion with rater or other staff, participants may disclose a level of risk of harming themselves or others that would require an evaluation for a higher level of care.

***Confidentiality/Loss of Privacy*:** As with any clinical research study, a risk of disclosure of personal material exists. These risks are present in interview and scale-based measures and technology-based data collection.

2.3.2 KNOWN POTENTIAL BENEFITS

Participants may benefit from participating in the experimental intervention by learning skills to improve introspective accuracy and may also experience, based on study hypotheses, improvements in functioning.

2.3.3 ASSESSMENT OF POTENTIAL RISKS AND BENEFITS

The risks and potential adverse effects associated with this research are minimal. We judge the importance of knowledge resulting from this study and the potential for improving functioning among people with psychotic disorders to be quite significant; therefore, the risk to benefit ratio is low.

3 OBJECTIVES AND ENDPOINTS

In in this pilot project, there are two objectives. Firstly, we will evaluate, in an open trial, whether iTEST is acceptable, indexed by adherence, and engages IA as a target, indexed by improved IA on trained

tasks and transfer on an untrained task in a different domain. Secondly, we will evaluate optimal dosing of the intervention, by whether go/no-go criterion for improvement is achieved at 8, 12, or 16 weeks.

Our go/no go criteria are threefold: 1) To demonstrate feasibility, the average adherence must be equal to or exceed 75%, 2) To show target engagement, we require a significant improvement in IA (*d*>.50) on both trained tasks, and 3) To show transfer of training, we require significant improvement in IA (*d*>.50) on the untrained Wisconsin Card Sorting Task. Criteria may be met prior to 16 weeks, at weeks 8 or 12 (Note: all participants continue through 16 weeks).

OBJECTIVES	ENDPOINTS	JUSTIFICATION FOR ENDPOINTS
Primary		
To evaluate the acceptability and preliminary impact on our target, IA	Adherence defined as the ratio of completed mobile prompts completed divided by the number possible – 75% is the criterion Clinically significant (Cohen's d > 0.5) improvement in trained introspective accuracy as measured by IA on our Verbal Learning and Memory Test and Mobile Electronic Test of Emotion Recognition tasks. These are administered 6x per week for 16 weeks. Clinically significant (Cohen's d > 0.5) improvement in untrained introspective accuracy (global resolution score) on the WCST. The WCST is administered at baseline, 8, 12 and 16 weeks.	Adherence at a level of 75% is generally accepted in mobile intervention studies such as this The reason that trained tasks are part of our Go/No Go criteria is because lack of improvement on trained tasks would call into question the ability of iTEST to modify IA, and would indicate need for program revisions. IA is calculated as the discrepancy (absolute) between each trial's correct performance from that estimated by the participant, which is nested within sessions and weeks. An additional Go/No Go component is transfer of training to an untrained task-based measure of IA, the Meta-Cognitive Wisconsin Card Sorting Test[13]. IA on this task is also associated with functioningand improvement on this task would support that iTEST

OBJECTIVES	ENDPOINTS	JUSTIFICATION FOR ENDPOINTS
		leads to generalized improvement in IA.
Secondary		
To identify the briefest duration of iTest that in which the sample reaches criterion	Reaching Criterion at 8, 12, or 16 weeks	Our rationale for the dosing framework is as follows. Our starting dose is that in-lab IA intervention in healthy adults occurred over 8 weeks and included 2160 total trials and 80 instances of IA feedback (in aggregate ~6 hours)[14]. iTEST at 8 weeks equates to 4320 trials and 188 rounds of feedback (8-10 hours), whichis about 2 times the dose of the in-person IA intervention in healthy adults[14]. We will then extend this period to 16 weeks with additional evaluations at 12 and 16 weeks, yielding a maximum of 16-20 hours of training in aggregate.
Tertiary/Exploratory		
To evaluate the impact of iTEST on functional impairment	The informant-rated *Specific Level of Function (SLOF)* scale will be used to measure functional outcome[15]. The SLOF is a 43-item interview administered as informant- or self-report. The scale assesses the participant's current functioning integrated into a single higher-order factor[16] integrating social, everyday, and vocational activities.	The SLOF emerged as the most reliable and externally valid scale-based measure of functioning in the team's prior validation study[17]. Informants ratings on the SLOF were also found to be sensitive to short term functional gains (12-24 weeks) in a treatment study[18].

4 STUDY DESIGN

Leave blank. Text should be included under the relevant subheadings below.

4.1 OVERALL DESIGN

Hypothesis:

Target Engagement Hypothesis: Participants' adherence and average changes in IA for the mobile tasks and pre- post differences in meta-cognitive WCST will meet or exceed our go/no go criteria

Optimal Dosing: Target engagement may occur prior to 16 weeks, at week 8 or 12. We would select the optimal dose for the R33 is determined by selecting the earliest assessment point that reaches go/no go criteria - all participants receive the same dose

Phase of the trial, if applicable: This is a pilot behavioral intervention and so study phases are somewhat difficult to apply but would be mostly consistent with a Phase Ib study that is investigating acceptability and a preliminary effect in a diseased population

Design of Trial: Single arm, unblinded, open pilot trial.

Methods used to minimize bias: Measures are standardized and accepted endpoints and the sample has few exclusion criteria for entry

Dosing: *All participants receive the same dose. We will investigate the point at which (among 8, 12, and 16 weeks) go/go criteria are met.*

Intervention groups: *This is a single arm trial and the duration of the intervention is 16 weeks.*

Sites: *There will be 2 sites – UCSD is the prime site and University of Texas at Dallas will be the secondary site with their own IRB.*

Study Intervention*: iTEST*

Interim Analysis: *No interim analysis is planned*

Stratification: *None*

Name of sub-studies, if any, included in this protocol*: None*

4.2 SCIENTIFIC RATIONALE FOR STUDY DESIGN

The rationale for the study design is to provide a test of acceptability and target engagement. The design was chosen according to the sponsor's (NIMH) experimental therapeutics framework, which supports testing of novel therapeutics with a sequence of trials first evaluating target engagement and then, if go/no go criteria are met, an efficacy trial of the intervention wherein outcomes are evaluated and the relationship between target engagement and outcomes is evaluated. This protocol is aimed at evaluating target engagement and so an open trial was selected. Since there are no prior studies evaluating whether introspective accuracy is modifiable in schizophrenia, we have selected an open pilot trial format.

4.3 JUSTIFICATION FOR DOSE

Our rationale for the dosing framework is as follows. Our starting dose is that in-lab IA intervention in healthy adults occurred over 8 weeks and included 2160 total trials and 80 instances of IA feedback (in aggregate ~6 hours)[14]. iTEST at 8 weeks equates to 4320 trials and 188 rounds of feedback (8-10 hours), whichis about twice the dose of the in-person IA intervention in healthy adults[14]. The reason to start with a more intensive "dose" of iTEST in psychotic disorders is that we would expect a higher level of impairment in iA than in a healthy population. We will then extend this period to 16 weeks with additional evaluations at 12 and 16 weeks, yielding a maximum of 16-20 hours of training in aggregate. . This *maximum* dose of training at 16 weeks is 75% of the mean training duration reported in a 2020 meta-analysis of available computerized cognitive training[19] and ~50% of the average duration of cognitive remediation programs[20]. When considering the amount of epochs of training, the dose is lower than the number of interactions required in our prior ecological momentary interventions[21-23].

4.4 END OF STUDY DEFINITION

A participant is considered to have completed the study if he or she has completed all phases of the study including the last visit or the last scheduled procedure shown in the Schedule of Activities (SoA), Section 1.3.

The end of the study is defined as completion of the last visit or procedure shown in the SoA in the trial globally.

5 STUDY POPULATION

Leave blank. Text should be included under the relevant subheadings below.

5.1 INCLUSION CRITERIA

There are two populations in this study. The iTEST study population consists of people who are adult outpatients with diagnoses of schizophrenia (n=60, divided equally between UCSD and U Texas at Dallas). There are also informants nominated by the participant who provide ratings on the two scales (one pertaining to functioning and the other to cognitive strategies). We detail itest participants first

Potentially eligible participants recruited through flyers, word of mouth or community presentations will complete a brief screening to review eligibility. That brief pre-screening details the study and consists of questions about basic eligibility. Thereafter an appointment is made for a lab visit provide that basic eligibility questions are met, and the participant is interested in the study.

During recruitment, all participants will receive an initial phone or in-person screening. The phone screening can involve a short survey via REDcap and a 1-on-1 screening call with a researcher. At this point, they will be asked basic demographic information and asked if they have a current psychiatric diagnosis. Only participants who are able to consent for themselves and fully understand the study procedures will be allowed to participate. For remote screening and consenting visits, consents will be obtained via a HIPAA compliant platform such as REDCap.

The screening and trial have the same informed consent.

The Screening which is held in person consists of:

A questionnaire concerning basic eligibility (e.g., current treatment stability)
A diagnostic assessment to confirm diagnosis (SCID)
A measure of reading level (WRAT-4)
A measure of psychiatric symptom severity (PANSS)

Participants are compensated for completing the screening at a rate of 20 dollars. The duration of the screening is estimated to take 30-40 minutes.

In order to be eligible to participate in this study, an individual must meet all of the following criteria

Inclusion Criteria:

(1) Voluntary informed consent to participate and capacity to consent as measured by the UCSD Brief Assessment of Capacity to Consent (UBACC)[24];
(2) Age 18 to 65;
(3) DSM-5 diagnosis of schizophrenia or schizoaffective disorder based on a SCID-5 interview and available medical record review;
(4) ≥ 6th grade reading level on the Wide Range Achievement Test-4 Reading (WRAT-4) subtest (needed to read instructions on device)[25];
(5) Stable co-treatments (no hospitalizations or medication class changes in 2 months before enrollment). We will determine symptom and medication stability by best-estimate history with information from medical records;
(6) Availability of a clinician (staff member, case manager, other mental health clinician) or close associate (family member, friend) with at least monthly contact who can be their informant – the purpose of informants is to complete the SLOF which is a rating scale measure of functional ability and the Cognitive Strategies which assess use of compensatory strategies in daily life;
(7) Minimum level of functional impairment based on milestones, excluding participants who are full-time employed and financially responsible for their household.

Informants are nominated by the participant. As such, we are seeking a partial waiver of consent to reach out them to invite them to participate in the research. Following a standard phone script (see script), informants are invited to participate in the study and their nature of participation is detailed. If verbal consent is obtained, the questionnaires are administered by telephone at the same time as consent.

Our justification for a partial waiver of consent for informants is as follows: 1) The research component involves no more than minimal risk to subjects; 2) The research could not be carried out practicably without the waiver since the informant must be nominated by the participant; 3) The waiver will not adversely affect the rights and welfare of the subjects; and 4) Where appropriate, the subjects will be provided with additional information about their participation.

If a participant's nominated informant cannot be reached or declines participation, we will ask the participant to provide another informant. In the unlikely event that no informants agree to participate, the participant will not be dropped from the study.

5.2 EXCLUSION CRITERIA

An individual who meets any of the following criteria will be excluded from participation in this study:

Exclusion Criteria:
(1) Greater than moderate disorganization on the PANSS (P2-Disorganization item >5)[26] because disorganized patients have had poorer outcomes in prior blended interventions in our lab due to difficulty learning to use mobile devices;
(2) DSM-5 alcohol or substance dependence in past 3 months based on the SCID;
(3) Level of care required interferes with outpatient therapy (e.g., hospitalized; severe medical illness); and
(4) Unable to adequately see or manually manipulate a smartphone**.**

This study will include individuals between the ages of 18 and 65. Children under the age of 18 are excluded for several reasons. Since the key outcome of the iTEST intervention is independent functioning, the research topic is not relevant to many children who reside with family members and who are not expected to achieve milestones of independence as are adults. Moreover, the mechanisms underlying introspective accuracy have not been fully evaluated in children with psychotic disorders, and so children with psychotic disorders represent a unique treatment population that is different from the adult patient population that attends outpatient community mental health clinics. At this point it is unclear if the treatment needs of children mirror those of adults. The upper end of the included age range is 65 years. This is due to a number of reasons including the upper limit on the available normative data for certain assessment measures and the need to avoid a confound with age-related physical declines in functioning. It would also be necessary to tailor the mobile application and adapt the in-person intervention to the unique needs of older adults (e.g., visual and hearing impairments, ageist beliefs/self-stigma, age-related role changes, loss and isolation, greater severity of cognitive impairments).

We considered a broader transdiagnostic focus but our preliminary data on direct comparisons strongly suggest people with psychotic disorders have significantly more profound deficits in IA[27], notably in responsiveness to feedback[28], than do people with bipolar disorder

5.3 LIFESTYLE CONSIDERATIONS

Not Applicable

5.4 SCREEN FAILURES

Screen failures are defined as participants who consent to participate in the clinical trial but are not subsequently assigned to the study intervention or entered in the study. A minimal set of screen failure information is required to ensure transparent reporting of screen failure participants, to meet the Consolidated Standards of Reporting Trials (CONSORT) publishing requirements and to respond to queries from regulatory authorities. Minimal information includes demography, screen failure details, eligibility criteria, and any serious adverse event (SAE).

Individuals who do not meet the criteria for participation in this trial (screen failure) because of a substance use disorder or disorganization on the PANSS may be rescreened after a minimum of one

month which is the duration follow back on these instruments. Rescreened participants should be assigned the same participant number as for the initial screening.

Screen failures will be provided with referrals to mental health services in the community by way of a resource list that contains information about walk-in, free and low cost services through the San Diego County Mental Health department.

6 STUDY INTERVENTION

Leave blank. Text should be included under the relevant subheadings below.

6.1 STUDY INTERVENTION(S) ADMINISTRATION

Leave Blank. Text should be included under the relevant subheadings below.

6.1.1 STUDY INTERVENTION DESCRIPTION

iTEST consists of several integrated components – coaching and mobile web-based software

***Coaching*:** iTEST begins with a 60-minute Introductory Coaching session and then weekly for 5 weeks thereafter. These sessions can be delivered remotely or in-person, per preference of the participant. Sessions are scripted and manualized (see *Fidelity* below) as our prior blended health interventions[21-23,29]. The introduction session provides the basis of iTEST (60 minutes) through psychoeducation on IA in functional recovery, application to compensatory strategies and goal setting, with a personalized goal identification activity derived from prior work[23,30]. Thereafter, coaching in iTEST involves the following structure: 1) review and feedback on mobile training adherence and IA scores (15 minutes), 2) training in application of improved IA to everyday activities (30 minutes), and 3) training, selection, and application of compensatory strategies to weekly goals (15 minutes). Each 60 min session begins with reviewing the week's IA scores and problem solving about any barriers to engagement. For the remainder of the session, Coaches develop action plans that incorporate IA concepts. Materials are derived and adapted from Action-Based Cognitive Remediation[30]. In each meeting, the Coach collaborates to identify and plan two functionally relevant activities that involve at least some cognitive challenge *("one thing you want to do and one thing you have to do in the next 2 weeks*"). The pair collaborates to generate a list of action steps via a cognitive "walkthrough" simulating the initiation, engagement, and completion of the goal-linked social or non-social activities. Challenges due to memory, social reasoning, or attention failures are thus minimized, and the Coach demonstrates and trains in use of compensatory strategies, including setting reminders, scheduling via calendars, and asking (rather than inferring) others' emotions and/or intentions. These are personalized (e.g., use of electronic or paper-based techniques). During the walkthrough of action steps, queries include identification of cues or other evidence that signifies progress or setbacks, which align with IA training in increasing attention to external cues (vs. self-generated impressions). The tasks of generating action steps and selecting compensatory strategies gradually shifts toward becoming fully participant driven. Coaching intensity is consistent with our past studies[21-23].

Fidelity and quality assurance for coaching: Participants are asked to provide consent for audio recordings for fidelity monitoring of coaching sessions. Audio files are then reviewed by Dr. Depp and other individuals who are part of the coach supervision team (who are included as personnel on the IRB). Audio files are kept on a secure server and rated according the process below. Once rated, audio

files are destroyed.

We will conduct a two-day training workshop for coaches and provide weekly supervision to the coaches. All treatment sessions, messaging, telephone contacts (iTEST and CC) are rated according to fidelity to the iTEST treatment model. Sessions will be rated using an adapted version of the *Cognitive Therapy Rating Scale for Psychosis (CTS-Psy[31])* Sessions have been alerted to capture the core elements of iTEST. We will evaluate these fidelity scores based on a threshold of 80% adherence to the iTEST model, and scores below 80% adherent would lead to discussions with the coaches to correct/modify their approach so as to improve fidelity. Coaches will be bachelor's or master's level with comparable experience to a community clinician and will receive weekly 1-hour supervision from Dr. Depp.

***Automated Drill-and-Practice Training with Mobile Web-based Platform*:** For onboarding, device training is provided on how to operate and charge the smartphone, the meaning of questions and response choices, personalizing training alerts, and how to access crisis lines. The participant completes three practice rounds of iTEST in which the focus on IA and scoring is explained. If participants do not evidence understanding of the basis of scoring (e.g., mistake performance vs. IA accuracy as the basis of scoring), additional rounds of practice will be completed. Next, the coach collaborates on a training schedule (time of day, preferred day off) and identifies potential barriers (e.g., interruptions, fatigue) and accompanying solutions, along with facilitators to practice (e.g., small self-selected available non-monetary rewards for completion). Finally, a schedule is developed for follow-up contacts and review of progress. Daily training will begin the day after this initial session.

Training involves prompts 6x per week to engage in the Task-Based training, which lasts 10-15 minutes per session and consists of three blocks each of IA on verbal memory and emotion recognition. The reason to deliver task-based training 6x per week was to provide one day off (which can be selected by the participant). Alarms can be silenced for 30-minute intervals (e.g., driving, appointments). If a participant misses a session, they will receive 2 reminders that day to complete it. If they miss two consecutive days (or one day during the first week), participants will be contacted by their Coach to engage in problem solving and motivational support. Our previous data on remote adherence suggested that early adherence was highly predictive of total adherence (r=.80)[32], so daily first week outreach will occur in case of non-adherence.

Our web-based smartphone compatible software is built around our previously validated VLMT and METER tasks. This flows from a) introductory/instruction screens, b) presentation of stimuli, c) performance assessment (initially with feedback on correctness), and d) IA queries (how many words/faces correct, confidence) e) feedback about IA. Items b-e are then repeated for the second and third trials of the tasks. Stimuli include variable numbers of words or faces with the number of stimuli randomly ordered per session and derive from a large normed corpus so that stimuli are not repeated. Training is graduated in four steps: (1) The first skill is correctly identifying the number of stimuli presented (task awareness); (2) After 3 consecutive sessions in which the number of stimuli is correctly identified, participants are provided *trial-by-trial* feedback on correctness of response and scoring is based on correct assimilation of trial-by-trial feedback to global estimation of performance (e.g., *"you indicated that you got 10 correct; however, you actually got 8 correct*"); (3) after 3 consecutive sessions/days of high IA (+/- 1 discrepancy between actual correct and self-assessed), training then progresses to presentation of stimuli without trial-by-trial feedback wherein more global estimates are

trained; (4) In the final stage, response options are expanded to include a level of certainty rating, and these responses are connected to scoring to incentivize choice for high- and low-confident responses across levels of difficulty. Feedback is provided in the form of discrepancy and its direction (i.e., "*you guessed a number correct that was 3 words higher than what you actually got correct*") which is translated to a score that is displayed in a "leaderboard" displaying progress during training. At the end of each session, questions ask what strategies the participant used to estimate their performance and what strategies could improve their score, with responses collaboratively developed with the Coach during coaching.

To deliver iTEST, we will use the NeuroUX platform, https://www.getneuroux.com/, a HIPAA compliant platform, for administration of iTEST. Information Technology infrastructure will securely collect the data gathered by smartphones and will store the information via transfer to an encrypted Amazon Web Service (AWS) server. AWS offers a simple and intuitive web-based user interface from user to server. It is an easy-to-use cloud computing platform that allows various storage services for backup, archiving, and disaster recovery use cases. AWS enables robust safeguards to help protect the privacy of the data with highly secure AWS Data centers. Specifically within AWS, data will be stored in the Amazon S3 service and will be used for large-scale analytics. Amazon S3 provides scalable and durable object storage in the cloud. S3 is where the experimenters will be storing the user information (phone number, start and end time) along with the survey links. Once the data is generated, the data will be stored securely in this server, which will only be accessible by study staff. The transmitted program elements are encrypted and separated from any identifiable information. No information is stored locally on the smartphone device and cannot be accessed by anyone if the device will be lost or stolen. The developer wrote a whitepaper for Institutional Review Boards that gives an executive summary of Amazon Web Service's security: https://aws.amazon.com/whitepapers/.

The research team will input the participant's phone number and the time of delivery of messages using an onboarding platform from Twilio. The onboarding platform passes the phone number and the delivery time windows to Dynamo DB (HIPAA Compliant), which then passes the info to a message queuing service, Amazon SQS (HIPAA Compliant). The NeuroUX platform then passes the queued messages to a cloud communications platform, Twilio (HIPAA Compliant). The same message is then forwarded to the downstream providers and to the destination carrier for the last leg of message delivery. With the scheduling dashboard, messages will have the option to be sent to email and/or to text messages.

As the training program is not an application on a smartphone, rather a hyperlink sent through an encrypted link to the text messages application, the assessments and cognitive training qualifies as a non-significant risk under FDA regulations and is experimental/investigational. The training is exempt from

IDE regulations for the following reasons, following recent 2022 FDA Policy for Device Software Functions and Mobile Medical Applications:

1) The software is not intended as an implant, nor does it support or sustain human life.
2) The use of the software is not important in diagnosing, curing, mitigating, treating disease, or preventing
the impairment of health.
3) The software does not cause significant harm to subjects – nor is it life threatening, cause permanent impairment of a body function, cause permanent damage to a body structure, nor could it necessitate

medical or surgical intervention to preclude permanent impairment of a body function or preclude permanent damage to body structure.
4). The participants will complete cognitive testing for the training and are not obligated to complete each survey session if they do not want to.
5). The training is not intended to be used for diagnostic purposes.
6). The training does not interact with any medical devices and does not impact the functionality or performance of traditional medical devices.
7). Furthermore, no blood draws or invasive collection methods will be used by the study owned smartphones or the participant's personal smartphones.
8) The subject does not need to go undergo any invasive procedures as part of the study.
9). The smartphone is used by the participant's fingers and will not enter the participant's body nor does it involve a break in skin, have contact with the mucous membrane or internal body cavity

Additional Screenshots of the software are attached to this protocol.

Transportation: For participants who are unable to meet in our research office space due to physical disability or financial reasons, we will use the UCSD Lyft Concierge service for transportation, provided that residence is within a 25-mile radius of UCSD. Participants will **NOT** be required to pay for the Lyft rides as the research study will cover the costs of these rides.

To use Lyft participants will be providing name, pick-up location, drop-off location, and phone number to the research staff, which will then be used by the Lyft driver. This information will only be given to the Lyft drivers after consent over the phone (during the phone screening procedures) and by providing written consent in the Informed Consent document. Participants phone number and the driver's phone number will be masked during the active ride through a third party system. Neither participant nor the driver will have access to each other's real phone numbers. Personal health information (PHI) will not be shared with Lyft.

If participants choose to their own mode of transportation, they are subject to any associated to charges or fees, however, they will not be required to pay for parking at our research office for the duration of the visit.

Compensation for assessments and screening:
Participants are paid up to $220.00 for in-lab assessments. A total of $20 is provided for completing the screening portion of Visit 1 and participants are compensated $50 for completing the rest of that visit. For Visit 2, 3, and 4 participants are compensated $50 for each.

We also have assessments delivered through the mobile device at Baseline (Visit 1) and each of the weeks that coincide with lab visits (i.e., weeks 9, 13, and 17).. Assessments will involve 6 days of completing brief mobile assessments that closely mirror the iTEST training, and participants will be paid 5 dollars for each of the 6 mobile assessments that they complete. The number of days completed is factored to the compensation and so participants can receive up to 30 dollars for completing all 6 assessments. Thus, over four assessment days, participants can be compensated up to 120 dollars (30 x 4 assessments).

Combining in-lab and mobile assessments, the maximum possible amount of compensation is $340.

Notably, participants are not compensated for meeting with the coach or for completing the training exercises on the mobile device. If participants choose to stop attending meetings with the coach or completing the mobile exercises, they may continue to participate in other aspects of the study and continue to receive compensation as above.

6.1.2 DOSING AND ADMINISTRATION

Administration is as above – Coaching is combined with 6 times per week completion of tasks on mobile devices for a total of six weeks

6.2 PREPARATION/HANDLING/STORAGE/ACCOUNTABILITY

Leave blank. Text should be included under the relevant subheadings below.

6.2.1 ACQUISITION AND ACCOUNTABILITY

The software used to deliver iTEST is developed by NeuroUX a vendor that generates computer software. The software is web-based. See above for the flow of information though the portal.

6.2.2 FORMULATION, APPEARANCE, PACKAGING, AND LABELING

Not Applicable

6.2.3 PRODUCT STORAGE AND STABILITY

Not Applicable

6.2.4 PREPARATION

Not Applicable

6.3 MEASURES TO MINIMIZE BIAS: RANDOMIZATION AND BLINDING

There is no randomization and blinding is not possible in this study.

6.4 STUDY INTERVENTION COMPLIANCE

Adherence is assessed by calculating the ratio of completed iTEST training sessions out of the number possible (6x per week by 16 weeks = 96). Adherence is monitored by a dashboard available to the

research team. If a participants misses two consecutive training sessions (or one in the first week) they are called by the research assistants to troubleshoot adherence.

6.5 CONCOMITANT THERAPY

Concomitant treatment is allowed, and no restrictions are placed. Concomitant medication and other treatments address other aspects of schizophrenia do not impact introspective accuracy. Based on our review of the literature there is no intervention, either medication or behavioral, that has been shown to lead to improvement or a negative impact on introspective accuracy in schizophrenia (which is an impetus for the study). Therefore, there is no cause for our trial of iTEST to restrict concomitant medications. Other behavioral interventions are also permitted as again introspective accuracy is a novel mechanism and would not overlap with other behavioral interventions.

6.5.1 RESCUE MEDICINE

Not Applicable

7 STUDY INTERVENTION DISCONTINUATION AND PARTICIPANT DISCONTINUATION/WITHDRAWAL

Leave blank. Text should be included under the relevant subheadings below.

7.1 DISCONTINUATION OF STUDY INTERVENTION

Discontinuation from iTEST does not mean discontinuation from the study, and remaining study procedures should be completed as indicated by the study protocol. If a clinically significant finding is identified (including, but not limited to changes from baseline) after enrollment, the investigator or qualified designee will determine if any change in participant management is needed. Any new clinically relevant finding will be reported as an adverse event (AE).

7.2 PARTICIPANT DISCONTINUATION/WITHDRAWAL FROM THE STUDY

Participants are free to withdraw from participation in the study at any time upon request.
An investigator may discontinue or withdraw a participant from the study for the following reasons:

- If any clinical adverse event (AE)) or other medical condition or situation occurs such that continued participation in the study would not be in the best interest of the participant

The reason for participant discontinuation or withdrawal from the study will be recorded on the Case Report Form (CRF). Research participants who sign the informed consent form but do not receive the study intervention will not be replaced. Research participants who sign the informed consent form, and receive the study intervention, and subsequently withdraw, or are withdrawn or discontinued from the study, will not be replaced.

7.3 LOST TO FOLLOW-UP

A participant will be considered lost to follow-up if he or she fails to return for any of the follow up visits scheduled visits and is unable to be contacted by the study site staff.

The following actions must be taken if a participant fails to return to the clinic for a required study visit:

- The site will attempt to contact the participant and reschedule the missed visit within two weeks and counsel the participant on the importance of maintaining the assigned visit schedule and ascertain if the participant wishes to and/or should continue in the study.
- Before a participant is deemed lost to follow-up, the investigator or designee will make every effort to regain contact with the participant (where possible, 3 telephone calls and, if necessary, a letter to the participant's last known mailing address or local equivalent methods. These contact attempts will be documented in the participant's study file.
- Should the participant continue to be unreachable, he or she will be considered to have withdrawn from the study with a primary reason of lost to follow-up.

8 STUDY ASSESSMENTS AND PROCEDURES

8.1 EFFICACY ASSESSMENTS

Primary Targets: In addition to adherence, the primary Go/No Go targets of the R61 include both IA on trained (METER, VLMT) and untrained task (WCST). During each VLMT and Meter administration, participants are presented with 3 trials where lists of words or faces are presented. For the VLMT, immediately following each exposure to the list, participants were shown target and recognition foil words one-by-one and asked to indicate whether or not the word appeared on the list. For the METER task, participants are asked to guess the emotion in a picture of a human face from among anger, sadness, happiness, no emotion, or fear. After completion of the task, scoring includes the percentage of correct identifications of target words or faces per trial. In embedded validity tests, after each administration, participants self-reported if they were distracted or interrupted.

The reason that trained tasks are part of our Go/No Go criteria is because lack of improvement on trained tasks would call into question the ability of iTEST to modify IA, and would indicate need for program revisions. IA is calculated as the discrepancy (absolute) between each trial's correct performance from that estimated by the participant, which is nested within sessions and weeks. IA on the VLMT and METER tasks predict functional outcomes and is stable over time without training.

An additional Go/No Go component is transfer of training to an untrained task-based measure of IA, the *Meta-Cognitive Wisconsin Card Sorting Test*[13]. Wisconsin Card Sorting Task (WSCT) (Heaton, Chelune, Talley, Kay & Curtiss, 1993). The WCST is a widely used measure of executive function that requires participants to match a presented card to one of four stimulus cards. There are several ways in which the card could be successfully sorted but participants are not informed of the criterion they should use to match the card. Instead, feedback is given after each sort (i.e. correct or incorrect), and participants must use this information to adjust their responses. After each sort, participants are asked to provide a judgement of correctness which are aggregated across sorts and used as the indicator of IA. IA on this task is also associated with functioning and improvement on this task would support that iTEST leads to

generalized improvement in IA. This adapted WCST was developed by Koren (2006)[13] and modified by our group[33]. The primary measure is the difference in accuracy judgments in relation to correctness[33]. Notably, IA on the WCST is stable over repeat administrations. Further, were performance on the WCST to improve with repeated exposures, IA is disconnected from performance (see above) and models will adjust for performance.

8.2 SAFETY AND OTHER ASSESSMENTS

Baseline Assessment: At the baseline assessment, we will collect demographic and social contextual information (e.g., living situation) and we will gather *medication data* including sedation/sleepiness side effects via the Karolinska scale[34] at all assessment points. All participants will continue to receive their current treatments, as determined by their clinical providers throughout the study. We will also administer the *Brief Assessment of Cognition in Schizophrenia-App Version (BACS)*[35] and *Virtual Reality Functional Capacity Assessment Tool*[36] *(VRFCAT)* to assess global cognition and functional capacity, respectively, at baseline[37,38]. In some past mobile interventions (with no coaching) adherence was poorer in patients with lower cognitive ability[39], so we will explore neurocognitive ability as a moderator. iTEST is designed to accommodate cognitive impairment by including device/app practice, a simplified guided user interface, and bi-weekly coaching. Finally, we will administer the *Treatment Rationale Scale*[40] to evaluate expectancies. Participants will receive 20 dollars for completing the screening portion of the baseline assessment and additional 30 dollars for completing the second part of the visit if eligible for a total of 50 dollars.

***Symptom Assessments*:** Depression and psychotic symptoms will be assessed with the Calgary Depression Scale for Schizophrenia (CDSS)[41] and Positive and Negative Syndrome Scale (PANSS[26]).

Table 3. Measures, Constructs, and Indicators

Construct	Measure	Operational Indicator
Sample Characterization/Exploratory Moderators		
Cognition	BACS (Baseline only)	Total Score
Functional Capacity	VRFCAT (Baseline only)	Total Score
Psychotic Symptoms	PANSS Positive and Negative	Total Score
Depressive Symptoms	Calgary Depression Scale (CDSS)	Total Score
Treatment Expectancies	Treatment Rationale Scale	Total Score
Trained Task-Based IA Target (R61,R33)		
Trained Task-based IA	IA on METER and VLMT	Estimated vs Actual correct
Transfer of Training Target (R61, R33)		
Untrained Task-based IA	Metacognitive WCST	Global resolution score
Primary Outcome (R33; Exploratory in R61)		
Functional Outcome	Informant-Rated SLOF	Total Score
Secondary Outcomes		
Life space Mobility	GPS	Daily distance from home
Exploratory Mechanisms		
Strategy Use; Motivation and Defeatist Beliefs	Compensatory Cognitive Strategies Scale; CAINS Motivation, Defeatist Performance Beliefs	Total scores
Acceptability and Side Effects		
Intervention Satisfaction	Client Satisfaction Questionnaire (CSQ)	Total Score
User Engagement	Trials and Coaching Sessions Completed	% of Total Possible
Adverse Reactions	Negative Effects Questionnaire (NEQ)	Total Score

***Primary Outcome*:** The informant-rated *Specific Level of Function (SLOF)* scale will be used to measure functional outcome[15]. The SLOF is a 43-item interview administered as informant- or self-report. The scale assesses the participant's current functioning integrated into a single higher-order factor[16] integrating social, every day, and vocational activities. The SLOF emerged as the most reliable and externally valid scale-based measure of functioning in the team's prior validation study[17]. Informants ratings on the SLOF were also found to be sensitive to short term functional gains (12-24 weeks) in a treatment study conducted by Dr. Harvey[18]. We will also administer the self-rated version of the SLOF and calculate self-informant discrepancies in functioning. Greater discrepancies predicted worse functioning , and we will explore whether IA improvements on tasks result in increased convergence between self- and informant SLOF ratings. We considered additional EMA of function[42], but decided not to due to the additional participant burden. See below for procedures related to informants. For completing all study visits and assessments, participants can earn up to $340.

***Exploratory Functional Outcome*:** As an additional objective digital biomarker of mobility pertinent to function (see[43,44]), we will measure GPS lifespace, by way of Neurologger, an open-source application that records coordinates at a data rate of 1 sample/minute, yielding approximately 13K data points per person per week. Accuracy tolerance will be set at +/- 100 feet. Note that participants can opt out of this exploratory measure and still participate in the trial. As in our prior reports in which mobility was associated with community function[43,44], outcomes are median daily distance travelled from home and % of samples at home.

***Exploratory Mechanistic Targets Linking Change in IA to Outcome*:** We will explore whether iTEST increases the use of strategies with the *Compensatory Cognitive Strategies Scale* (completed by informants and participants)[45] as well as motivation measured with *Clinical Assessment Interview for Negative Symptoms (CAINS) Motivation and Pleasure (MAP) subscale*[46]. We will also explore Defeatist Performance Beliefs (DPAS)[47], which predict effort[48] and may improve with more accurate self-awareness of abilities.

Informant Ratings: We have attached to this protocol a standard script for recruitment of informants nominated by the participant. The script details the nature of the study and what is asked of the informant. The script is intended to be read over the phone to the participant and if willing, the questionnaires for informants (SLOF; Compensatory Cognitive Strategies Scale) are administered over the phone immediately thereafter (15-20 minutes). Informants then provide an address where a pay card can be sent. Informants are recruited at baseline and at follow up 16 weeks and receive 15 dollars per occasion for a total of 30 dollars.

Intervention Satisfaction: An adapted *Client Satisfaction Questionnaire*[49] keyed to both coaching sessions and mobile intervention components

For safety, we will administer the *Columbia Suicide Severity Rating Scale*[50]**. See suicide safety protocol.** Since change in IA could theoretically lead to adverse experiences in some individuals (e.g., *greater awareness>depressed mood*), we will administer the *Negative Effects Questionnaire (NEQ*[51]*)* at follow-ups. These are descriptive in nature.

We will also measure Clinical Worsening, which we define a criterion of a score of 96 on the PANSS Total Score or a 30% worsening from previous assessment. These criteria are linked to clinical global impression markers and are standard in acceptance. We will use the same approach as the Suicide

Safety Protocol to response to evidence or worsening as below:

***Monitoring and Responding to Suicidal Ideation* or Clinical Worsening During Lab-Visits** Our safety monitoring protocol is standardized and details these procedures, how they are recorded, and what the staff, facility, and investigator responsibilities are. Whereupon participants endorse active suicidal ideation on the *Columbia Suicide Severity Rating Scale* (CSSRS ≥ 2; up to past 1 month) or recent suicidal behavior (past up to past 3 months), the investigators at the corresponding site will be contacted. Additionally, a score exceeding 96 on the PANSS total (severe symptoms) or 30% increase in symptoms from prior would also lead to contacting the investigator. Research staff will document recent treatment contacts and the resources available to participant, including whether or not their clinical providers are aware of the participants current status. The investigator (licensed provider) or delegated licensed provider will perform an in-person evaluation to determine if immediate emergency services are necessary. Also, at each study visit and particularly during our final assessment with participants we will provide resources lists and support linkage to available services, crisis lines, and other resources. These efforts will be critical in the final study visit in order to mitigate against any potential loss of benefit of contact with research staff that might be experienced after the conclusion of study participation. Participants who have completed the protocol and who are no longer in the study will be allowed to and encouraged to reach out to study investigators in case of questions about community resources pertinent to mental health care or suicide prevention. These plans are developed to be in full accordance with the NIMH's Guidance on Conducting Research with Persons at Elevated Risk for Suicide.

Monitoring or Responding to Suicidal Ideation during Remote or Mobile Device Interactions: We have taken several steps to ensure participant safety during the periods of remote assessment and interventions. All participants will be provided with staff contact information and emergency resources in a printed manual, and they will be instructed that they may call research staff at any time in case of questions or emergencies. Participants must provide release of information for our staff to contact their provider or key informant, so any emergent crises will be relayed to that participant's selected provider. Tile "shortcuts" are provided on the device home screen to resources (e.g., 24-hour Crisis Lines available in the region and to study staff). We, reporting to the Data and Safety Monitoring Board, will monitor these protocols during the course of the study and will make adjustments to study procedures should problems arise. We note that we are not directly querying about suicidal ideation or behavior on the device and so data elements entered by participants on the device are not specially linked to clinical responses. However, we will contact participants by telephone in case of lack of response, remote telephone coaching or check-in interactions with the study therapist will occur weekly, and we will conduct evaluations regarding potential crises upon contacting participants by phone using the same steps as outlined above in in-person contacts.

8.3 ADVERSE EVENTS AND SERIOUS ADVERSE EVENTS

Leave blank. Text should be included under the relevant subheadings below.

8.3.1 DEFINITION OF ADVERSE EVENTS (AE)

The FDA definition of an Adverse event is any untoward medical occurrence associated with the use of an intervention in humans, whether or not considered intervention-related (21 CFR 312.32 (a)).

8.3.2 DEFINITION OF SERIOUS ADVERSE EVENTS (SAE)

For this study, the following standard AE definitions are used:

Adverse event: Any unfavorable and unintended sign, symptom or disease temporally associated with the use of a medical treatment or procedure, regardless of whether it is considered related to the medical treatment or procedure.

Serious Adverse Event: Any AE that results in any of the following outcomes:

- Death
- Life-threatening event requiring inpatient hospitalization or prolongation of existing hospitalization
- Persistent or significant disability/incapacity

8.3.3 CLASSIFICATION OF AN ADVERSE EVENT

Leave blank. Text should be included under the relevant subheadings below.

8.3.3.1 SEVERITY OF EVENT

All AEs will be assessed by the study clinician using a protocol defined grading system.

For safety, we will administer the *Columbia Suicide Severity Rating Scale*[50] (see *Suicide Protocol*). Since change in IA could theoretically lead to adverse experiences in some individuals (e.g., *greater awareness>depressed mood*), we will administer the *Negative Effects Questionnaire (NEQ*[51]*)* at follow-ups. These are descriptive in nature.

For adverse events (AEs) not included in the protocol defined grading system, the following guidelines will be used to describe severity.

- **Mild** – Events require minimal or no treatment and do not interfere with the participant's daily activities.
- **Moderate** – Events result in a low level of inconvenience or concern with the therapeutic measures. Moderate events may cause some interference with functioning.
- **Severe** – Events interrupt a participant's usual daily activity and may require systemic drug therapy or other treatment. Severe events are usually potentially life-threatening or incapacitating. Of note, the term "severe" does not necessarily equate to "serious".]

8.3.3.2 RELATIONSHIP TO STUDY INTERVENTION

All adverse events (AEs) must have their relationship to study intervention assessed by the clinician who examines and evaluates the participant based on temporal relationship and his/her clinical judgment. The degree of certainty about causality will be graded using the categories below. In a clinical trial, the study product must always be suspect.

- **Related** – The AE is known to occur with the study intervention, there is a reasonable possibility that the study intervention caused the AE, or there is a temporal relationship between the study intervention and event. Reasonable possibility means that there is evidence to suggest a causal relationship between the study intervention and the AE.
- **Not Related** – There is not a reasonable possibility that the administration of the study intervention caused the event, there is no temporal relationship between the study intervention and event onset, or an alternate etiology has been established.

8.3.3.3 EXPECTEDNESS

The PI will be responsible for determining whether an adverse event (AE) is expected or unexpected. An AE will be considered unexpected if the nature, severity, or frequency of the event is not consistent with the risk information previously described for the study intervention

8.3.4 TIME PERIOD AND FREQUENCY FOR EVENT ASSESSMENT AND FOLLOW-UP

The occurrence of an adverse event (AE) or serious adverse event (SAE) may come to the attention of study personnel during study visits and interviews of a study participant.
All AEs including local and systemic reactions not meeting the criteria for SAEs will be captured on the appropriate case report form (CRF). Information to be collected includes event description, time of onset, clinician's assessment of severity, relationship to study product (assessed only by those with the training and authority to make a diagnosis), and time of resolution/stabilization of the event. All AEs occurring while on study must be documented appropriately regardless of relationship. All AEs will be followed to adequate resolution.

Changes in the severity of an AE will be documented to allow an assessment of the duration of the event at each level of severity to be performed. AEs characterized as intermittent require documentation of onset and duration of each episode.

Research staff will record all reportable events with start dates occurring any time after informed consent is obtained until 7 (for non-serious AEs) or 30 days (for SAEs) after the last day of study participation. At each study visit, the investigator will inquire about the occurrence of AE/SAEs since the last visit. Events will be followed for outcome information until resolution or stabilization.]

8.3.5 ADVERSE EVENT REPORTING

Adverse Event Definitions and Protocol: Any AE that is reported by the participant is noted by a study team member and will be reported to the PI and to the UCSD (or UTD or UM) Human Research Protection Office (HRPP) per their reporting policy and procedures. Any additional follow-up evaluations required by the event will also be evaluated by the PI and sent to HRPP per their reporting policy as well as to the DSMB. Any participant who develops an SAE will be followed-up through the resolution of the event and for one month thereafter. A log is kept of all AEs in a secure location at UCSD. Trained

research personnel will assess for any serious medical or psychological conditions that may require immediate medical attention or psychiatric treatment. Should attention be required, staff will provide these individuals with information about suitable referrals, including local hospitals, clinics, and/or relevant health-care and mental health professionals, and they the PIs will strongly encourage the individuals to follow up with these referrals. SAEs are reported to the DSMB, NIMH, and the HRPP within 24 hours

8.3.6 SERIOUS ADVERSE EVENT REPORTING

The study PI will immediately report to the sponsor any serious adverse event, whether or not considered study intervention related, including those listed in the protocol or investigator brochure and must include an assessment of whether there is a reasonable possibility that the study intervention caused the event. Study endpoints that are serious adverse events (e.g., all-cause mortality) must be reported in accordance with the protocol unless there is evidence suggesting a causal relationship between the study intervention and the event (e.g., death from anaphylaxis). In that case, the investigator must immediately report the event to the sponsor.

All serious adverse events (SAEs) will be followed until satisfactory resolution or until the site investigator deems the event to be chronic or the participant is stable.

8.3.7 REPORTING EVENTS TO PARTICIPANTS

Not Applicable

8.3.8 EVENTS OF SPECIAL INTEREST

Not Applicable

8.3.9 REPORTING OF PREGNANCY

Not Applicable

8.4 UNANTICIPATED PROBLEMS

Leave blank. Text should be included under the relevant subheadings below.

8.4.1 DEFINITION OF UNANTICIPATED PROBLEMS (UP)

The Office for Human Research Protections (OHRP) considers unanticipated problems involving risks to participants or others to include, in general, any incident, experience, or outcome that meets **all** of the following criteria:

- Unexpected in terms of nature, severity, or frequency given (a) the research procedures that are described in the protocol-related documents, such as the Institutional Review Board (IRB)-

approved research protocol and informed consent document; and (b) the characteristics of the participant population being studied;
- Related or possibly related to participation in the research ("possibly related" means there is a reasonable possibility that the incident, experience, or outcome may have been caused by the procedures involved in the research); and
- Suggests that the research places participants or others at a greater risk of harm (including physical, psychological, economic, or social harm) than was previously known or recognized.

8.4.2 UNANTICIPATED PROBLEM REPORTING

The investigator will report unanticipated problems (UPs) to the reviewing Institutional Review Board (IRB). The UP report will include the following information:

- Protocol identifying information: protocol title and number, PI's name, and the IRB project number;
- A detailed description of the event, incident, experience, or outcome;
- An explanation of the basis for determining that the event, incident, experience, or outcome represents an UP;
- A description of any changes to the protocol or other corrective actions that have been taken or are proposed in response to the UP.

To satisfy the requirement for prompt reporting, UPs will be reported using the following timeline:

- UPs that are serious adverse events (SAEs) will be reported to the IRB and to the DCC/study sponsor within 24 hours of the investigator becoming aware of the event.
- Any other UP will be reported to the IRB within 10 days of the investigator becoming aware of the problem.

8.4.3 REPORTING UNANTICIPATED PROBLEMS TO PARTICIPANTS

Not Applicable

9 STATISTICAL CONSIDERATIONS

9.1 STATISTICAL HYPOTHESES

The statistical hypotheses are that adherence will exceed the criterion of 75% and that there will be a pre-post change in the VLMT, METER and WCST tasks that will exceed that of 0.5 Cohen's d with the baseline sample wide standard deviation serving as the denominator.

9.2 SAMPLE SIZE DETERMINATION

We used the Monte Carlo simulation capabilities of Mplus version 8.4[52] to perform our power analyses. To determine our levels of power to detect changes in IA for the list-learning task or emotion recognition tasks equivalent to d = |0.50| for R61 Aim 1, we specified three-level models with 1000 replications each wherein trials (t = 18) were nested within weeks (w = 8, 12, or 16) that were further nested within persons (n = 60 or 48 in the case of 20% attrition). These models were similar to Figure 1 with the exception that changes in IB and IA at Level 3 were not regressed on intervention condition. As can be seen in Table 1, our power levels were excellent (i.e., > .90) across all combinations of sample size and week size, ranging from .91 for detecting changes in IA of d = -0.50 at 8 weeks with 48 participants to .97 for detecting changes in IA of d = 0.50 at 16 weeks with 60 participants.

To determine our levels of power to detect transfer-of-training for R61 Aim 1 and optimal dosing for Aim 2, we again specified three-level models with 1000 replications each wherein trials (t = 18) were nested within weeks (w = 8, 12, or 16) that were further nested within persons (n = 60 or 48 in the case of 20% attrition). However, we modified these models such that they included an additional Truth and Bias model for the WCST transfer of training target at Level 2 that had a change in IB of d = -0.50 across the weeks. To reflect the fact that this in-laboratory tasks would only be measured at 0, 4, 8, 12, and 16 weeks, we specified that these measurements should have 62.5% missing data in the simulations (e.g., only completing 3 out of 8 weeks). As can be seen in Table 2, we should have more-or-less adequate levels of power to detect a change in IB for the in-lab assessment at 16 weeks with 48 participants (power = .77) and adequate power to detect a change in IB for the in-lab assessment at 12 and 16 weeks with 60 participants (power = .78 and .83, respectively).

To determine our levels of power to detect exploratory predictors of changes in IA for the list-learning and emotion recognition tasks, as well as to detect the effects of changes in IA for the list-learning task and emotion recognition on functional outcomes, we modified the original models for R61 Aim 1 such that we either regressed changes in IA for the mobile tasks task on a predictor, or regressed a functional outcome on changes in IA, respectively.

9.3 POPULATIONS FOR ANALYSES

All analyses will be conducted on Intention-to-Treat (ITT) Analysis Dataset (i.e., all participants)

9.4 STATISTICAL ANALYSES

9.4.1 GENERAL APPROACH

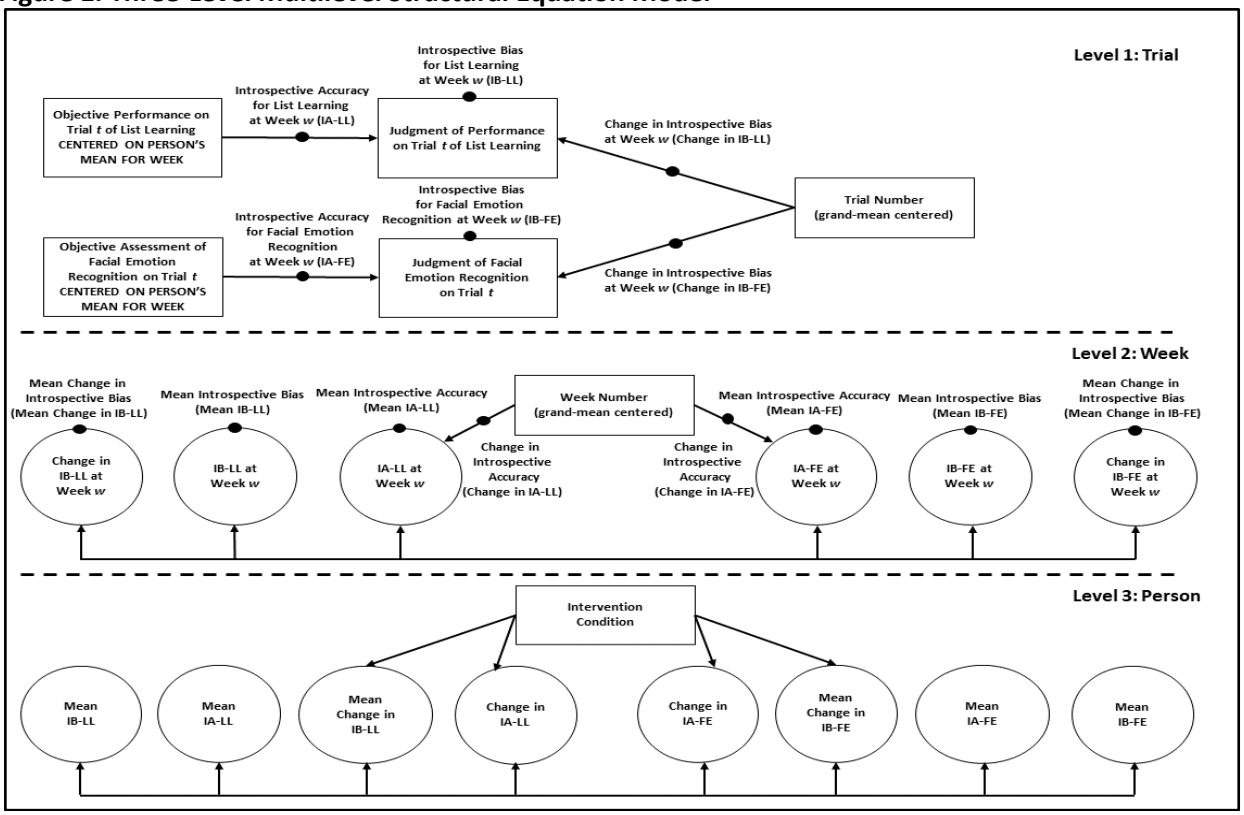
As this is a pilot study, there is no formal SAP

Since our analyses focus on clinically significant magnitude of effects or achievement of thresholds, significance testing is not relevant

We will use Multilevel Structural Equation Modeling (MLSEM) with three levels to estimate changes in levels of Introspective Accuracy (IA) across trials (level 1) and/or weeks (level 2) within persons (level 3). Figure 1 presents an example three-level model wherein changes in IA for the *Variable Length Memory Test* (VLMT) list learning and *Mobile Electronic Test of Emotion Recognition* (METER) emotion recognition are simultaneously estimated. Given the relatively large number of random effects, we will use Bayesian

estimation rather than maximum likelihood estimation for the analyses[53]. Although some task or ability domains will not be measured at every trial or every week, specifying trial as Level 1 (e.g., for IA on mobile task) and week as Level 2 (e.g., for WCST) will still enable us to estimate changes in IA for these domains.

Figure 1. Three-Level Multilevel Structural Equation Model



As shown in Figure 1, at Level 1 (trials within weeks), participants' judgments of their performance on a task at trial *t* on week *w* will be regressed on their objective performance on that same task at trial *t* on week *w* (e.g., participants' judgments of their performance on trial *t* of the list-learning/METER task at week *w* will be regressed on their objective performance on trial *t* of that same list-learning task at week *w*). In line with the Truth and Bias model framework[54] participants' judgments of their performance will be centered on the grand-mean for objectively assessed performance, and their scores for objective performance will be centered on the mean for their objective performance that week. This centering strategy renders the intercept for participants' judgments of their performance to be the average trial-to-trial difference between their judgment of their performance and the grand-mean for objective performance at week *w* (e.g., IB for list learning/METER at week *w*; see IB-LL/FE in Figure 1). Moreover, the slope for this regression captures participants' capacity to accurately track the lows and highs of their performance across the trials for that week (e.g., IA for list learning/METER at week *w*; see IA-LL/FE in Figure 1). To estimate systematic change in IA across the trials (see, e.g., Change in IA-LL/FE in Figure 1), we will regress participants' judgments of their performance at trial *t* on week *w* (which is technically IA given the centering strategy) on trial number (grand-mean centered).

The black dots placed on parameters at Level 1 in Figure 1 indicate the estimation of random effects to permit for the heterogeneity of those Level 1 effects across the weeks in the study for participants. Thus, at Level 2 (weeks within persons), we will estimate week-to-week differences in participants' average

levels of IB across the trials, participants' levels of IA, and participants' average changes in IB across the trials (see, e.g., the circles for IB-LL/FE at week *w*, IA-L/FE at week *w*, and Change in IB-LL/FE at week *w* in Figure 1, respectively). Further, to estimate systematic change in IA across the weeks (see, e.g., Change in IA-LL/FE in Figure 1), we will regress participants' IA at week *w* on week number (grand-mean centered). Because IA is operationalized as a slope (as opposed to a mean-level difference like IB), we estimate change in IA at Level 2 so that reliable estimates of IA can be obtained each week (e.g., 18 trials contribute to IA for list learning). Although not depicted in Figure 1, we will include additional Truth and Bias models at Level 2 for WCST targets of transfer wherein participants' judgments of their abilities assessed at week *w* are regressed on their objective performance for that domain at week *w* (e.g., participants' judgments of how many WCST items they got correct at week *w* will be regressed on their objective performance on the WCST at week *w*). We will use the same centering strategy described previously to render the intercepts and slopes from the regressions as indices of IB and IA, respectively.

The black dots placed on parameters at Level 2 in Figure 1 indicate the estimation of random effects to permit for the heterogeneity of those Level 2 effects across participants in the study. Thus, at Level 3 (persons), we will estimate individual differences in participants' average levels of introspective bias across the trials and weeks, participants' average levels of introspective accuracy across the weeks, participants' average changes in introspective bias across the trials and weeks, and participants' average changes in introspective accuracy across the weeks (see, e.g., the circles for Mean IB-LL/FE, Mean IA-LL/FE, Mean Change in IB-LL/FE, and Change in IA-LL/FE in Figure 1, respectively). We will evaluate R61 Aim 1 by inspecting the standardized mean-level differences for Change in IA-LL/FE in Figure 1 at weeks 8, 12, and 16. Similarly, we will evaluate R61 transfer of training by inspecting the standardized mean-level differences for changes in IB and/or IA for the WCST.

9.4.2 ANALYSIS OF THE PRIMARY EFFICACY ENDPOINT(S)

These are detailed in 9.4.1

9.4.3 ANALYSIS OF THE SECONDARY ENDPOINT(S)

ToTo evaluate the briefest dose of the intervention, we will inspect changes at weeks 8, 12, and 16. We w will evaluate which of these time points reaches the criterion (Cohen's d=0.5) for our primary endpoints.

We will also evaluate whether change in the Specific Level of Functioning (SLOF) scale exceeds d=0.5 improvement. Analyses will follow the same approach as in 9.4.1, employing linear mixed models in the intent to treat sample.

9.4.4 SAFETY ANALYSES

Safety analyses will be descriptive and include rates of ideation according to the Columbia Suicide Severity Rating Scale (CSSR-S) and scores on the Negative Effects Questionnaire. Since this is an open trial, and we do not link these ratings to change in the intervention (e.g. discontinuation) these are for descriptive purposes and for triage in the case of the CSSR-S (see Suicide Safety protocol)

9.4.5 BASELINE DESCRIPTIVE STATISTICS

Not applicable as this is an open single arm trial.

9.4.6 PLANNED INTERIM ANALYSES

Not Applicable

9.4.7 SUB-GROUP ANALYSES

There are no plans to evaluate endpoints by age, sex, race/ethnicity, or other demographic characteristic(s). This is a pilot study and therefore the study is underpowered to detect an effect variation by subgroup.

9.4.8 TABULATION OF INDIVIDUAL PARTICIPANT DATA

Individual participant data will not be listed by measure and time point.

9.4.9 EXPLORATORY ANALYSES

Not Applicable

10 SUPPORTING DOCUMENTATION AND OPERATIONAL CONSIDERATIONS

10.1 REGULATORY, ETHICAL, AND STUDY OVERSIGHT CONSIDERATIONS

Leave blank. Text should be included under the relevant subheadings below.

10.1.1 STUDY DISCONTINUATION AND CLOSURE

The only foreseeable reasons for termination/suspension would be if an unexpected significant risk was revealed. As this is a pilot study other criteria (futility, etc.) are not relevant.

Circumstances that may warrant termination or suspension include, but are not limited to:

- Determination of unexpected, significant, or unacceptable risk to participants

Study may resume once concerns about safety, protocol compliance, and data quality are addressed, and satisfy the sponsor, IRB and/or Food and Drug Administration (FDA).

10.1.2 CONFIDENTIALITY AND PRIVACY

Participant confidentiality and privacy is strictly held in trust by the participating investigators, their staff, and the sponsor(s) and their interventions. This confidentiality is extended to cover testing of biological samples and genetic tests in addition to the clinical information relating to participants. Therefore, the study protocol, documentation, data, and all other information generated will be held in

strict confidence. No information concerning the study or the data will be released to any unauthorized third party without prior written approval of the sponsor.

All research activities will be conducted in as private a setting as possible.

The NIH and the Institutional Review Board (IRB), may inspect all documents and records required to be maintained by the investigator, including but not limited to, medical records (office, clinic, or hospital) for the participants in this study. The clinical study site will permit access to such records.

The study participant's contact information will be securely stored at each clinical site for internal use during the study. At the end of the study, all records will continue to be kept in a secure location for as long a period as dictated by the reviewing IRB, Institutional policies, or sponsor requirements.

Study participant research data, which is for purposes of statistical analysis and scientific reporting, will be transmitted to and stored at the National Data Archive. This will not include the participant's contact or identifying information. Rather, individual participants and their research data will be identified by a unique study identification number called a GUID. The study data entry and study management systems used by clinical sites and staff will be secured and password protected. At the end of the study, all study databases will be de-identified and archived at the National Data Archive.

Certificate of Confidentiality

To further protect the privacy of study participants, a Certificate of Confidentiality will be issued by the National Institutes of Health (NIH). This certificate protects identifiable research information from forced disclosure. It allows the investigator and others who have access to research records to refuse to disclose identifying information on research participation in any civil, criminal, administrative, legislative, or other proceeding, whether at the federal, state, or local level. By protecting researchers and institutions from being compelled to disclose information that would identify research participants, Certificates of Confidentiality help achieve the research objectives and promote participation in studies by helping assure confidentiality and privacy to participants.

Confidentiality for In-Lab Assessments: In order to protect confidentiality all participant data will be de-identified by providing each participant a unique ID in computer files and all physical files from the study will be kept in a locked cabinet in our research offices. All electronic data and files will be stored on a secure server also housed at UCSD. Only study investigators and trained research personnel linked to this study and registered with the IRB will have access to this data. All documentation will identify participants by their Unique ID number and not other personally identifying information.

Confidentiality and Privacy Protections for Mobile Device Data: Mobile health interactions are transmitted in an encrypted fashion and devices used in this research will not contain or transmit personally identifying information with study data. Since the study platform uses a web-based data collection engine, there is no data that is stored on the device. We will use HIPAA compliant Amazon Web Services (AWS) to schedule the iTEST skills training, surveys and for data usage. AWS offers a simple and intuitive web-based user interface from user to server. It is an easy-to-use cloud-computing platform that allows various storage services for backup, archiving, and disaster recovery use cases. AWS enables robust safeguards to help protect the privacy of the data with highly secure AWS Data centers.

Amazon S3 provides scalable and durable object storage in the cloud. S3 is where the experimenters will be storing the user information (study ID number, phone number, start and end time) along with the survey links. Once the data is generated, the data will be stored securely in this server, which will only be accessible by study staff. Study staff will utilize queries in Amazon's Athena to extract key variables of interest in the mobile cognitive tests. We will also be using Postman to send the surveys to our users in the form of text messages. Postman is a popular API client that makes creating, sharing, and documenting APIs simpler. This is done by creating and saving HTTPs requests via python. A group of saved requests can be organized into folders or "collections". From this collection, experimenters are able to adjust the code to send the survey links to certain participants at a specific start date and end date to their mobile numbers. By saving these API requests, we will not have to remember the endpoint of the API, API key, etc. – as such the information is not connected to PII. All PIs and study staff will have access to these platforms and will ensure that the data is secure.

Finally, it is possible that some participants may store data (e.g., notes) on the device provided to them if they opt not to use their own device or do not have one. Individuals are informed of this risk, and should the participants lose the study device, we have taken the additional step of incorporating the Autowipe application on all study deices which enables remote factory reset of the device, erasing all data. An additional safeguard for privacy loss is the use of a homescreen "lock" such that participants must enter a brief code to operate the equipment. If participants lose or damage study device, they will not be held responsible for the cost of the device or any other costs.

10.1.3 FUTURE USE OF STORED SPECIMENS AND DATA

It is a requirement of the sponsor that all data collected that is capable of de-identification (i.e., excluding PII, video or audio) is shared with the National Data Archive. Participants are informed of this in the Informed Consent Process.

In the event that funding for this application is granted, a resource sharing plan has been developed that will ensure that the findings and implications of this research will be available to the public in a timely and understandable fashion.

- We will deposit and share data in the NIMH National Data Archive and we have budgeted staff time in our data management team 100 hours per year to accomplish the preparation, curation, and uploading of semi-annual de-identified data as the study progresses (linked with Global Unique Identifiers) to the online system in collaboration with NIH staff.
- We will develop documents and tools to support this dissemination, in particular for the sharing ecological momentary assessment data which presents data management challenges in its volume and complexity, including within-person aggregation scripts that can be used to analyze the data at the day and week level.

10.1.4 KEY ROLES AND STUDY GOVERNANCE

The primary governing body of this project will be a Steering Committee, which will be co-chaired by the PIs (Depp and Pinkham) and comprised of the PIs and co-investigators (Moore, Harvey, Ackerman) and

consultant (Bowie). The Steering Committee will convene via teleconference initially to hold weekly study management calls with all key personnel, the project manager and all staff during the start-up period, tapering off to monthly calls with only the Steering Committee after therapist training is completed, rater reliability is established, and enrollment is initiated. Sub-committees for the project will include bi-weekly separate Coordinator meeting, for training and monitoring, led by Dr. Pinkham and weekly Therapist meeting, supervision, and fidelity monitoring, led by Dr. Depp. The team reports to the Data Safety Monitoring board twice in the first year of the study and then annually thereafter. The schedule of meetings may be increased based on the requirements of the scientific management of this study.

Roles of Individual Team Members:

Colin Depp, Principal Investigator will in partnership with Dr. Pinkham will and oversee all aspects of the project, including budget responsibility and clinical trial protocol deployment. He will have primary responsibility for designing, implementing and monitoring the iTEST blended intervention, including a) coordinating and troubleshooting mobile intervention components, b) training therapists in delivering the intervention and using clinical dashboards, and c) establishing the fidelity monitoring protocol and providing supervision of therapists. He along with Dr. Pinkham will be responsible for communicating with the NIMH (See MPI plan for justification for MPI framework).

Amy Pinkham, MPI, will coordinate recruitment and retention at UT Dallas (R61), and she will be the lead investigator in a) training in, assessing reliability, and implementing study protocol to sustain rigor in masked assessments in the clinical trial, and b) developing and minting data sharing, outcome, and safety recording systems for the clinical trial, including adverse events. Dr. Depp and Dr. Pinkham's roles are complementary and distinct for the benefit of rigor, aligned with Dr. Depp's supervision of interventionist staff and data recording involved intervention fidelity and deployment, and Dr. Pinkham's supervision of staff engaged in recruitment and assessment who are masked to randomization.

Philip Harvey, Co-Investigator, is a Professor at the University of Miami and he will participate in analyses of de-identified data, work with Drs Pinkham and Depp in resolving procedural issues with the iTEST intervention, and collaborate with the team on next steps for the project.

Raeanne Moore, Co-Investigator, will lead the deployment of the mobile cognitive assessment components, training research assistants in technical aspects of data collection, generating best practices for data transfer and secure storage, preparation of data for sharing, and integration and synchronization of data streams. Dr. Moore will be primarily responsible for managing study workflows with our software partner to ensure that the proposed patient-facing, program-facing, and data repository elements, respectively, are rigorously deployed and efficiently packaged for Data Sharing.

Robert Ackerman, Co-Investigator, will serve as the lead statistician and will work in collaboration with the PIs on the respective study components to carry out the proposed study. He will oversee responsibilities and tasks performed in data management by RAs, and he will conduct interim and outcome analyses for the R61 and R33 for both regulatory (e.g., DSMB) and scientific reporting.

Christopher Bowie, Consultant, will provide consultation as an expert in the implementation, evaluation, and dissemination of cognitive rehabilitation for psychotic disorders. This includes both

supporting the developmental activities in the R61 phase, refinement of the intervention, and then deployment and outcome interpretation in the R33 phase.

Research Coordinators, UCSD, UTD A full-time Research Coordinator at each site (study-wide 2 total for R61) will be responsible for recruitment, consenting and enrollment, data uploading, and delivering rigorous masked assessments of study outcomes. Dr. Pinkham will supervise the group of Coordinators to ensure cross-site standardization of procedures.

iTEST Interventionist, UCSD, UTD Part-time study therapists at each of the three sites will work under the clinical supervision of Dr. Depp and will be responsible for delivery and documentation of the study intervention and recording for fidelity monitoring purposes. Therapists will have at least a bachelor's or master's level of training, and must have experience with provision of training/learning based interventions in people with serious mental illnesses. However, they are not doctoral level practitioners and so will be similar in experience to that of community mental health workers.

10.1.5 SAFETY OVERSIGHT

Data Safety Monitoring Board (DSMB) Description: The study will have an DSMB. The PIs will monitor and collect information on adverse events (AEs), serious adverse events (SAEs), and/or unanticipated problems. A Data Safety Monitoring Board (DSMB) comprised of independent members will monitor study safety. We will use an established DSMB that monitors clinical studies in our laboratory. The DSMB consists of three members who are external to UCSD, UTD or UM, chaired by a leading expert in clinical trials in serious mental illness, a biostatistician and clinical trials methodologist, and a director of a mental health advocacy organization.

DSMB Provisional Charter: The Board will 1) Meet semi-annually in the first year, and then annually via teleconference (and email correspondence as needed) to first comment on the protocol prior to any enrollment, and finalize the DSMB Charter. Subsequent meetings will then review reports on protocol progress and adverse events; 2) Review developing outcome data. The PIs will provide subject data tables for primary target and outcome variables, including reasons for missing data, forms completion, and review data Quality Control/Quality Assurance procedures at annual meetings. In addition summary responses to the *Negative Experiences Questionnaire,* a measure of side effects in digital health, will be included for both conditions; 3) Review AEs/SAEs according to the convention detailed below; 4) Review participant enrollment progress and dropouts: PI will provide current CONSORT diagram with reasons for dropouts/losses to follow-up and attached dropout table at annual meetings. The DSMB will make recommendations based on review of 1-4 above, regarding appropriate protocol, operational, or other changes, and including reporting deviations or other AEs/SAEs/Unexpected problems related to the study protocol to the appropriate agency/institute (IRB, NIH). The PI will ensure that the IRB is notified of any AEs/SAEs/Unexpected problems, and that the NIH is informed of actions, if any, taken by the DSMB or IRB as a result of its continuing review. The PI will generate minutes of DSMB meetings, including a summary of any discussion of AEs/SAEs or Unexpected problems, and enrollment/dropouts, which will be reviewed and approved by the DSMB Chair, and these minutes will be transmitted to the IRB by the PI or Sub-PI at each site at the time of continuing annual IRB review of the protocol.

We note that we will provide the finalized DSMB charter once approved by the DSMB.

10.1.6 CLINICAL MONITORING

Not Applicable

10.1.7 DATA HANDLING AND RECORD KEEPING

Leave blank. Text should be included under the relevant subheadings below.

10.1.7.1 DATA COLLECTION AND MANAGEMENT RESPONSIBILITIES

Data collection is the responsibility of the clinical trial staff at the site under the supervision of the site investigator. The investigator is responsible for ensuring the accuracy, completeness, legibility, and timeliness of the data reported.

All source documents should be completed in a neat, legible manner to ensure accurate interpretation of data.

Hardcopies of the study visit worksheets will be provided for use as source document worksheets for recording data for each participant enrolled in the study.

Clinical data (including adverse events (AEs), concomitant medications, and expected adverse reactions data) and clinical laboratory data will be entered into Redcap a 21 CFR Part 11-compliant data capture system. The data system includes password protection and internal quality checks, such as automatic range checks, to identify data that appear inconsistent, incomplete, or inaccurate.

10.1.7.2 STUDY RECORDS RETENTION

Study documents will be retained a minimum of three years from the date of Federal Financial Report (FFR) submission. The NIH does not specify beyond this when records can be destroyed.

10.1.8 PROTOCOL DEVIATIONS

A protocol deviation is any noncompliance with the clinical trial protocol, International Conference on Harmonisation Good Clinical Practice (ICH GCP), or Manual of Procedures (MOP) requirements. The noncompliance may be either on the part of the participant, the investigator, or the study site staff. As a result of deviations, corrective actions are to be developed by the site and implemented promptly.

These practices are consistent with ICH GCP:

- 4.5 Compliance with Protocol, sections 4.5.1, 4.5.2, and 4.5.3
- 5.1 Quality Assurance and Quality Control, section 5.1.1
- 5.20 Noncompliance, sections 5.20.1, and 5.20.2.

It is the responsibility of the site investigator to use continuous vigilance to identify and report deviations within 2 working days of identification of the protocol deviation, or within 1 working days of the scheduled protocol-required activity. All deviations must be addressed in study source documents, reported to NIMH Program Official. Protocol deviations must be sent to the reviewing Institutional Review Board (IRB) per their policies. The site investigator is responsible for knowing and adhering to

the reviewing IRB requirements. Further details about the handling of protocol deviations will be included in the MOP

We will maintain appropriate medical and research records for this trial, in compliance with ICH GCP and regulatory and institutional requirements for the protection of confidentiality of participants.

10.2 ADDITIONAL CONSIDERATIONS

Not Applicable

10.3 ABBREVIATIONS

AE	Adverse Event
BACS	Brief Assessment of Cognition in Schizophrenia-App Version
C-SSRS	Columbia Suicide Severity Scale
CDSS	Calgary Depression Scale
CONSORT	Consolidated Standards of Reporting
CRF	Case Report Form
CSQ	Client Satisfaction Questionnaire
DCC	Data Coordinating Center
DPAS	Defeatist Performance Beliefs
DSMB	Data and Safety Monitoring Board
EMA	Ecological Momentary Assessment
FDA	Food and Drug Administration
FFR	Federal Financial Report
GPS	Global Positioning System
GUID	Globally Unique Identifier
HRPP	Human Research Protection Office
IA	Introspective Accuracy
IC	Informed Consent
ICH GP	International Conference on Harmonisation Good Clinical Practice
IRB	Institutional Review Board
iTEST	Improved Thinking through Everyday Self-Assessment Training
ITT	Intention-to-Treat
MAP	Motivation and Pleasure subscale
METER	Mobile Electronic Test of Emotional Recognition
MLSEM	Multilevel Structural Equation Modeling
MOP	Manual of Procedures
MPI	Multiple Principal Investigator
NDA	National Data Archive
NEQ	Negative Effects Questionnaire
NEQ	Negative Effects Questionnaire
NIH	National Institutes of Health
NIMH	National Institute of Mental Health
OHRP	Office for Human Research Protections
PANSS	Positive and Negative Syndrome Scale
PI	Principal Investigator
PII	Personal Identifiable Information
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SCID-5	Structured Clinical Interview for DSM-5 Disorders 5
SLOF	Specific Levels of Functioning Scale
SOA	Schedules of Activities
UBACC	UCSD Brief Assessment of Capacity to Consent
UCSD	University of California, San Diego
UM	University of Miami
UP	Unanticipated Problems
US	United States
UTD	University of Texas at Dallas

VLMT	Mobile Variable Difficulty List Memory Test
VRFCAT	Virtual Reality Functional Capacity Assessment Tool
WCST	Wisconsin Card Sorting Task
WRAT-4	Wide Range Achievement Test-4

10.4 PROTOCOL AMENDMENT HISTORY

The table below is intended to capture changes of IRB-approved versions of the protocol, including a description of the change and rationale. A Summary of Changes table for the current amendment is located in the Protocol Title Page.

Version	Date	Description of Change	Brief Rationale

11 REFERENCES

1. Charlson FJ, Ferrari AJ, Santomauro DF, et al. Global epidemiology and burden of schizophrenia: findings from the global burden of disease study 2016. *Schizophrenia bulletin*. 2018;44(6):1195-1203.

2. Harvey PD, McGurk SR, Mahncke H, Wykes T. Controversies in computerized cognitive training. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*. 2018;3(11):907-915.

3. Prikken M, Konings MJ, Lei WU, Begemann MJ, Sommer IE. The efficacy of computerized cognitive drill and practice training for patients with a schizophrenia-spectrum disorder: a meta-analysis. *Schizophrenia research*. 2019;204:368-374.

4. Bowie CR, Depp C, McGrath JA, et al. Prediction of real-world functional disability in chronic mental disorders: a comparison of schizophrenia and bipolar disorder. *American Journal of Psychiatry*. 2010;167(9):1116-1124.

5. Silberstein J, Harvey PD. Impaired introspective accuracy in schizophrenia: an independent predictor of functional outcomes. *Cognitive neuropsychiatry*. 2019;24(1):28-39.

6. Harvey PD, Pinkham A. Impaired self-assessment in schizophrenia: Why patients misjudge their cognition and functioning. *Current Psychiatry*. 2015;14(4):53-59.

7. Gould F, McGuire LS, Durand D, et al. Self-assessment in schizophrenia: Accuracy of evaluation of cognition and everyday functioning. *Neuropsychology*. 2015;29(5):675-682. doi:10.1037/neu0000175

8. Bowie CR, Bell MD, Fiszdon JM, et al. Cognitive remediation for schizophrenia: An expert working group white paper on core techniques. *Schizophrenia research*. 2020;215:49-53.

9. Parrish EM, Kamarsu S, Harvey PD, Pinkham A, Depp CA, Moore RC. Remote Ecological Momentary Testing of Learning and Memory in Adults With Serious Mental Illness. *Schizophrenia Bulletin*. 2020;

10. Depp CA, Kamarsu S, Filip TF, et al. Ecological momentary facial emotion recognition in psychotic disorders. *Psychol Med*. Jan 12 2021:1-9. doi:10.1017/s0033291720004419

11. Morgan O, Strassnig MT, Moore RC, et al. Accuracy of immediate self-assessment of neurocognitive test performance: Associations with psychiatric diagnosis and longitudinal psychotic symptoms. *J Psychiatr Res*. Dec 2022;156:594-601. doi:10.1016/j.jpsychires.2022.10.069

12. Posner K, Brent D, Lucas C, et al. Columbia-suicide severity rating scale (C-SSRS). *New York, NY: Columbia University Medical Center*. 2008;10

13. Koren D, Seidman LJ, Goldsmith M, Harvey PD. Real-world cognitive—and metacognitive—dysfunction in schizophrenia: a new approach for measuring (and remediating) more “right stuff”. *Schizophrenia Bulletin*. 2006;32(2):310-326.

14. Carpenter J, Sherman MT, Kievit RA, Seth AK, Lau H, Fleming SM. Domain-general enhancements of metacognitive ability through adaptive training. *Journal of Experimental Psychology: General*. 2019;148(1):51.

15. Schneider LC, Struening EL. SLOF: a behavioral rating scale for assessing the mentally ill. Oxford University Press; 1983:9-21.

16. Bowie CR, Best MW, Depp C, et al. Cognitive and functional deficits in bipolar disorder and schizophrenia as a function of the presence and history of psychosis. *Bipolar disorders*. 2018;20(7):604-613.

17. Harvey PD, Raykov T, Twamley EW, Vella L, Heaton RK, Patterson TL. Validating the measurement of real-world functional outcomes: phase I results of the VALERO study. *American Journal of Psychiatry*. 2011;168(11):1195-1201.

18. Bowie CR, Grossman M, Gupta M, Oyewumi LK, Harvey PD. Cognitive remediation in schizophrenia: efficacy and effectiveness in patients with early versus long-term course of illness. *Early Intervention in Psychiatry*. 2014;8(1):32-38. doi:*https://doi.org/10.1111/eip.12029*

19. Bonnechère B, Langley C, Sahakian BJ. The use of commercial computerised cognitive games in older adults: a meta-analysis. *Scientific Reports*. 2020/09/17 2020;10(1):15276. doi:10.1038/s41598-020-72281-3

20. Cella M, Preti A, Edwards C, Dow T, Wykes T. Cognitive remediation for negative symptoms of schizophrenia: a network meta-analysis. *Clinical psychology review*. 2017;52:43-51.

21. Depp C, Ehret B, Villa J, Perivoliotis D, Granholm E. A brief mobile-augmented suicide prevention intervention for people with psychotic disorders in transition from acute to ongoing care: protocol for a pilot trial. *JMIR research protocols*. 2021;10(2):e14378.

22. Depp CA, Ceglowski J, Wang VC, et al. Augmenting psychoeducation with a mobile intervention for bipolar disorder: a randomized controlled trial. *Journal of affective disorders*. 2015;174:23-30.

23. Depp CA, Perivoliotis D, Holden J, Dorr J, Granholm EL. Single-session mobile-augmented intervention in serious mental illness: a three-arm randomized controlled trial. *Schizophrenia Bulletin*. 2019;45(4):752-762.

24. Jeste DV, Palmer BW, Appelbaum PS, et al. A New Brief Instrument for Assessing Decisional Capacity for Clinical Research. *Archives of General Psychiatry*. 2007;64(8):966-974. doi:10.1001/archpsyc.64.8.966

25. Wilkinson GS, Robertson GJ, Psychological Assessment Resources I. *WRAT 4 : wide range achievement test ; professional manual*. Psychological Assessment Resources, Inc.; 2006.

26. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull*. 1987;13(2):261-76. doi:10.1093/schbul/13.2.261

27. Tercero BA, Perez MM, Mohsin N, et al. Using a Meta-cognitive Wisconsin Card Sorting Test to measure introspective accuracy and biases in schizophrenia and bipolar disorder. *J Psychiatr Res*. Aug 2021;140:436-442. doi:10.1016/j.jpsychires.2021.06.016

28. Badal VD, Depp CA, Moore R, Ackerman RA, Harvey PD, Pinkham A. Confidence, Accuracy Judgments and Feedback in Schizophrenia and Bipolar Disorder: A Time Series Network Analysis *Psychol Med*. In Press;

29. Granholm E, Holden J, Dwyer K, Mikhael T, Link P, Depp C. Mobile-assisted cognitive behavioral therapy for negative symptoms: open single-arm trial with schizophrenia patients. *JMIR mental health*. 2020;7(12):e24406.

30. Bowie CR, Grossman M, Gupta M, Holshausen K, Best MW. Action-based cognitive remediation for individuals with serious mental illnesses: Effects of real-world simulations and goal setting on functional and vocational outcomes. *Psychiatric Rehabilitation Journal*. 2017;40(1):53.

31. Haddock G, Devane S, Bradshaw T, et al. An investigation into the psychometric properties of the Cognitive Therapy Scale for Psychosis (Cts-Psy). *Behavioural and Cognitive Psychotherapy*. Apr 2001;29(2):221-233. doi:10.1017/S1352465801002089

32. Jones SE, Moore RC, Pinkham AE, Depp C, Harvey PD. A cross-diagnostic study of Adherence to Ecological Momentary Assessment: Comparisons across study length and daily survey frequency find that early adherence is a potent predictor of study-long adherence. Under review;

33. Tercero BA, Perez MM, Mohsin N, et al. Using a meta-cognitive Wisconsin card sorting teset to measure introspective accurcy in schizophreina and bipolar diosrder. *Journal of psychiatric research*. In Press;

34. Shahid A, Wilkinson K, Marcu S, Shapiro CM. Karolinska sleepiness scale (KSS). *STOP, THAT and one hundred other sleep scales*. Springer; 2011:209-210.

35. Keefe RS, Harvey PD, Goldberg TE, et al. Norms and standardization of the Brief Assessment of Cognition in Schizophrenia (BACS). *Schizophrenia research*. 2008;102(1-3):108-115.

36. Keefe RSE, Davis VG, Atkins AS, et al. Validation of a Computerized test of Functional Capacity. *Schizophr Res*. Aug 2016;175(1-3):90-96. doi:10.1016/j.schres.2016.03.038

37. Nuechterlein KH, Green MF, Kern RS, et al. The MATRICS Consensus Cognitive Battery, part 1: test selection, reliability, and validity. Comparative Study

Multicenter Study

Research Support, Non-U.S. Gov't

Research Support, U.S. Gov't, P.H.S. *The American Journal of Psychiatry*. Feb 2008;165(2):203-13. doi:10.1176/appi.ajp.2007.07010042

38. Patterson TL, Goldman S, McKibbin CL, Hughs T, Jeste DV. UCSD Performance-Based Skills Assessment: development of a new measure of everyday functioning for severely mentally ill adults. *Schizophr Bull*. 2001;27(2):235-45. doi:10.1093/oxfordjournals.schbul.a006870

39. Granholm E, Ben-Zeev D, Link PC, Bradshaw KR, Holden JL. Mobile Assessment and Treatment for Schizophrenia (MATS): a pilot trial of an interactive text-messaging intervention for medication adherence, socialization, and auditory hallucinations. Research Support, N.I.H., Extramural. *Schizophrenia Bulletin*. May 2012;38(3):414-25.

40. Devilly GJ, Borkovec TD. Psychometric properties of the credibility/expectancy questionnaire. *J Behav Ther Exp Psychiatry*. Jun 2000;31(2):73-86. doi:10.1016/s0005-7916(00)00012-4

41. Addington J, Shah H, Liu L, Addington D. Reliability and validity of the Calgary Depression Scale for Schizophrenia (CDSS) in youth at clinical high risk for psychosis. *Schizophrenia research*. 2014;153(1-3):64-67. doi:10.1016/j.schres.2013.12.014

42. Granholm E, Holden JL, Mikhael T, et al. What Do People With Schizophrenia Do All Day? Ecological Momentary Assessment of Real-World Functioning in Schizophrenia. *Schizophr Bull*. Feb 26 2020;46(2):242-251. doi:10.1093/schbul/sbz070

43. Depp CA, Bashem J, Moore RC, et al. GPS mobility as a digital biomarker of negative symptoms in schizophrenia: a case control study. *NPJ digital medicine*. 2019;2(1):1-7.

44. Parrish EM, Depp CA, Moore RC, et al. Emotional determinants of life-space through GPS and ecological momentary assessment in schizophrenia: What gets people out of the house? *Schizophrenia Research*. 2020;224:67-73.

45. Becker H, Stuifbergen AK, Henneghan A, Morrison J, Seo EJ, Zhang W. An initial investigation of the reliability and validity of the Compensatory Cognitive Strategies Scale. *Neuropsychol Rehabil*. 2019;29(5):739-753. doi:10.1080/09602011.2017.1329154

46. Kring AM, Gur RE, Blanchard JJ, Horan WP, Reise SP. The clinical assessment interview for negative symptoms (CAINS): final development and validation. *American Journal of Psychiatry*. 2013;170(2):165-172.

47. Weissman AN, Beck AT. Development and validation of the Dysfunctional Attitude Scale: A preliminary investigation. 1978;

48. Reddy LF, Horan WP, Barch DM, et al. Understanding the Association Between Negative Symptoms and Performance on Effort-Based Decision-Making Tasks: The Importance of Defeatist Performance Beliefs. *Schizophr Bull*. Oct 17 2018;44(6):1217-1226. doi:10.1093/schbul/sbx156

49. Attkisson CC, Zwick R. The client satisfaction questionnaire: Psychometric properties and correlations with service utilization and psychotherapy outcome. *Evaluation and Program Planning*. 1982/01/01/ 1982;5(3):233-237. doi:*https://doi.org/10.1016/0149-7189(82)90074-X*

50. Posner K, Brent D, Lucas C, et al. Columbia-suicide severity rating scale (C-SSRS). *New York, NY: Columbia University Medical Center*. 2008;

51. Rozental A, Kottorp A, Boettcher J, Andersson G, Carlbring P. Negative Effects of Psychological Treatments: An Exploratory Factor Analysis of the Negative Effects Questionnaire for Monitoring and Reporting Adverse and Unwanted Events. *PLoS One*. 2016;11(6):e0157503. doi:10.1371/journal.pone.0157503

52. Muthén B, Muthén L. *Mplus*. Chapman and Hall/CRC; 2017.

53. Hamaker EL, Klugkist I. Bayesian estimation of multilevel models. *Handbook for advanced multilevel analysis*. Routledge/Taylor & Francis Group; 2011:137-161. *European Association for Methodology series.*

54. West TV, Kenny DA. The truth and bias model of judgment. *Psychological review*. 2011;118(2):357.