

Evaluating Efficacy of a Digital Game Therapeutic for Children With Autism

PROTOCOL NUMBER: 60181

IND#: N/A

**INVESTIGATIONAL
PRODUCT:** GUESSWHAT

PROTOCOL VERSION / DATE: Version 1: 21Sept2023
Original Version 1: 21Sept2023

SPONSORED BY: Dr. Dennis Wall PhD

CONTACT INFORMATION:
Dennis Wall, dpwall@stanford.edu

Multi-Site Coordinator:
Kaiti Dunlap, kdunlap2@stanford.edu

Protocol Version and Amendment Tracking

Version Number/Amendment	Approval Date
V.01	9/14/2023

Summary of Changes

PROTOCOL SIGNATURE PAGE

Protocol Title: Evaluating Efficacy of GuessWhat Mobile System Protocol Number:
Protocol Version/Date: Version 1 / 08Aug2023
Sponsor Name: Dennis Wall

Declaration of Investigator

I confirm that I have read the above-mentioned protocol and its attachments. I agree to conduct the described trial in compliance with all stipulations of the protocol, regulations and ICH E6 Guideline for Good Clinical Practice (GCP).

Principal Investigator Name: Dennis Wall


Principal Investigator Signature:  _____

Date: 09/14/2023
Date (MM/DD/YYYY)

Statement of Compliance

The signature below provides the necessary assurance that this study will be conducted according to all stipulations of the protocol including statements regarding confidentiality, and according to local legal and regulatory requirements, US federal regulations, and ICH E6(R2) GCP guidelines.

Site Coordinator: Kaitlyn Dunlap

Site Coordinator Signature:  _____
Name

Date: 09/14/2023
Date (MM/DD/YYYY)

List of Abbreviations

• AE	• Adverse event
• eCRF	• Electronic case report form
• EDC	• Electronic data capture
• FDA	• Food and Drug Administration
• GCP	• Good Clinical Practice

1. SYNOPSIS

Title	Evaluating Efficacy of GuessWhat Mobile System
Design	Randomized, Investigator-Blinded, Cross-Over with Waitlist Control
Objectives	Determine whether GuessWhat Dtx can improve socialization in children with ASD after 8 weeks of use.
Endpoints	<p>Primary: Change from week 0 to week 8 in the Vineland Adaptive Behavior Scales-3 Socialization Subscale, Comprehensive Parent Report</p> <p>Secondary: Change from week 0 to week 8 in Parent Stress Scale, and Child Emotion Recognition Score</p> <p>Exploratory/Descriptive: Descriptive analysis of child/parent surveys, treatment regimen compliance, demographics, missing data, and any resulting treatment related AEs</p>
Study Sites	ACES ABA clinic sites including but not limited to California, Texas, Oklahoma, North Carolina.
Planned Enrollment	A minimum of 500 to a maximum of 2,000 participants will be enrolled to achieve a minimum of 360 to a maximum 900 completed participants (n=180 - 450 in each of 2 arms) evaluable
Study Population	Autistic children 3-12 years old.
Subject Entry Criteria	<p>Inclusion Criteria</p> <p>(1) Parent who is at least 18 years old of child with autism.</p> <p>(2) Child with autism is between 3 and 12 years old at time of baseline data collection.</p> <p>(3) A professional diagnosis of Autism that is able to be confirmed by negative score on a machine learning based algorithm using items from Mobile Autism Risk Assessment and video-based classifier/or by a copy of diagnostic report or by a confirmation from the therapy provider (i.e. ACES ABA).</p> <p>Exclusion Criteria</p> <p>(1) Parent does not have an Android or iOS smartphone compatible with GuessWhat App.</p> <p>(2) The parent is unable to speak/read in available translations of the study app: English, Spanish, Brazilian Portuguese.</p> <p>(3) We will exclude participants who have a self-reported concurrent psychiatric or neurologic diagnosis including genetic syndromes and epilepsy.</p>

Randomization	Block randomization stratified by child sex.
Treatment Regimen (Standard dose)	App is used 3 times a week for 15 minutes for 8 weeks
Duration of Study Participation	16 weeks with 8-week crossover
Statistical Methods	Intent-to-Treat
Study Measurements: Efficacy, Safety	Efficacy: Change in Vineland Adaptive Behavior Scales-3 Socialization Subscale week 0 to week 8 Safety: Adverse event reporting

2. TABLE OF CONTENTS

1. Synopsis	5
3. Introduction	9
3.1 Background Information	9
3.2 Previous studies	9
3.3 Rationale for the Current Study	9
4. Design	10
4.1 Study Design	10
4.2 Study Sites	10
5. Objectives	10
5.1 Primary Objective	10
5.2 Secondary Objectives	10
6. Subject Population	11
6.1 Inclusion Criteria	11
6.2 Exclusion Criteria	11
6.3 Removal of Subjects from Treatment	11
6.4 Compensation	11
7. Study Treatment - GuessWhat	11
7.1 Study Intervention	11
Standard Dose	12
7.2 Administration of GuessWhat	12
Treatment Compliance	13
Treatment Precautions	13
Concomitant Therapy	13
8. Study Procedures	13
8.1 Screening and Enrollment Steps	13
Week 0/Screening and Baseline Assessment	13
Week 1: Enrollment and Allocation	14
Weeks 1-8: Intervention	14
Week 4 Post-Test 1	14
Week 8 Post-test 2 & Cross Over	14
Week 12 Post-test 3 Follow-Up	15
Week 16 Post-Test 3	15
9. Efficacy and Safety Assessments	15
9.1 Efficacy Assessments	15
Primary Measurement (Efficacy):	15

Secondary Measurements (Efficacy):	15
9.2 Descriptive/Exploratory Measurements	15
9.3 Safety Assessments	16
Adverse Events.....	16
10. Statistical methods and analysis plan.....	17
10.1 General Considerations	17
10.2 Sample Size Justification and Power Calculation	18
10.3 Demographic and Baseline Characteristics	18
10.4 Analysis Populations	18
10.5 Stratification, Subgroup Analysis and Pooled Analysis	18
Interim Analysis.....	18
Efficacy Endpoint Analysis	19
Eight-Week Follow-up Analysis.....	19
Descriptive analyses.....	19
Safety Analysis	20
11. Recording and collection of data	20
11.1 Case Report Form	20
11.2 Study Files and Subject Source Documents	20
11.3 Monitoring	21
11.4 Audit	21
11.5 Retention of Data	21
12. Ethics	21
12.1 Ethics Committee.....	21
12.2 Subject Information and Consent	21
13. SCHEDULE OF ASSESSMENTS.....	22

3. INTRODUCTION

3.1 Background Information

We have built a prototype mobile app (guesswhat.stanford.edu) that engages children in a fluid social interaction with a social partner (e.g. mom or dad), turning the focus of the camera on the child and reinforcing prosocial learning while simultaneously measuring the child's progress in socialization. Our pilot statistics demonstrate that our development efforts are making the application more popular and engaging, with the 28-day active user base currently at over 500 users (~80% of whom are on iOS as the leading version of the app and average 4-week retention up to 25% and climbing too, respectively). To date, the app has generated more than 16,000 game sessions yielding 3,000+ videos (75+ hours) and more than 2 million highly emotive frames that demonstrate the app's potency for data acquisition. This early success supports the prediction that GW will generate enough data to build computer vision algorithms that can automatically classify key features including, but not limited to, emotion, eye gaze, and gestures. In addition to the potential for AI model development is GW's potential for therapeutic intervention.

3.2 Previous studies

We examined the therapeutic feasibility of GuessWhat in 72 children (75% male, average age 8 years 2 months) with autism who were asked to play the game for three 90-second sessions per day, 3 days per week, for a total of 4 weeks. The group showed significant improvements in Social Responsiveness Score-2 (SRS-2) total (3.97, $p < .001$) and Vineland Adaptive Behavior Scales-II (VABS-II) socialization standard (5.27, $p = 0.002$) scores.

These results demonstrated clinically significant improvements consistent with or better than previously reported behavioral therapy intervention outcomes in similarly aged participants. Overall, families played an average of 18 90-second game sessions over the four weeks, with an average of three 90-second gameplay sessions per day of use and average of six days of use over the four weeks. These sessions generated almost 1,000 videos (24 hours in total) used to generate 300,000+ frames labeled with a host of app and sensor data such as choice of game deck, number of correct guesses, etc., providing strong evidence for the feasibility of using GW to generate a rich behavioral database for training pediatric neurodevelopmental AI models.

3.3 Rationale for the Current Study

While multiple studies have demonstrated that the delivery of behavioral therapy as early as possible can effectively treat and even eliminate an ASD diagnosis,^{9–13} wait times to receive a diagnostic evaluation can,¹⁴ exceed 14 months,¹⁵ and out-of-pocket expenses can reach \$80,000 per year¹⁶ leaving one in three autistic children in the United States unable to receive the standard care. Geographic isolation and coronavirus disease 2019 (COVID-19) related lockdowns present further limitations in access to consistent and adequate treatment.

We previously demonstrated that wearable augmented reality glasses such as Superpower Glass, can deliver significant therapeutic benefits to children with autism using a fun exchange between the autistic child and a caregiver through games such as “Capture The Smile” where the child must get the caregiver to express a series of emotions. Other mobile interventions, such as AKL-T01 (EndeavorRx, Akili Interactive Labs, San Francisco, California, United States), the first game-based therapeutic to receive market clearance by the FDA in June 2020, have shown therapeutic effectiveness in attention deficit/hyperactivity disorder (ADHD), anxiety, depression, and other mental health conditions. This growing body of evidence highlights the potential for scalable digital therapies to provide access to care to patient populations in the United States and globally.

4. DESIGN

4.1 Study Design

This is a Phase 2 randomized cross over design with waitlist control. The intervention period is 8 weeks. Total study length is 16 weeks. Outcomes assessors will not be blinded, but the investigator will be blinded.

4.2 Study Sites

We will recruit from ACES ABA clinic sites across the U.S. via email marketing.

Requirements for site selection and site registration: Sites will be evaluated by Stanford team and the main PI prior to selection or activation.

Recruitment and enrollment will be decentralized and handled by a single study coordinator located at Stanford.

We will enter into formal collaboration/DUA agreements with sites who wish to be involved in the study/have access to the study data.

We will evaluate potential study sites by

1/Patient population and demographics: are there patients who meet study criteria who will be willing/able to participate?

2/ Is the study site willing to post promotional materials on site and support communication via initial recruitment email/contact?

5. OBJECTIVES

5.1 Primary Objective

To evaluate the effects of a social game-based digital therapeutic, GuessWhat, on adaptive socialization when used consistently over 8 weeks in children with autism 3-12 years compared to a waitlist control group.

5.2 Secondary Objectives

To see if use of a social game-based digital therapeutic can lead to changes in emotion recognition and parent stress when used consistently over 8 weeks.

To understand whether there is a difference in clinically meaningful changes if the device is used 4 weeks and/or 8 weeks continuously.

To understand what demographic variables (if any) make patients more likely to respond to the treatment.

To understand the safety of device when used at home with a parent/caregiver.

6. SUBJECT POPULATION

6.1 Inclusion Criteria

- (1) Parent who is at least 18 years old of child with autism.
- (2) Child with autism is between 3 and 12 years old at time of baseline data collection.
- (3) A professional diagnosis of Autism that is able to be confirmed by negative score on a machine learning based algorithm using items from Mobile Autism Risk Assessment and video-based classifier/or by a copy of diagnostic report or by a confirmation from the therapy provider (i.e. ACES ABA).

6.2 Exclusion Criteria

- (1) Parent does not have an Android or iOS smartphone compatible with GuessWhat App.
- (2) The parent is unable to speak/read in available translations of the study app: English, Spanish, Brazilian Portuguese.
- (3) We will exclude participants who have a self-reported concurrent psychiatric or neurologic diagnosis including genetic syndromes and epilepsy.

6.3 Removal of Subjects from Treatment

We will remove subjects from treatment if they experience any negative effects. We will also remove subjects from treatment if there is suspicion of fraudulent activity (i.e. impersonating a parent of child, etc.)

6.4 Compensation

Following completion of measures after completion of the 8-week intervention period, participants will receive \$50 e-gift card for their participation in the study. Participants who go on to complete T measures will receive a \$20 e-gift card. There will be opportunities for continued follow-up from T3. Participants will be able to volunteer for these additional follow-ups.

7. STUDY TREATMENT - GUESSWHAT

7.1 Study Intervention

We invented a digital treatment, GuessWhat, that operates on smartphones to create a fluid social engagement with an autistic child and his/her social partner to reinforce prosocial learning while simultaneously collecting game metric and video data that can be used to measure the child's progress and symptoms. The GuessWhat app challenges the child to interpret and imitate social and emotion-centric prompts shown on the screen of a smartphone held just above the eyes of a parent/guardian. The parent flips the phone upward/downward if able/unable to correctly guess what the child is acting, switching to the next image in the deck with each flip. The phone's front camera (facing the child) records the entire gameplay session and after each gameplay session, caregivers can choose to share the video with our research team.

Preliminary work to-date resulted in positive user feedback, evidence of engagement for both the parents and children, and meaningful gains in socialization in the child.

Parents and children choose from a variety of different game decks, with decks showing faces emoting the standard Ekman emotion expressions, cartoon characters playing sports activities, occupations, etc. Key mechanisms include facial expression labeling, imitation of varied social emotional prompts, and joint attention with a social partner during play to act out prompts and earn points.

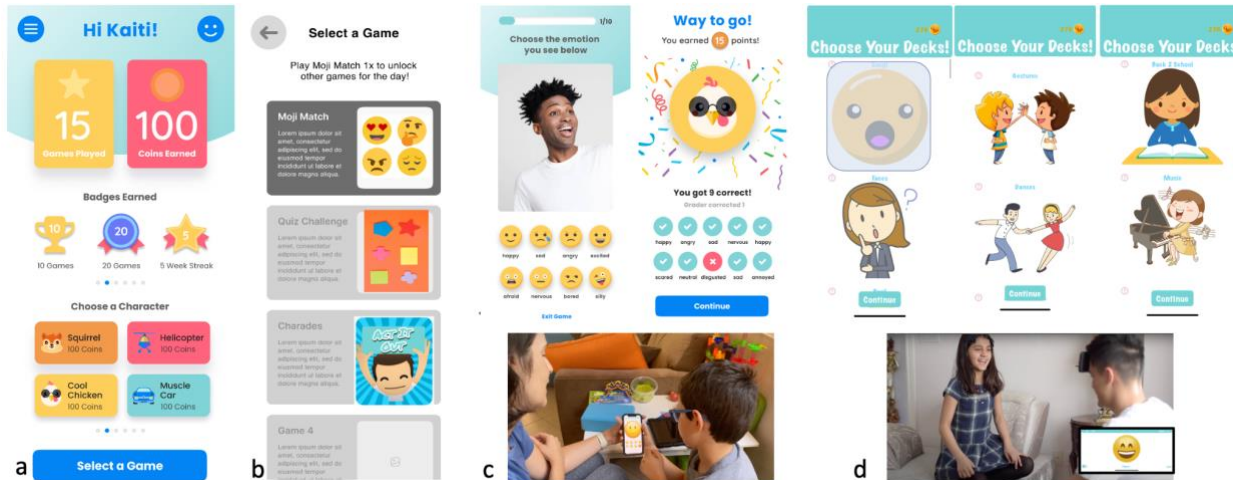


Figure 1. Mobile Intervention User Experience. GuessWhat is a mobile game available for any smartphone device that uses varied a) reward systems, and b) social emotional games to help children practice facial expression and emotion recognition, joint attention, and play. In a typical game session, c) the parent and child play a round of “MojiMatch” where the child matches an emoji to an animated GIF of a facial expression. Afterwards, the parent reviews and makes their own choices about the expression. d) Then they play “Act it Out,” where the parent holds the smartphone to their forehead and tries to guess the emotion mimicked by the child in response to the prompt shown on the phone’s screen. Upon guessing, the parent tilts the phone to proceed to the next prompt through the end of the session. After each game, parent and child can review a video of the session together. In-app game modes, unlocking deck and character choices based on coins earned, and activity-based achievement badges reinforce positive progression and ensure optimal child engagement through time.

Standard Dose

GuessWhat is a digital therapeutic application intended to improve socialization skills including joint attention and facial expression/emotion recognition in children with Autism Spectrum Disorder (ASD) between the ages of 3 and 12 years old. It is designed to be used regularly with a **caregiver 3-4 times a week for 15-20 minutes**. GuessWhat is intended for use as a part of a therapeutic program that may include clinician directed therapy, medication, and/or education programs, which further address symptoms of ASD.

7.2 Administration of GuessWhat

We will include several screening steps prior to enrollment. These will include: (1) completion of a short screening questionnaire, (2) Completion of baseline measures online, these measures will be screened for MARA score in the autism range (<2.0), IP address, Demographics age matches baseline DOB, Time to complete baseline >20 minutes (not required, but will be flagged if not met), Missing data <40% (not required, but will be flagged if not met).

A coordinator will complete the above screening steps, and if the participant meets all requirements, will send a link to download the app and corresponding access code.

Upon download, access code entry, and account creation on the app, participants will be automatically randomized to treatment or control using block randomization.

GuessWhat is administered by the parent initiating play with the child. The parent interacts with the child throughout gameplay to administer the treatment for the recommended dose.

The current trial's control is a treatment as usual group. Control participants will be asked to continue their treatment, and to inform the study team of any changes in treatment regimen during follow up assessments.

Treatment Compliance

After randomization and enrollment, research coordinators will monitor activity and contact participants by phone or email if there is no activity within the first 7 days. Participants will receive 3 reminders before being marked as lost to follow-up. The app will also send automatic notifications and automated email reminders will be sent weekly to remind parents to engage with the app.

Treatment Precautions

We identified no critical tasks during our use related risk analysis. We used simulated use to identify user tasks and determine associated hazards. To date No Adverse Events have been reported through testing. Following completion of Clinical Trial, we will include all adverse event information. We determined that use of the app presents no greater risk than a child and parent encounters in daily life.

Concomitant Therapy

We will not prohibit any concomitant therapy, pharmaceutical or behavioral. We will collect information on concomitant therapy and treatment at week 4, 8, 12, and 16.

8. STUDY PROCEDURES

- Refer also to [Schedule of Assessments](#)

8.1 Screening and Enrollment Steps

Week 0/Screening and Baseline Assessment

Participants will be screened, consented, and enrolled entirely online using electronic data capture. Parents/Guardians and coordinators will complete the following steps prior to enrollment via online forms:

- Screening questionnaire completed online
- Informed Consent administered and obtained online (waiver of documentation approved)
- Baseline questionnaires completed (Demographics, Mobile Autism Risk Assessment (MARA), Parent Stress Scale, VABS-3 Socialization subscale parent report, Emotion Recognition Child online Task)

- review of baseline data completed by coordinator
 - MARA Score <2 (required)
 - IP address matches State (not required, but will be flagged if mismatch)
 - Demographics age matches baseline DOB (required)
 - Time to complete baseline >20 minutes (not required, but will be flagged if not met)
 - Missing data <40% (not required, but will be flagged if not met)
- 1 or more of the not required flags→email+phone call to confirm participant. Coordinator sends email followed by 1 phone call attempt with no response→ participant fails screen.
- Coordinator marks Participants Screening Phase 1 CLEAR or NOT CLEAR in RedCap
- Within 1 day of being marked “CLEAR,” Participants immediately receive automated message via (1) Email, and then (2) Twilio if not downloaded within 3 days.
- If denied by coordinator: no notification.

Week 1: Enrollment and Allocation

- Upon receiving notification with download link and download access code, participant downloads the app and is automatically randomized to treatment or control using a pre-defined block randomization scheme developed by the statistician.

Weeks 1-8: Intervention

If randomized to Control, participants will continue treatment as usual. When they log-in to the app, they will receive a message that they will wait a few more weeks before beginning treatment.

If randomized to Treatment, participants will be asked to play the app 3 times a week for 10-15 minutes for the next 8 weeks and they will receive weekly push notifications and email reminders.

Week 4 Post-Test 1

Treatment and Control will be asked to complete all outcome measures and will be asked about any changes in treatment or safety events at week 4. Participants will also be asked to complete a question assessing time spent playing with their child over the last month in emotion-centric theatrical play.

Week 8 Post-test 2 & Cross Over

At week 8 from initial app download date, participants will be asked via email to complete all outcome measures as well as check-in on any changes in treatment and any safety events. These will be collected via RedCap online surveys.

At week 8 from initial download date, treatment participants' access to the app will be removed, and control participants' access to the app will be provided. This will occur automatically in the app. We will use a script that runs daily to manage crossover participants' app experience. After an additional 8 weeks (week 16), treatment participants will regain access to the app.

Week 12 Post-test 3 Follow-Up

Treatment and Control will be asked to complete all outcome measures and will be asked about any changes in treatment or safety events at week 4. Participants will also be asked to complete a question assessing time spent playing with their child over the last month in emotion-centric theatrical play.

Week 16 Post-Test 3

Treatment and Control will be asked to complete all outcome measures and will be asked about any changes in treatment or safety events at week 4. Participants will also be asked to complete a question assessing time spent playing with their child over the last month in emotion-centric theatrical play. Treatment participants will regain access to the app.

9. EFFICACY AND SAFETY ASSESSMENTS

9.1 Efficacy Assessments

Primary Measurement (Efficacy):

1. Change in Vineland Adaptive Behavior Socialization Parent Report from baseline (week 0) to week 8

Vineland Adaptive Behavior Scales, 3rd edition (VABS-3) Socialization subscale of the Parent/Caregiver Comprehensive form will be administered online to the parents. Scores from the socialization domain of the VABS-3 reflects one's functioning in social situations. The socialization subscale is up to 112 items depending on age and development, where raw scores are converted to IQ-type standard scores--v-scale scores (M=15, SD=3) where scores range from 1 to 24, and factor in age equivalents, growth scale values, and higher scores indicate better adaptive functioning

Secondary Measurements (Efficacy):

1. Change in Parent Stress Scale from baseline (week 0) to week 8

The Parental Stress Scale is an 18-item parent report scale that evaluates the extent of parent stress. Scores range from 18-90, where higher scores indicate a higher level of Parental stress.

2. Change in Child Emotion Recognition score from baseline (week 0) to week 8

Study team will use an image-based emotion recognition task where participants will select the emotion that corresponds to an emotional facial image in the form of a GIF (Graphics Interchange Format) image. Participants will select 1 of 8 available Ekman emotion labels for 16 facial emotional image stimuli presented in random order. Correct responses will be measured against a predetermined majority rules consensus of the emotional content of the GIFs.

3. Change in Vineland Adaptive Behavior Socialization Parent Report from baseline (week 0) to week 4
4. Change in Parent Stress Scale from baseline (week 0) to week 4
5. Change in Child Emotion Recognition score from baseline (week 0) to week 4

9.2 Descriptive/Exploratory Measurements

1. Game metric Analysis for Treatment participants
 - Average duration (average time playing per day)
 - Average frequency (days played per week) of play

- % Moji Match GIFs with disagreement between parent and child
 - Average emotion confusion for each of the Ekman emotions pre vs. post
 - Average time to guess in Act in Out
 - Time spent acting out a prompt, by prompt type (Sports, Emotions, Occupations, etc.)
 - Frequency of correct vs. incorrect guesses during Act it Out
2. treatment regimen compliance
 3. demographics
 4. missing data

9.3 Safety Assessments

Adverse Events

Definition of an Adverse Event

The FDA Safety Guidance, referencing 21CFR312.32(a), defines an Adverse Event as follows:

Adverse event means any untoward medical occurrence associated with the use of a drug in humans, whether or not considered drug related.

An adverse event (also referred to as an adverse experience) can be any unfavorable and unintended sign (e.g., an abnormal laboratory finding), symptom, or disease temporally associated with the use of a drug and does not imply any judgment about causality. An adverse event can arise with any use of the drug (e.g., off-label use, use in combination with another drug) and with any route of administration, formulation, or dose, including an overdose.

Evaluating and Reporting of Adverse Events

All AEs will be promptly documented on the Adverse Event in RedCap and assessed by the Principal Investigator. Details of the event must include the dates of onset and resolution, severity, relationship to study drug, seriousness, and whether the event caused the subject to withdraw from the study, outcome and timing with regard to administration of the study drug.

Severity: Severity should be graded and recorded as follows:

- Mild: Awareness of event but easily tolerated
- Moderate: Discomfort enough to cause interference with usual activity
- Severe: Inability to carry out usual activity, incapacitating, requires medical intervention

Relationship: The relationship of the Adverse Event to the study device will be determined initially by the Investigator, and assessed using the following definitions:

- **Related:** There is a distinct temporal relationship between the event onset and administration of the study drug. There is a known reaction to agent or chemical group or predicted by known pharmacology. The event cannot be explained by subject's clinical state or other factors.
- **Unrelated:** Evidence exists that the AE has an etiology other than the study drug (e.g., pre-existing condition, underlying disease, intercurrent illness, or concomitant medication).

Final determination of relatedness will be made by the Stanford Institutional Review Board during an unanticipated problem or adverse event submission.

Expectedness: Whether or not the event was expected or unexpected. Further details on this can be found in section 8.5.

These criteria, in addition to good clinical judgment, should be used as a guide for determining the causal assessment. If it is felt that the event is not related to study therapy, then an alternative explanation should be provided.

Serious Adverse Events (SAEs)

Suspected Unexpected Serious Adverse Reactions (SUSARs)

A SUSAR carries specific and time-based reporting requirements for the Sponsor of a clinical trial. Thus, after a Site Investigator reports an SAE, the FDA expects the Sponsor will determine whether it meets the definition of a SUSAR.

A SUSAR is defined according to 3 criteria:

1. The AE is deemed a “suspected adverse reaction” if there is a reasonable possibility that the study drug caused the AE. A “reasonable possibility” means there is evidence to suggest a causal relationship between the drug and adverse event.
2. The AE is “Serious” if it meets the definition of an SAE.
3. The AE is deemed “unexpected” if it is not listed in the Investigator’s Brochure (IB) or if in the IB, has not been reported at the severity observed.

In cases where the Investigator deems a SUSAR has occurred, it must file an IND Safety Report with the FDA. FPHU will require the assistance and cooperation of the Investigator and staff to provide accurate and complete information on the subject and observed SAE so that reporting requirements to the FDA can be met.

Reporting SUSARs to the FDA: IND Safety Reports

IND safety reports are used to submit reports of SUSARs to the FDA. There are 2 types of reports:

- A “15-day report” is used when the reported SAE is a SUSAR and requires that as much information as is available to the investigator and the sponsor, be submitted to the FDA in on the appropriate form. For US trials, the appropriate form is the FDA Form 3500A also commonly known as a “MedWatch” form.
- A “7-day report” is used when the SUSAR is considered to be fatal or life-threatening.

The 7-day and 15-day timelines begin the day that the Sponsor determines the information qualifies for reporting and are counted in calendar days – not business days. Therefore, it is important that the Site investigator carefully follow the reporting requirements.

10. STATISTICAL METHODS AND ANALYSIS PLAN

10.1 General Considerations

For the primary comparison, the following will be tested:

- Null hypothesis: There will be no difference between the observed change in VABS-3 Socialization scores from baseline to week 8 between treatment and control participants.
- Alternative hypothesis: We will observe a minimally clinically important increase on the VABS-3 Socialization score from baseline to week 8 in the treatment group, but not in the control group.

Hypothesis tests will be two sided and conducted at an overall $\alpha = 0.05$ level of significance.

10.2 Sample Size Justification and Power Calculation

We will enroll up to 2,000 participants to achieve a 50% enrollment rate and 36% retention rate.

We aim to successfully retain 36% of enrolled families and have 360 parents and their child with autism aged 3-12 years old complete all study procedures (180 Treatment, 180 Control).

We will close out the study after enrolling 360 participants with completed endpoints, estimated for 90% power and 2-sided significance of .05 to detect minimal differences in VABS-III Socialization Standard Score of 3.1 points, consistent with the results of a GuessWhat feasibility study previously reported by our lab.

10.3 Demographic and Baseline Characteristics

Summary statistics will be provided per treatment group for demographic (race, gender, income/SES, age) and other initial subject characteristics (e.g., concomitant disorders, concomitant treatments) will be provided per treatment group and for the total group. The ITT population will be used for the summaries.

10.4 Analysis Populations

The following analysis populations will be defined for the study:

- (1) The intent-to-treat (ITT) population will include all randomized patients. Patients will be analyzed according to their assigned treatment arm. All efficacy analyses will be completed in the ITT population.
- (2) Per-Protocol (Treatment Compliant): This population is defined as any participant that began the initial 8-week treatment or waitlist control phase, completed both the pre and post assessments, and met a minimum compliance level of 50%. Participants that received incorrect intervention after randomization will be excluded from this population.
- (3) The safety population will include all patients who receive study treatment. Patients will be analyzed according to actual treatment received. All safety analyses will be completed in the safety population.

10.5 Stratification, Subgroup Analysis and Pooled Analysis

Following consent, those subjects meeting all entry criteria will be randomized in a 1:1 ratio. Random blocks of size 2 or 4 will be utilized.

Interim Analysis

An interim analysis will be conducted on the primary outcome, VABS-3 Socialization, to determine if the minimum sample size of 360 participants is sufficient to detect a statistical difference between Treatment and Waitlist Control groups. The population for this analysis will follow the same guidelines used for inclusion into the ITT population (See [Analysis Populations](#)). The specific results of the interim analysis will remain blinded and will not be used to declare success of the primary objective at the interim; Un-blinded staff will conduct this analysis for the purpose of finalizing the sample size only. To maintain blinding for all other study personnel and affiliates, the results of this analysis will only indicate the sample size necessary to sustain the power of the study to detect an effect size of 0.4. Enrollment in the study will continue until at least a minimum of 360 patients have been enrolled and the study has achieved the number of enrolled patients necessary to sustain the power of the study. The study will not exceed 2,000 enrolled patients.

Efficacy Endpoint Analysis

To test the efficacy of GW plus ABA vs ABA alone, we will use a generalized linear mixed-effects regression model. The model will include all available measurements for all participants randomized in the study, including withdrawn participants and those who did not adhere to the intervention. Specifically, for outcome measure M in the measures section above, and individual i at time of measure t , we will apply the following model: $M_{it} = \gamma_0 + \gamma_{0i} + \gamma_1 \text{ weeks}_{it} + \gamma_2 \text{ treatment}_i + \gamma_3 \text{ weeks}_{it} \times \text{treatment}_i + \varepsilon_{it}$, where the individual-specific random effect, γ_{0i} , is included to account for correlation across repeated measurements within an individual over time, weeks_{it} represents the number of weeks elapsed since randomization for the i th individual at the t th time point, and treatment_i is an indicator representing whether the i th individual belongs to the treatment-first group. To test for significance, we will use a 2-sided Wald test on γ_3 (representing the impact of treatment over time).

We will further assess whether age or sex are modifiers of treatment in the regression analysis. For this purpose, we will augment the above analysis to include the modifier (e.g., age) and interaction terms with treatment, week, as well as a 3-way interaction term with treatment and week. Likelihood ratio tests will then assess whether the modifier significantly impacted the treatment effect. In addition to the primary ITT cohort, we will apply the same mixed-effects model analysis to the per-protocol cohort to evaluate the impact of the intervention on outcomes among those who adhered to the study (per-protocol group).

Eight-Week Follow-up Analysis

To test if a return to baseline occurred in the treatment group, we will apply a model that tracks gains from intake to posttest 2 and posttest 4 only among those assigned to the treatment group: $M_{it} = \gamma_0 + \gamma_{0i} + \gamma_1 \text{ weeks}_{it} \times [\text{posttest 1}]_{it} + \gamma_2 \text{ weeks}_{it} \times [\text{posttest 2}]_{it} + \varepsilon_{it}$, where a Wald test on γ_2 will allow us to test whether gains at 16 weeks are (still) significantly different from the intake baseline. The sign and magnitude of the coefficient indicates the point value of the change. We will also perform a follow-up analysis on the entire treatment cohort, including control participants who crossed over.

Secondary Measures

We will apply the ITT model from the primary analysis to the secondary measures, Parent Stress Scale and Child Emotion Recognition measure. We will also perform the same moderator analysis on the secondary outcome measures.

Treatment Cohort Analysis

We will perform further exploratory analyses on the treatment-first cohort and then on the full treatment cohort, which included the crossover group. Using mixed-effects linear regression methods, we will compare changes in outcome measures to baseline values of age, usage, and ABA therapy hours to evaluate additional trends, such as increased effect of GW use on particular age groups.

Descriptive analyses

Descriptive statistics (proportions for categorical variables, means, medians, standard deviations, and interquartile ranges for continuous variables) will be reported for all key patient variables, including baseline and demographic characteristics, use of medications, compliance, and study completion status. Data that are missing on key patient characteristics and the outcome will be fully described, including any patterns of missingness (i.e., any relationships between missingness of a variable and patient characteristics).

A CONSORT diagram displaying the number of patients screened, eligible, and consented along with reasons for ineligibility will be provided. Graphical tools such as histograms, boxplots, and scatterplots will be created to assess quality of data and to display patterns over time.

Safety Analysis

Treatment related adverse events for GuessWhat intervention will be presented. Any adverse events occurring during any phase of the study judged by clinical sites' Primary Investigator (PI) to be related to the intervention will be recorded and presented in a table listing. The Safety population (see [analysis population details](#)) will be used for this analysis that will include randomized participants. The severity of each event will be evaluated by the PI and presented with the recorded events. If there are frequently occurring treatment related adverse events, a frequency table will be included, to rank events from most frequent to most rare. Subjects will be included in the actual treatment group received.

11. RECORDING AND COLLECTION OF DATA

11.1 Case Report Form

The Investigator or designee will record all data collected on the electronic Case Report Form (eCRF) provided for that purpose. For this study, REDCAP will be the eCRF. The site will be suitably trained on the use of the eCRF and appropriate site personnel will be provided electronic signatures.

All site entries will be made in a secured web site and the Principal Investigator will review the record for completeness. Upon completion of the review, the PI will sign electronically in the signature page of the eCRF.

The Investigator or designee will make necessary eCRF corrections. The investigator must authorize the corrections to the entered data on eCRF.

Specific instructions on use of the EDC system and guidelines for data entry and correction will be provided to the sites.

11.2 Study Files and Subject Source Documents

Subject confidentiality is strictly held in trust by the participating investigators, research staff, the Sponsor and their designees. This confidentiality is extended to cover testing of biological samples in addition to the clinical information relating to subjects. Authorized representatives of the Sponsor may inspect all documents and records required to be maintained by the Investigator, including but not limited to, medical records (office, clinic or hospital) and pharmacy records for the subjects in this study. Any data, specimens, forms, reports, and other records that leave the site will be identified only by a subject identification number to maintain confidentiality.

The Investigator must maintain adequate and accurate records to enable the conduct of the study to be fully documented and the study data to be subsequently verified. These documents include Investigators' Study Files and original subject clinical source documents generated at the study site. The term "original" means the first recording of the data.

The Investigator will ensure the site master files are maintained, including the study protocol and its amendments, IRB and regulatory approvals with associated correspondence, informed consents, study drug records, staff curriculum vitae, all correspondence, and other appropriate documents.

Subject clinical source documents may include, but are not limited to, subject hospital/clinic records. The Investigator must assure that all original source documents are available to support monitoring activities.

11.3 Monitoring

Monitoring will be conducted according to the applicable ICH and GCP guidelines to ensure protocol adherence, quality of data, drug accountability, compliance with regulatory requirements and continued adequacy of the investigational site and its facilities. The site study staff will cooperate in the monitoring process by ensuring the availability of the eCRFs, source documents and other necessary documents at the time of monitoring and by prompt attention to any matters brought to their attention by the monitor.

11.4 Audit

ICH guidelines for GCP require independent inspection of clinical program activities. Such inspections may be performed at any time - before, during and/or after the study. The site Investigator and site study staff are responsible for maintaining the site master file containing all study-related regulatory documentation as outlined by the Sponsor that will be suitable for inspection at any time by the Sponsor, its designees, and/or regulatory agencies. The Investigator understands and agrees to give access to the necessary documentation and files.

11.5 Retention of Data

The site PI will maintain the records of device/tablet disposition, final eCRFs, worksheets, and all other study-specific documentation (e.g., study file notebooks or source documentation) until notified by the sponsor that records may be destroyed. To avoid error, the investigator will contact the sponsor or their designated representative before the destruction of any records pertaining to the study to ensure they no longer need to be retained. In addition, the sponsor will be contacted if the site PI plans to leave the institution so that arrangements can be made for the transfer of records.

12. ETHICS

12.1 Ethics Committee

A properly constituted, valid IRB/IEC must review the treatment plan and procedures, the Investigator's informed consent document, and related subject information. It is the responsibility of the Investigator to ensure that all aspects of institutional review are conducted in accordance with current regulations governing the jurisdiction where the study is conducted. The Sponsor (or designee) must receive a letter documenting IRB/IEC approval that specifically identifies the title of the treatment plan, subject information sheet, and ICF.

12.2 Subject Information and Consent

It is the responsibility of the site Investigator to ensure that written informed consent is obtained from the subject or legal representative before any activity or procedure is undertaken that is not part of routine care. The informed consent must comply with local regulations.

The background of the study, the procedures, the potential benefits and risks of the treatment, and the fact that treatment is voluntary for the subject must be explained to the subject or legal representative. The subject or representative must be given sufficient time to consider whether to receive compassionate treatment. A copy of the ICF, signed and dated by the subject/representative and the site Investigator (or designee), must be given to the subject/representative.

Each consent form should contain an authorization allowing the Investigator and the Sponsor (or designee) to use and disclose protected health information (PHI) (i.e., subject-identifiable health information) in compliance with local law.

13. SCHEDULE OF ASSESSMENTS

	Screening Week 0	Enrollment Week 1	Check-In 1 Week 4	Primary Assessment Week 8	Check-In 2 Week 12	Crossover follow-up Week 16
Screening						
Pre-Screening Consent	X					
Screening Questionnaire	X					
Informed Consent	X					
Assessments						
Demographics Questionnaire	X					
Mobile Autism Risk Assessment	X					
Parent Stress Scale	X		X	X	X	X
Vineland Adaptive Behavior Scales-3 Socialization Subscale Parent Report	X		X	X	X	X
Child Emotion Task	X			X		X
App Download	X			X		
Change in treatment/play questionnaire			X	X	X	X
Allocation/Randomization		X				
Intervention period is 8 weeks		X	X	X		
Cross Over period is 8 weeks				X	X	X