SOCIAL COGNITION PSYCHOMETRIC EVALUATION (SCOPE) STUDY FOR EARLY PSYCHOSIS

Kelsey A. Ludwig

A thesis submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of doctor of philosophy in the Department of Psychology and Neuroscience at the University of North Carolina at Chapel Hill.

Chapel Hill
2017

Approved by:

David Penn

Gabriel Dichter

Keith Payne

ABSTRACT

Kelsey A. Ludwig: Social Cognition Psychometric Evaluation (SCOPE) for Early Psychosis
(Under the direction of David L. Penn)

Social cognition is an important outcome in schizophrenia research. Unfortunately, there has been a lack of consensus regarding which measures of social cognition best capture this domain of functioning. The Social Cognition Psychometric Evaluation (SCOPE) study was developed to address the need for a battery of measures that have sound psychometric properties and can be implemented in clinical trials for individuals with chronic schizophrenia. The current study expands upon the SCOPE study by examining the psychometric properties of the eight candidate measures administered to individuals early in the course of psychosis. Thirty-eight stable outpatients with first episode psychosis (FEP) and thirty-nine healthy controls completed the battery at baseline and one-month follow-up assessments. The SCOPE battery was evaluated on a collection of psychometric properties, including: (1) Reliability – including test-retest and internal consistency, (2) Between group differences – including direct comparisons between first episode patients, the chronic schizophrenia sample from SCOPE, and both demographically-matched control groups, (3) Utility as a repeated measure, (4) Convergent and discriminant validity, (5) Relationship to social and occupational functioning, (6) Incremental validity – variance in functioning beyond neurocognition, and (7) Feasibility – including practicality of administration and tolerability. Social cognition accounted for substantially more variance in functional outcome than neurocognition. Participants with FEP outperformed chronic schizophrenia patients on the majority of candidate measures of social cognition. Only one measure, the Hinting task, displayed adequate psychometric properties to

be recommended for use in clinical research with first episode psychosis. The remaining candidate measures would require modifications before implementation or cannot be recommended for use in clinical research with first episode psychosis.

To my mentor, advisor, and long-time friend: I cannot even imagine enduring this process without you. Thank you for all of your guidance and comic relief along the way. You strike an impressive balance between supporting and challenging your students, which I will never forget and always appreciate. Thank you for helping me flourish throughout this process.

ACKNOWLEDGEMENTS

We would like to thank all individuals who participated in this investigation, including all research subjects and research assistants at the participating sites for their significant effort and contribution to the SCOPE study. This work was supported by the National Institute of Mental Health (R01 MH093432 awarded to Drs. Harvey, Penn and Pinkham). Drs. Harvey, Penn and Pinkham (Principal Investigators) should also be acknowledged for designing the study and writing the protocol for the project on which this study was based. A special thank you should be extended to Dr. Amy Pinkham for her valuable insights and helpful suggestions to improve this thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ABBREVIATIONS

FEP	First Episode Psychosis
SCZ	Schizophrenia
HC	Healthy Control
NOS	Not Otherwise Specified
AIHQ	Ambiguous Intentions and Hostility Questionnaire
BLERT	Bell-Lysaker Emotion Recognition Task
ER-40	Penn Emotion Recognition Test
Eyes	Reading the Mind in the Eyes Test
TASIT	The Awareness of Social Inferences Test
Hinting	Hinting Task
RAD	Relationships Across Domains
Trust	Trustworthiness Task
MCCB	MATRICS Consensus Cognitive Battery
WRAT	Wide Range Achievement Test
UPSA-B	UCSD Performance-Based Skills Assessment-Brief
SSPA	Social Skills Performance Assessment
PANSS	Positive and Negative Syndrome Scale

INTRODUCTION

Social cognition, defined as the mental processes underlying people's capacity to perceive, process and comprehend social information, is related to quality of life, daily living skills and occupational functioning in schizophrenia (Frith, 2008; Green, Hellemann, Horan, Lee, & Wynn, 2012; Kunda, 1999; Mancuso, Horan, Kern, & Green, 2011). Social cognition accounts for additional variance in functioning than various cognitive factors (Brüne, 2007), and mediates the relationship between neurocognition and functioning in psychosis (Fett et al., 2011; Schmidt, Mueller & Roder, 2011). Based on its relation to functional outcome, social cognition in schizophrenia has garnered considerable research interest over the past few decades and is increasingly considered a viable target for treatment (Couture, Penn, & Roberts, 2006; Fett et al., 2011; Green & Leitman, 2008; Penn, Corrigan, Bentall, Racenstein, & Newman, 1997).

Despite burgeoning interest in studying social cognition, studies investigating these constructs vary greatly in the tasks employed, many of which may lack a strong empirical foundation and involve unknown or questionable psychometric properties (Couture & Penn, 2012; Fett et al., 2011; Savla, Vella, Armstrong, Penn, & Twamley, 2013; Thompson, Bartholomeusz, & Yung, 2011). The absence of a validated battery of social cognitive measures is problematic as inadequate and inconsistent measurement can jeopardize the validity, reproducibility, and comparability of findings, and may lead to effective treatments being discarded or ineffective treatments pursued (Drost, 2011).

To address this need, an ongoing NIMH project called the Social Cognition Psychometric Evaluation (SCOPE) study was initiated (Pinkham et al., 2014; Pinkham et al., 2015). SCOPE is

a multiphase project that involves identifying the currently accepted domains of social cognition, selecting the best available measures to assess these domains, and administering tasks to a large sample of stable outpatients with schizophrenia and demographically-matched controls.

Findings from the initial validation study suggested the Bell-Lysaker Emotion Recognition Task (*BLERT*; Bell, Bryson, & Lysaker, 1997), Penn Emotion Recognition Task (*ER-40*; Kohler et al., 2003), Reading the Mind in the Eyes Test (*Eyes*; Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2003), The Awareness of Social Inferences Test (*TASIT*; McDonald, Flanagan, Rollins, & Kinch, 2003), and Hinting Task (*Hinting*; Corcoran, Mercer, & Frith, 1995), displayed acceptable reliability and validity for implementation in clinical research. Remaining measures, including: Ambiguous Intentions Hostility Questionnaire (*AIHQ*; Combs, Penn, Wicher, & Waldherer, 2007), Relationships Across Domains (*RAD*; Sergi et al., 2009), and Trustworthiness Task (*Trust*; Adolphs, Tranel, & Damasio, 1998), demonstrated weaker characteristics and were deemed inadequate for use in clinical trials targeting social cognition (Pinkham et al., 2015), although subsequent findings support continued development and use of the AIHQ Blame Score (Buck et al., 2017, in press).

Importantly, SCOPE included a predominantly middle-aged, chronic sample typical of many treatment studies. There is some debate as to whether first episode psychosis (FEP) and chronic schizophrenia patients should exhibit the same types and degree of social cognitive impairment (Savla et al., 2013; Thompson, Bartholomeusz, & Yung; 2011; Ventura et al., 2015). Some research suggests attenuated or unremarkable deficits earlier in the course of illness (An et al., 2010; Bora & Pantelis, 2013; Romero-Ferreiro et al., 2016; Sprong et al., 2007), though findings are mixed (Barkl et al., 2014; Green et al., 2012; Horan et al., 2012; Zaytseva, Burova, Garakh, & Gurovich, 2013). FEP samples may also be more heterogeneous than many chronic

schizophrenia samples (Birchwood et al., 1998), and differences in social cognition across phase of illness may stem from variations in clinical stability (Bora & Pantelis, 2013; Green et al. 2012) and age-related changes in neurocognitive abilities (Hartshorne & Germine, 2015). Consequently, the results of SCOPE may not accurately represent younger individuals with FEP.

The purpose of the current study was to extend Pinkham et al.'s (2015) psychometric investigation of the SCOPE battery with a younger FEP sample. Paralleling SCOPE, we report on: (1) Reliability: test-retest, internal consistency, (2) Between-group differences: including directly comparing first episode patients, chronic schizophrenia patients from SCOPE, and both demographically-matched control groups, (3) Utility as a repeated measure, (4) Convergent and discriminant validity, (5) Relationship to social/occupational functioning, (6) Incremental validity: variance in functioning beyond neurocognition, and (7) Feasibility: practicality of administration and tolerability.

Investigations comparing first episode and chronic schizophrenia patients frequently involve control groups that differ in important demographics from at least one clinical group (Romero-Ferreiro et al., 2016). This thesis provides a direct comparison of performance on the social cognition battery when administered to chronic schizophrenia, first episode psychosis, and their respective, demographically matched control groups. Recommendations regarding suitability of the candidate measures for implementation in clinical research with first episode psychosis outpatients have also been provided.

METHOD

Participants

The study took place at the University of North Carolina at Chapel Hill. FEP patients were primarily recruited from the Outreach and Support Intervention Services (OASIS) clinic in Carrboro, NC. Patients required a diagnosis of schizophrenia, schizoaffective disorder, schizophreniform disorder, or psychosis NOS, confirmed by the Structured Clinical Interview for DSM-IV Axis-I Disorders, Patient Edition (*SCID-P*; First et al., 2002). OASIS clinicians and/or a trained research assistant at UNC-CH conducted all diagnostic interviews.

Participants were excluded if diagnosed with psychosis for greater than five years, or had been hospitalized within the last two-months. Deterioration is most common before illness onset and during the first few years of psychosis (Birchwood et al., 1998). Furthermore, evidence indicates a subsequent illness "plateau," during which a level of relative stability is established 2-5 years after illness onset (Srihari et al., 2012). Thus, a cut-off of five years for illness duration was used. Participants were required to be on a stable medication regimen for a minimum of the two-month hospitalization-free period, though they were not excluded if psychiatrically stable while not receiving antipsychotics.

Control participants were recruited through community flyers and online advertisements. Controls were selected for similarities in age/gender to FEP outpatient participants. Controls were precluded from participation for meeting criteria for any Axis I/II disorder according to the DSM-IV, or if they had a first-degree family member with a history of psychosis. Chronic outpatient and matched control participants from the original SCOPE study were included in

analyses examining possible performance differences across stage of illness only. For detailed demographic information, please refer to Pinkham et al., 2015.

All participants were considered ineligible based on: 1) presence/history of mental retardation, 2) presence/history of brain injury and/or neurological disorder (e.g., seizures, multiple sclerosis), 3) sensory limitation that would interfere with assessment (e.g., blindness/deafness), and/or 4) evidence of non-nicotine substance dependence in the past six-months, with substance use not being exclusionary. Evidence of substance dependence was collected from patients' healthcare providers, via chart review, and/or through substance use disorder modules from the SCID (DSM-IV; First et al., 2002).

Measures

Measures of Social Cognition

We administered identical versions of the eight candidate measures of social cognition from the initial psychometric evaluation (Pinkham et al., 2015), including:

Attributional style

The Ambiguous Intentions and Hostility Questionnaire, abbreviated version (*AIHQ*) measures the extent to which an individual tends toward a hostile or aggressive interpretation of everyday situations or common occurrences (Combs et al., 2007). For this task, participants were asked to read a set of five ambiguous situations and imagine how each scenario might actually happen to them. They were first asked to provide (a) a brief explanation for *why* the person in the scenario acted the way they did. Then, they answered three Likert-style questions to gauge (b) to what extent the participant believed the person in the story behaved the way they did *on purpose* (from 1 = "Definitely No" to 6 = "Definitely Yes"), (c) *how angry* the participant would feel (from 1 = "Not at All Angry" to 5 = "Very Angry"), and (d) the degree to which

he/she would *blame* the other person (from 1 = "Not at All" to 5 = "Very Much"). Finally, they were asked to provide (e) a brief description of a probable behavioral response (i.e., what they would do about the situation).

Likert-style questions, averaged per item and aggregated across five scenarios, comprise a Blame Score (BS, range: 3-16). Two independent raters with high concordance (ICCs ranged .834 to .967 for the individual items) coded participants' responses to the two open-ended questions (items a and e) to calculate Hostility Bias (HB, range: 1-5) and Aggression Bias (AB, range: 1-5). Higher scores reflect more hostile or aggressive attributional styles.

Emotion processing

The Bell Lysaker Emotion Recognition Task (*BLERT*) measures an individual's ability to correctly identify the feeling or emotion expressed during a series of 21 short video clips (Bell, Bryson, & Lysaker, 1997). Participants were asked to view 10-second recordings of a male actor providing facial, vocal, and bodily indications of seven distinct emotions: sadness, anger, disgust, happiness, surprise, fear or no emotion. The dependent measure was the total number of correct responses, ranging from 0 to 21.

The Penn Emotion Recognition Test (*ER-40*) examines the ability of individuals to distinguish between various facial expressions of emotion (Kohler et al., 2003). In this task, participants are instructed to choose which feeling or emotion is being expressed in a set of forty photographic stimuli. The individuals represented are ethnically diverse, stem from multiple age groups, and include an equal number of male-female actors. Photographs are balanced for each emotion category (i.e., happy, sad, anger, fear, no emotion) and vary in terms of intensity of the expression. Participants were shown each photograph sequentially and asked to select an

emotion before moving onto the next slide. The dependent variable was the total number of correct responses, ranging from 0 to 40.

Theory of mind

The Reading the Mind in the Eyes Test (*Eyes*) measures how well a person is able to conceptualize and discern the mental state of another person based on a set of static visual stimuli (Baron-Cohen et al., 2001). The set of stimuli is comprised of 36 black-and-white photographs depicting the eye-region of various actors' and actresses' faces. Participants were shown each stimulus separately and asked to choose the most appropriate of four possible mental state descriptors. Participants were provided with a glossary to peruse whenever he or she felt unclear about the meaning of particular term. The dependent measure was the total number of correct responses, ranging from 0 to 36.

The Awareness of Social Inferences Test, Part III (*TASIT*) assesses the extent to which an individual comprehends interpersonal contextual cues; i.e., sarcasm or dishonesty (McDonald et al., 2003). Participants viewed 16 short video clips depicting conversations between two or more people, and answered four basic questions about each exchange. The questions gauged the participant's understanding of the beliefs, meanings, intentions and feelings expressed by the actors. All participants viewed vignettes from Form A during the first assessment and watched clips from Form B during the second assessment. The dependent variable was the total number of correct responses, ranging from 0 to 64.

The Hinting Task (*Hinting*) examines the ability of a participant to decipher the true meaning of indirect verbal cues, or hints (Corcoran, Mercer, & Frith, 1995). The experimenter reads a series of 10 brief conversations to the subject. Each conversation involves one of two characters hinting something to the other person, and the participant is asked to provide an

explanation of each indirect verbal cue. If the subject misinterprets the situation, an additional clue is given so as to provide the participant the opportunity to earn partial credit. The dependent measure was the total number of correct responses, ranging from 0 to 20.

Social perception

The Relationships Across Domains, abbreviated version (*RAD*) is a paper-and-pencil based measure of social perception (Sergi et al., 2009). The abbreviated version includes a set of fifteen vignettes that describe a male-female relationship to which participants were asked to answer three yes/no questions related to each dyad. This task was used to measure participants' implicit understanding of social relationships and ability to predict others' behaviors according to four relational models that influence social behavior: communal sharing, authority ranking, equality matching, and market pricing. The dependent variable was the total number of correct responses, ranging from 0 to 45.

Novel

The Trustworthiness Task (*Trust*), labeled novel as it does not fit neatly under any of the four currently accepted domains of social cognition, gives an indication of individuals immediate social judgments of unknown others (Adolphs, Tranel, & Damasio, 1998). Participants were shown a set of static visual stimuli, 42 black-and-white photos of strangers' faces, and were asked to judge the extent to which he/she would trust the person in the photograph. They provided ratings on a 7-point scale (from -3 = "Very Untrustworthy" to +3 = "Very Trustworthy"). Rather than summing the number of "correct" responses, the average rating of trustworthiness provided for the set of visual stimuli served as the dependent measure.

Neurocognitive Measures

Neurocognition was measured using a subset of *The MATRICS Consensus Cognitive Battery* (*MCCB*): Trail-Making Test-Part A, BACS-Symbol Coding, Category Fluency-Animal Naming, Letter-Number Span, and Hopkins Verbal Learning Test-Revised (Nuechterlein et al., 2008). Consistent with SCOPE, subtests were selected according to correlations with composite scores of neurocognitive performance (Keefe et al., 2006; Pinkham et al., 2015). With the exception of the Trails task, for which reverse scores were used, higher scores reflect better performance on the selected measures of the MCCB. Neurocognitive scores were calculated by a) computing *t*-scores for each subtest using the following formula: $T = (X - M_{FEP}) / (s / \sqrt{(n_{FEP})}$, and b) computing a standardized composite score by taking the average of the *t*-scores for each subtest: $MCCB_{Composite} = (1/5)(T_{LNS} - T_{Trails} + T_{Sym} + T_{HVLT\text{-}R} + T_{Animals})$.

The Wide Range Achievement Test (WRAT-3) Reading subscale assesses an individual's reading ability (Weickert et al., 2000). For the purposes of this study, this measure was used to supplement the MCCB and provide an estimate of IQ.

Measures of Social and Occupational Functioning

Community and daily living skills were assessed using *The UCSD Performance-Based Skills Assessment-Brief* (*UPSA-B*; Mausbach et al., 2007), a performance-based measure of functional capacity, and *The Specific Level of Functioning Scale, Self-Report* (*SLOF*; Schneider & Struening, 1983).

Social skills were assessed with *The Social Skills Performance Assessment* (*SSPA*; Patterson et al., 2001). Participant and experimenter acted out two social situations: meeting a new neighbor and persuading a landlord to fix a bathroom leak. Scenes were audio-recorded and rated by a blind-to-diagnosis, expert coder involved in all previous ratings for SCOPE.

The mean score across both role-plays was used as the dependent measure, with a possible range from 1 to 5.

Symptomatology

The Positive and Negative Syndrome Scale (*PANSS*) is an interview-based assessment that evaluates symptom severity for people with psychosis (Kay, Opler, & Fiszbein, 1997). The rating scale, comprised of three sub-scales, measures positive symptoms, negative symptoms and symptoms of general psychopathology.

Feasibility

Practicality was operationalized as administration time, which was measured using an electronic stopwatch. Research assistants were instructed to begin timing immediately upon initiating a social cognitive task, and to discontinue timing upon completion of the test.

To assess tolerability, participants were asked to rate each measure on a Likert-type scale that ranged from 1 (*very unpleasant*) to 7 (*very pleasant*) immediately following the completion of each social cognitive measure. Ratings of 4 indicated a neutral response of neither pleasant nor unpleasant.

Procedure

Participants completed two assessments: baseline and a retest assessment scheduled to occur approximately 4 weeks later. With the project approved by the UNC-CH Institutional Review Board, participants provided signed informed consent and completed social cognitive, neurocognitive, and functional outcome measures at baseline. Task block order and the order of individual tasks within the social cognitive battery were counterbalanced. A rater trained using the same procedures employed in SCOPE conducted diagnostic and symptomatic interviews. Symptoms were reassessed in patients at retest. With the exception of TASIT, for which an

alternative form was available, identical social cognitive tasks were administered at retest, in a different counterbalanced order. In accord with the original SCOPE protocol, Version-A was administered at baseline, Version-B at retest.

Statistical Analyses Overview

Data analyses were performed using the Statistical Package for the Social Science (SPSS) version 23. Statistical significance was defined as $p < .05$.

We followed the psychometric validation process employed for the initial validation study (Pinkham et al., 2015). Score distributions of the social cognitive measures were inspected for normality by examining skew and kurtosis statistics and visually inspecting histograms. Though no measures were transformed, one participant in the clinical group was an outlier on the BLERT (> 3SD from the mean). Consequently, these data were excluded from further analyses.

Test-Retest Reliability and Internal Consistency

Pearson's *r* correlation coefficients were calculated for each social cognition measure to examine test-retest reliability. Pearson's *r* values greater than or equal to 0.6 for test-retest reliability were considered adequate (Kraemer et al., 2012; Pinkham et al., 2015). To determine the extent to which items within a measure are related, Cronbach's alpha was used to determine internal consistency for each task. Alpha values greater than or equal to 0.8 were regarded as acceptable (Nunnally & Bernstein, 1994).

Between Group Differences

To directly determine the extent to which phase and clinical group membership contributed to performance on each social cognitive measure, we conducted a 2x2 between-subjects factorial ANOVA. Group (clinical sample or healthy controls) served as one factor and phase (chronic or first episode study) was included as the second independent variable. The

influences of participant age and estimated IQ were controlled for in the analyses. Consistent with Pinkham et al., subjects were removed from the analyses if they were identified as outliers in their respective group, or if a subject did not complete a particular task. This process was carried out separately for each candidate measure in the SCOPE battery.

Post hoc analyses probing interactions and main effects involved conducting pairwise comparisons using the Bonferroni correction method. Cohen's *d* values were used to measure the magnitude of differences in performance between FEP patients, chronic patients, and both demographically matched control groups. Effect sizes were evaluated according to the ranges recommended by Cohen (1988): small (d=0.2), medium (d=0.5), and large (d=0.8).

Utility as a Repeated Measure

The clinical utility of these tasks as repeated outcome measures was analyzed by assessing evidence of practice effects and any indication of floor/ceiling effects. As participants completed each task twice, we conducted paired-samples t-tests and calculated Cohen's d_z values. Effect sizes indicating statistically significant improvement (Cohen's d_z values greater than .30) between the first and second administrations of a particular task were considered suggestive of clinically relevant practice effects. The number of participants scoring at or below chance levels operationalized floor effects. The number of participants earning a perfect score on a particular task operationalized ceiling effects.

Convergent and Discriminant Validity

Convergent validity was examined to determine the extent to which the eight social cognitive tasks were related. We calculated correlations between participants' total scores on the social cognitive tasks completed at baseline. We reported statistically significant associations between measures, with an emphasis on moderate to strong correlations between tasks purported

to assess capabilities in the same domain of social cognition. To examine discriminant validity, we reported on correlations between participants' total scores on the social cognition measures and composite score on the MATRICS Consensus Cognitive Battery (MCCB) administered at baseline.

Relationship to Social and Occupational Functioning

To investigate the relationship between social cognition and functional outcome, we computed correlations between social cognitive capacity and performance on functional outcome measures, including: the UCSD Performance-Based Skills Assessment (UPSA), the Social Skills Performance Assessment (SSPA), and the Specific Level of Functioning (SLOF) scale.

Incremental validity – Variance in Functioning beyond Neurocognition

To assess incremental validity beyond neurocognitive abilities, we constructed a regression model in blocks. The first block consisted of the composite score on the MATRICS Consensus Cognitive Battery (MCCB). To examine the increase in predictive power of social cognition, a second block was created that included all social cognitive measures shown to be significantly related to each functional outcome measure. We used *R*-squared and adjusted *R*-squared statistics to determine the extent to which the addition of social cognition accounted for variance in functional outcome beyond neurocognition.

Feasibility – Practicality of Administration and Tolerability

Descriptive statistics were used to assess practicality of administration and tolerability for patients. We provided mean and standard deviation values for practicality, operationalized as administration time, including instruction review and task completion; and tolerability, operationalized as self-report ratings of enjoyability on a Likert-type scale from 1 (*very unpleasant*) to 7 (*very pleasant*).

Decision-making Process Regarding SCOPE Battery for First Episode Psychosis

To determine the suitability of the SCOPE battery for FEP, we emphasized test-retest reliability, relationship to functional outcome, and ability to distinguish patient and control performance. We followed a similar process to Pinkham et al. (2015) such that we consulted with the principal investigators of SCOPE before making final recommendations. We also used the following groupings to meaningfully classify candidate measures into three disparate categories:

Acceptable at Present signifies the measure displayed acceptable reliability and validity in the current study, and would not require modifications before use. To be considered acceptable for implementation in clinical trials at present, a measure displayed adequate test-retest reliability, demonstrated a significant relationship with at least one functional outcome measures, and effectively distinguished between patients and controls.

Acceptable with Concerns indicates specific attributes of the task were concerning and warrant further investigation before implementation in clinical research. Areas of concern included failure to distinguish early psychosis patients and nonclinical controls, as well as intolerability for patients. Although feasibility is an important indicator of the utility of a measure, a task was not precluded from recommendation solely for requiring additional time to administer and/or for being less enjoyable provided that the measure also contributed valuable information.

Currently Unacceptable signifies a task displayed weak psychometrics overall and cannot be recommended for use in clinical research with this population. Inadequacies included poor test-retest reliability and/or limited relation to functional outcome.

RESULTS

Participants

Thirty-eight FEP patients and 39 age- and gender-matched controls completed the baseline assessment. Thirty-five patients and 36 controls returned to complete visit two. Average time between administrations was comparable for both groups (M_{FEP}=33.08 days, SD=5.65; M_{HC}=31.61 days, SD=4.81; $t(70)$=-1.190, P=.238). FEP patients and controls did not differ in regard to gender, race, ethnicity, age, or estimated IQ (see Tables 1a and 1b). FEP patients completed significantly fewer years of education than controls, whereas patients' parents completed significantly more years of education than the control sample.

Separate 2x2 factorial ANOVAs were conducted to examine differences demographic, clinical and cognitive characteristics across studies. Results revealed a significant main effect of Phase on age (F (1, 356) = 131.02, $p < 0.001$) and IQ (F (1, 356) = 45.56, $p < 0.001$). Regarding education, analyses also revealed significant main effects of Phase (F (1, 356) = 45.47, $p < 0.001$) and Group (F (1, 356) = 17.37, $p < 0.001$). Participants involved in SCOPE-FEP were significantly younger, better educated, and had higher IQ than participants in the original SCOPE study (Tables 1a and 1b). Educational attainment also varied across groups, with controls ($M_{edu\text{-}hc}$ = 14.0 years, SD = 1.7, n = 143) completing significantly greater years of education than patients ($M_{edu\text{-}scz\&fep}$ = 12.9 years, SD = 1.8, n = 217).

No statistically significant Phase x Group interactions were found for age (F (1, 356) = 0.71, $p > .05$), IQ (F (1, 356) = 0.01, $p > .05$), or education (F (1, 356) = 2.21, $p > .05$).

An additional 2x2 factorial ANOVA examining patient symptoms across phase (chronic

or first episode study) and assessment visit (baseline or retest) was conducted. A significant main effect was found for phase (F (1, 215) = 12.77, $p < 0.001$), but not assessment type (F (1, 206) = 1.72, $p > .05$). Post hoc analyses revealed FEP patients presented with more symptoms than chronic patients at baseline and retest (Tables 1a and 1b). Significant relationships between symptoms and social cognition were found for only two candidate measures: the TASIT (r = -.156, $p < .05$) and Trust task (r = -.149, $p < .05$).

Test-Retest Reliability

Hinting, RAD, and AIHQ (BS) demonstrated acceptable levels of test-retest reliability (Pearson's r values ≥ 0.6) for FEP patients. BLERT, ER-40, Eyes, RAD, and two AIHQ subscales (AB/BS) showed adequate values among younger controls (Table 2).

Internal Consistency

For FEP patients, few candidate measures approached/exceeded acceptable Cronbach's alpha values ($\alpha \geq 0.8$). Exceptions included Trust (.943), TASIT (.795) and AIHQ-BS (.857). Internal consistency was generally lower for controls, with values for all tasks below target standards (Table 2).

Group Differences

Significant main effects of Phase (chronic or first episode study) were found for the Eyes, Hinting, RAD, TASIT and Trust tasks (Table 3b). Controlling for age and estimated IQ, patient and control participants in SCOPE-FEP outperformed clinical and non-clinical participants in the original SCOPE study on the Eyes, Hinting, RAD, TASIT and Trust tasks. Results also revealed significant main effects of Group (patients or controls) for the AIHQ (HB/BS), BLERT, ER40, Eyes, Hinting, RAD, and TASIT (Table 3b). With the exception of the Trust task and one subscale of the AIHQ (AB), control participants outperformed patients on all measures in the

candidate battery when controlling for relevant covariates. Findings indicated statistically significant Phase x Group interactions for several tasks, including: the AIHQ (BS), BLERT, Eyes and RAD (Table 3b). All *F*-statistics and mean performance values have been summarized in Tables 3a and 3b.

Post hoc pairwise comparisons were conducted using a Bonferroni correction to further examine statistically significant main effects and interactions. Results revealed that chronic patients scored significantly worse on the AIHQ (BS), BLERT, Eyes and RAD than first episode patients and both control groups when controlling for the effects of age and estimated IQ. Scores on the AIHQ (BS), BLERT, Eyes and RAD for the remaining three groups did not significantly differ from one another. Additional analyses revealed performance differences between FEP patients and their demographically matched control group on only two measures: the Hinting and TASIT. Similarly, performance between FEP patients and the control sample from SCOPE significantly differed on the ER40 only.

Effect sizes (Cohen's *d*) indicating magnitude of differences between chronic patients and the remaining three groups were moderate-to-large for two subscales of the AIHQ (HB/BS) (range: d = .65 to .90), BLERT (range: d = .76 to 1.43), ER40 (range: d = .71 to .94), Eyes (range: d = .67 to 1.33), and Hinting (range: d = .66 to 1.37); and large for the RAD (range: d = .93 to 1.85) and TASIT (range: d =1.05 to 2.12). Similarly, performance differences between FEP patients and their demographically-matched control group were large for both the Hinting (d = .81) and TASIT (d = 1.04). Though significant, the magnitude of the difference between FEP patients and SCOPE controls on the ER40 was small (d = .21).

Utility as a Repeated Measure

For FEP patients, three of eight measures demonstrated statistically significant differences between assessments (Tables 4a and 4b). Patient performance on ER-40 and Hinting improved, whereas TASIT performance worsened from baseline to retest. Effect sizes were moderate (Cohen's d_z range: .414-.642). Compared to the initial psychometric evaluation (Pinkham et al., 2015), floor/ceiling effects were less evident for this sample. A maximum of two patients (<6%) received perfect or chance-level scores on any measure.

Alternatively, younger controls performed significantly better on BLERT, ER-40, RAD, and AIHQ-BS at retest. Similar to FEP patients, control performance across versions of the TASIT worsened significantly from baseline to retest. Practice effects varied, with effect sizes in the small-to-medium range (Cohen's d_z range: .212-.732). Only one control scored at/below chance levels on any task (Eyes) during either visit.

With the exception of the second administration of BLERT and first administration of Hinting, <8% of controls scored at ceiling for any candidate measure. Five (12.8%) received perfect scores on BLERT (visit 2), whereas four (10.3%) scored at ceiling on Hinting (visit 1).

Convergent and Divergent Validity

Correlations between candidate measures of social cognition, as well as associations between these tasks and the neurocognitive battery when administered to a younger sample have been provided in Tables 5a and 5b. With the exception of the Eyes task, remaining tasks demonstrated significant relationships with one or more additional measures in the SCOPE battery. The BLERT demonstrated the greatest number of relationships to other measures in the battery (Tables 5a and 5b). Significant associations were all in the expected direction, and of small-to-medium magnitude (range: r=.327-.527). The Eyes task was also the only measure to demonstrate a significant relationship with neurocognition (r=.327).

Relationship to Functional Outcome

Correlations between social cognitive and neurocognitive tasks, and functional outcome measures for FEP are provided in Table 6. With the exception of BLERT, Trust, and AIHQ-AB, most measures demonstrated significant relationships with one or more outcome measures. Significant associations were in the expected direction, and of medium magnitude (range: r=.344-.475).

Incremental Validity

Social cognition explained additional variance in functional outcome above and beyond neurocognition. Simple linear regression analyses indicated neurocognition, when entered as a single predictor variable, accounted for 22% of variance in UPSA-B total scores (adjusted R^2 = .218, $F(1,36) = 11.33$, $P < .01$) and 12% of variance in SSPA ratings (adjusted R^2 = .123, $F(1,35) = 6.039$, $P < .05$), but was not a significant predictor of SLOF self-report values (adjusted R^2 = -.012, $F(1,36) = .550$, $P > .05$) (Table 7). Sequential regression analyses revealed social cognition, entered after neurocognition as a second block, accounted for an additional 20% of variance in community living skills (UPSA-B; R^2 change = .199, $P < .05$), 19% of variance in social skills (SSPA; R^2 change = .193, $P < .05$), and 21% of variance in real-world functioning (SLOF-SR; R^2 change = .214, $P < .01$) (Table 7).

Practicality and Tolerability

Excluding BLERT administration time ($t(75) = 5.78$, $P = .019$, $d = .499$) and TASIT enjoyability ratings ($t(74) = 5.06$, $P = .027$, $d = -.379$), practicality and tolerability did not differ significantly between FEP patients and controls (Table 8). Most measures required <8 minutes to complete. Participants rated all tasks as relatively pleasant (range: M = 4.29 - 5.62).

Recommendations

Regarding suitability for FEP, Hinting was the sole measure to be considered **Acceptable at Present**. RAD was categorized as **Acceptable with Concerns**. Remaining candidate measures (AIHQ, BLERT, ER40, Eyes, TASIT, Trust) were regarded as **Currently Unacceptable**. A comparison between our recommendations and the outcome of the initial psychometric evaluation has been provided in Table 9. A discussion of convergence and divergence between studies is provided below.

DISCUSSION

The current study evaluated the psychometric properties of the SCOPE battery for first episode psychosis. Our findings suggest one measure, the Hinting task, was considered **Acceptable at Present**, or appropriate for use with FEP patients. In addition to displaying adequate test-retest reliability and effectively distinguishing between patients and controls, the Hinting task also exhibited significant relationships with both performance-based measures of functioning.

The RAD was classified as **Acceptable with Concerns** and may be cautiously considered for use with FEP. This measure demonstrated adequate test-retest reliability, a significant relationship to functioning, minimal practice effects, and limited floor/ceiling effects. However, this task was one of the longest to administer and was rated the least enjoyable by patients and controls. RAD's failure to distinguish patients from controls also tempers enthusiasm for this measure. Modification efforts to develop a shortened version may prove beneficial.

The remaining candidate measures were deemed **Currently Unacceptable** for use with FEP and warrant careful consideration if employed in future clinical trials. Though relatively quick and easy to administer, the BLERT and Trust displayed the weakest psychometric properties overall, including poor test-retest reliability, failure to differentiate individuals with/without psychosis, and limited relation to functioning. For the ER40 and TASIT, the primary concern was inadequate test-retest reliability. Based on moderate and significant practice effects observed for the TASIT, there was also concern about possible interference or

non-equivalence between versions.

Though AIHQ and Eyes demonstrated significant associations with real-world functioning, predominantly low test-retest reliability estimates and inability to distinguish patients from controls precluded these tasks from recommendation. Notably, however, one subscale of the AIHQ, the BS, was strong on all metrics except group differences. Prior research indicates this subscale of the AIHQ demonstrates adequate psychometric properties, including acceptable internal consistency and test-retest reliability estimates; distinguishes patients from controls; displays significant relationships to functional outcome, and exhibits associations with relevant clinical variables in chronic samples (e.g., hostility and suspiciousness) (Buck et al., 2017, in press). The AIHQ-BS may therefore benefit from further examination and use; however it will be important to determine if it can be used independently of the rest of the measure.

Note that measures were required to demonstrate adequate test-retest reliability among patients, distinguish patients from controls, and exhibit significant relationships to functioning in order to be recommended for use in clinical trials targeting social cognition in FEP. Given the small sample size, we suggest careful consideration of these recommendations and thoughtful interpretation of the present findings. In particular, our recommendations may not be as applicable to other research goals (e.g., cross-sectional designs).

Consistent with Pinkham et al., our data demonstrate that the Hinting task is a psychometrically valid theory of mind measure that should be considered appropriate for implementation in psychosis research regardless of stage of illness. Importantly, both Pinkham et al. and the present study utilized a more stringent scoring manual. We emphasize the reported psychometric properties as limited to this revised scoring system (available from AEP upon

request). Analyses are underway to determine whether the psychometric properties of the task may change if the original scoring criteria are utilized.

Also consistent with the original SCOPE study, our findings substantiate the claim that social cognition accounts for more variance in functional outcome than various cognitive factors. When measures of social cognition were included in the analyses, the explanatory power of neurocognition dropped significantly. These findings corroborate previous research suggesting social cognition mediates the relationship between neurocognition and functioning in psychosis (Fett et al., 2011; Schmidt, Mueller & Roder, 2011). Together, findings from this study provide strong support for the importance of social cognition in FEP.

In contrast, our findings diverge from the initial psychometric evaluation in a number of ways. Although the BLERT displayed some of the strongest properties in SCOPE, it was one of the weakest measures when administered to FEP outpatients. Whereas only two AIHQ subscales (HB/AB) showed inadequate test-retest reliability among patients in SCOPE, the Hinting, RAD and one subscale of the AIHQ (BS) were the only measures to reach acceptable levels when administered to FEP. Reliability estimates were generally lower for controls than patients in SCOPE, while the opposite was observed in our sample. In addition, excluding one AIHQ subscale (AB), all social cognitive tasks adequately differentiated between clinical and normative groups in Pinkham et al. Alternatively, significant group differences were observed for only two measures when administered to a younger sample.

Certain procedural incongruences and sample differences between our study and the original SCOPE study may have contributed to lower test-retest reliability estimates, differential sensitivity to group differences, and limited relationship to functional outcome. Effect sizes indicating meaningful changes in performance between visits suggested clinically relevant

practice effects for half the battery when administered to a younger sample (Table 4b). Memory and practice effects have been shown to adversely affect test-retest reliability (Abner et al., 2012; Broglio et al., 2007; Greig et al., 2004). In fact, educational attainment and general intelligence for our patient and control samples were significantly higher than chronic and control participants in SCOPE. Specific differences between our clinical sample and that of SCOPE may explain why measures did not reliably differentiate patients and controls, and clarify the limited value of most tasks as independent predictors of functional outcome. It is also possible that the outcome measures used to assess social functioning and daily living skills may be inappropriate for younger patients.

Psychosis onset typically occurs during late adolescence and early adulthood, a period of developmental transition and social/lifestyle changes that may contribute to less stable social cognition early in the course of illness (Horan et al., 2012). To assess the possibility that changes in symptom severity between visits may have impacted social cognitive performance, we recalculated test-retest correlations controlling for symptom fluctuations. Values were unchanged, thus indicating it is unlikely symptom variability accounted for lower test-retest reliability estimates.

Directly comparing performance between chronic and first episode patient samples revealed superior social cognitive abilities for individuals earlier in the course of illness. FEP patients outperformed chronic patients on the majority of tasks, including the AIHQ (BS), Eyes, Hinting, RAD and TASIT. Chronic and FEP patients performed similarly on the remaining tasks, including two subscales of the AIHQ subscales (HB/AB), BLERT, ER40 and Trust task. When the effects of age and general intelligence were controlled for in the analyses, individuals with first episode psychosis exhibited more severe social cognition deficits than matched

controls on only two measures of social cognition, the Hinting and TASIT. These findings suggest social cognitive deficits may be less prominent early in the course of illness, are potentially mitigated by higher general intelligence, and/or worsen with age. These findings support prior research indicating theory of mind deficits may precede the onset of psychosis and should be considered a vulnerability marker of psychosis (Bora & Pantelis, 2013). It is also possible that using total scores, particularly for measures on which items are scored dichotomously for correct/incorrect responses, may be scaled in units too gross to detect true, though subtler social cognition deficits exhibited by FEP patients in other domains of social cognition (Browne et al., 2016; Miller, 2016). Additional research is needed to clarify the extent to which abilities may change over time. Future studies should also include dual-normative comparison groups, carefully consider the impact of relevant covariates on social cognitive performance, and longitudinally compare patient performance at different stages of illness.

An examination of convergent and divergent validity produced several interesting findings. First, we found a significant, moderate-sized correlation between the Eyes task and the neurocognition battery ($d = .327$, $p < .05$). The Eyes task was also shown to be unrelated to other candidate measures of social cognition when administered to a younger sample. These findings echo a concern raised by SCOPE investigators and expert consultants indicating performance on this task may rely on vocabulary and reading level (Pinkham et al., 2015). Thus, it is unclear whether the Eyes test truly measures theory of mind in psychosis, or if it merely functions as an additional measure of neurocognition.

Similarly, the BLERT was associated with tasks purported to assess various domains of social cognition. These findings align closely with prior research examining the factor structure of social cognition in schizophrenia spectrum disorders (Buck, Healey, Gagen, Roberts & Penn,

2016). Specifically, research suggests emotion processing and less complex theory of mind abilities tend to be correlated and comprise the "lower-level cue detection" factor of social cognition in psychosis (Buck et al., 2016; Mancuso et al., 2011). Although Pinkham et al. did not examine the extent to which the eight social cognitive tasks were related, nor did they present correlations between the SCOPE battery and neurocognition, it is unclear whether these interpretations can be appropriately translated to chronic schizophrenia research.

Finally, certain limitations must be considered. First, the inclusion of a relatively small sample, especially compared to the original psychometric evaluation, is a noteworthy limitation of the present study. Additionally, data were collected from a relatively homogenous sample of predominantly white, well-educated males from one of the fastest-growing metropolitan areas in the United States. FEP patients were recruited from a coordinated specialty care clinic focused on early intervention and recovery in Chapel Hill, a small university town; and may qualitatively differ from clinical samples recruited from more traditional community mental health centers in Dallas and Miami, both large urban areas. Given the various demographic, cognitive, clinical and regional differences between our sample and participants involved in SCOPE, interpretations of the present findings should be regarded with caution.

In summary, the present study indicates social cognitive assessment needs to be approached differently for individuals early in the course of illness, and investigators should use caution when employing tasks that have been used primarily with chronic samples. This underscores the need for the development of new measures for use with FEP, as well as a better understanding of how social cognition and functioning may differ across stage of illness. In addition to improving the validity, reproducibility, and comparability of research findings, we may use this information to tailor treatment and develop targeted interventions for FEP.

REFERENCES

Addington, J, & Piskulic, D. Social Cognition Early in the Course of the Illness. In: Roberts D., Penn D. (Eds.), Social Cognition in Schizophrenia. Oxford University Press, New York, pp. 245-262.

Addington, J, Saeedi, H, & Addington, D (2006). Influence of social perception and social knowledge on cognitive and social functioning in early psychosis. *Br. J. Psychiatry*, 189 (2006), pp. 373–378.

Adolphs, R., Tranel, D., & Damasio, A. R. (1998). The human amygdala in social judgment. *Nature*. 393(6684), 470–474.

Barkl, SJ, Lah, S, Harris, A, & Williams, LM. (2014). Facial emotion identification in early-onset and first-episode psychosis: A systematic review with meta-analysis. *Schizophrenia Research*, 159(1), 62-69.

Baron-Cohen S., Wheelwright S., Hill J., Raste Y., & Plumb I. The 'Reading the Mind in the Eyes' Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *J Child Psychol Psychiatry*, 42, 241–251.

Bell, M., Bryson, G., & Lysaker, P. (1997). Positive and negative affect recognition in schizophrenia: A comparison with substance abuse and normal control subjects. *Psychiatry Research*, *73*(1-2), 73-82.

Bliksted, V, Fagerlund, B, Weed, E, Frith, C, & Videbech, P (2014). Social cognition and neurocogntive deficits in first-episode-schizophrenia. Schizophrenia Research, 153(1-3), 9-17.

Bora, E., & Pantelis, C. (2013). Theory of mind impairments in first-episode psychosis, individuals at ultra-high risk for psychosis and in first-degree relatives of schizophrenia: Systematic review and meta-analysis. *Schizophrenia Research*, 144(1-3), 31-36.

Bora, E., Yucel, M., & Pantelis, C. (2009). Theory of mind impairment in schizophrenia: meta-analysis. *Schizophrenia Research*, 109(1–3), 1–9.

Bowie, C.R., Grossman, M., Gupta, M., Oyewumi, L.K., & Harvey, P.D. (2014). Cognitive remediation in schizophrenia: efficacy and effectiveness in patients with early versus long-term course of illness. *Early Interv. Psychiatry,* 8(1), 32–38.

Brüne, M., Abdel-Hamid, M., Lehmkamper, C., & Sonntag, C. (2007). Mental state attribution, neurocognitive functioning, and psychopathology: what predicts poor social competence in schizophrenia best? *Schizophrenia Research*, 92, 151–159.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Erlbaum; Hillsdale, NJ. Cohen, J. D., Braver, T. S., & O'Reilly, R. C. (1998). A computational

approach to prefrontal cortex, cognitive control, and schizophrenia: Recent developments and current challenges. In A. C. Roberts, T. W. Robbins, L. Weiskrantz, A. C. Roberts, T. W. Robbins, L. Weiskrantz (Eds.), *The prefrontal cortex: Executive and cognitive functions* (pp. 195-220). New York, NY, US: Oxford University Press.

Combs, D. R., Penn, D. L., Wicher, M., & Waldheter, E. (2007). The ambiguous intentions hostility questionnaire (AIHQ): A new measure for evaluating hostile social-cognitive biases in paranoia. *Cognitive Neuropsychiatry*, 12(2), 128-143.

Corcoran R., Mercer G., & Frith C. D. (1995). Schizophrenia, symptomatology and social inference: investigating "theory of mind" in people with schizophrenia. *Schizophrenia Research*, *17*, 5–13.

Corigliano, V., De Carolis, A., Trovini, G., Dehning, J., Di Pietro, S., Curto, M., ... & Comparelli, A. (2014). Neurocognition in schizophrenia: From prodrome to multi-episode illness. *Psychiatry research*, *220*(1), 129-134.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, 10*(7), 1-9.

Couture, S. M., Penn, D. L., Losh, M., Adolphs, R., Hurley, R., & Piven, J. (2010). Comparison of social cognitive functioning in schizophrenia and high functioning autism: more convergence than divergence. *Psychological Medicine*, *40*(4), 569–579.

Couture, S. M., Penn, D. L., & Roberts, D. L. (2006). The functional significance of social cognition in schizophrenia: a review. *Schizophrenia bulletin*, *32*(suppl 1), S44-S63.

Couture S, & Penn D. Introduction to social cognition in schizophrenia. In: Roberts D., Penn D. (Eds.), Social Cognition in Schizophrenia. Oxford University Press, New York, pp. 1-16.

Drost EA. Validity and Reliability in Social Science Research. *Education Research and Perspectives*, 38(1): 105-123.

Fett, AJ, Viechtbauer W, Dominguez M, Penn DL, van Os J, Krabbendam L. (2011). The relationship between neurocognition and social cognition with functional outcomes in schizophrenia: A meta-analysis. *Neuroscience and Biobehavioral Reviews*, 35, 573-588.

First, M. B., Spitzer, R. L., Gibbon M., & Williams, J. B.W. (2002). Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Patient Edition. (SCID-I/P) New York: Biometrics Research, New York State Psychiatric Institute.

Frith, S. (2008). Social cognition. Philosophical Transactions of the Royal Society, 363(1499): 2033-2039.

Furr, R.M., & Bacharach, V.R. (2008). *Psychometrics: an introduction*. Los Angeles: Sage Publications.

Fusar-Poli, P., Smieskova, R., Kempton, M. J., Ho, B. C., Andreasen, N. C., & Borgwardt, S. (2013). Progressive brain changes in schizophrenia related to antipsychotic treatment? A meta-analysis of longitudinal MRI studies. *Neuroscience And Biobehavioral Reviews*, *37*(8), 1680-1691.

Green, M. F., Bearden, C. E., Cannon, T. D., Fiske, A. P., Hellemann, G. S., Horan, W. P., & ... Nuechterlein, K. H. (2012). Social cognition in schizophrenia, part 1: Performance across phase of illness. *Schizophrenia Bulletin*, *38*(4), 854-864.

Green, M. F., Hellemann, G. S., Horan, W. P., Lee, J., & Wynn, J. K. (2012). *Arch Gen Psychiatry, 69*(12): 1216–1224.

Green, M. F., & Leitman, D. I. (2008). Social cognition in schizophrenia. *Schizophrenia Bulletin*, 34(4), 670–672.

Heinrichs, R. W., & Zakzanis, K. K. (1998). Neurocognitive deficit in schizophrenia: A quantitative review of the evidence. *Neuropsychology, 12,* 426–445.

Ho, B.C., Andreasen, N.C., Ziebell, S., Pierson, R., & Magnotta, V. (2011). Long-term Antipsychotic Treatment and Brain Volumes: A Longitudinal Study of First-Episode Schizophrenia. *Archives of General Psychiatry*, *68*(2), 128–137.

Hooker, C. I., Bruce, L., Lincoln, S. H., Fisher, M., & Vinogradov, S. (2011). Theory of mind skills are related to gray matter volume in the ventromedial prefrontal cortex in schizophrenia. *Biological Psychiatry*, *70*(12), 1169–1178.

Horan, W. P., Green, M. F., DeGroot, M., Fiske, A., Hellemann, G., Kee, K., & ... Nuechterlein, K. H. (2012). Social cognition in schizophrenia, part 2: 12-month stability and prediction of functional outcome in first-episode patients.

Kay, S. R., Flszbein, A., & Opfer, L. A. (1987). The Positive and Negative Syndrome Scale (PANSS) for Schizophrenia. *Schizophrenia Bulletin*, *13*(2), 261-276.

Kern, R. S., Nuechterlein, K. H., Green, M. F., Baade, L. E., Fenton, W. S., Gold, J. M., ... & Stover, E. (2008). The MATRICS Consensus Cognitive Battery, part 2: co-norming and standardization. *American Journal of Psychiatry*.

Kohler, C. G., Walker, J. B., Martin, E. A., Healey, K. M., & Moberg, P. J. (2009). Facial emotion perception in schizophrenia: a meta-analytic review. *Schizophrenia bulletin*, 192.

Kohler C. G., Turner T. H., Bilker W. B., Brensinger, C. M., Siegel, S. J., Kanes, S. J., Gur, R.E., & Gur, R. C. (2003). Facial emotion recognition in schizophrenia: intensity effects and error pattern. *Am J Psychiatry*,160, 1768–1774.

Krstev, H., Jackson, H., & Maude, D. (1999). An investigation of attributional style in first-episode psychosis. *British Journal Of Clinical Psychology*, *38*(2), 181-194.

Kunda, Z. (1999). Social cognition: Making sense of people. MIT Press; Cambridge, MA.

Kurtz, M. M., & Mueser, K. T. (2008). A Meta-Analysis of Controlled Research on Social Skills Training for Schizophrenia. *Journal of Consulting and Clinical Psychology*, 76(3), 491–504.

Lee, R.S., Redoblado-Hodge, M.A., Naismith, S.L., Hermens, D.F., Porter, M.A., & Hickie, I.B. (2013). Cognitive remediation improves memory and psychosocial functioning in first-episode psychiatric outpatients. *Psychol. Med, 43(6)*, 1161–1173.

Mancuso, F., Horan, W. P., Kern, R. S., & Green, M. F. (2011). Social cognition in psychosis: multidimensional structure, clinical correlates, and relationship with functional outcome. *Schizophrenia research*, *125*(2), 143-151.

Marchesi, C., Affaticati, A., Monici, A., De Panfilis, C., Ossola, P., Ottoni, R., & Tonna, M. (2015). Decrease of functioning in remitted and non-remitted patients 16 years after a first-episode schizophrenia. *Journal Of Nervous And Mental Disease*, *203*(6), 406-411.

Martens, L., & Baker, S. (2009). *Promoting Recovery from First Episode Psychosis: A guide for families*. Toronto, Canada: Center for Addition and Mental Health.

Mausbach, B. T., Harvey, P. D., Goldman, S. R., Jeste, D. V., & Patterson, T. L. (2007). Development of a brief scale of everyday functioning in persons with serious mental illness. *Schizophrenia Bulletin*, *33*(6), 1364-1372.

McCleery, A., Ventura, J., Kern, R. S., Subotnik, K. L., Gretchen-Doorly, D., Green, M. F., & ... Nuechterlein, K. H. (2014). Cognitive functioning in first-episode schizophrenia: MATRICS Consensus Cognitive Battery (MCCB) profile of impairment. *Schizophrenia Research*, *157*(1-3), 33-39.

McDonald S., Flanagan S., Rollins J., & Kinch J. (2003). TASIT: a new clinical tool for assessing social perception after traumatic brain injury. *J Head Trauma Rehabil, 18*, 219–238.

Mesholam-Gately, R. I., Giuliano, A. J., Goff, K. P., Faraone, S. V., & Seidman, L. J. (2009). Neurocognition in first-episode schizophrenia: A meta-analytic review. *Neuropsychology*, *23*(3), 315-336.

Mueser, K.T., Deavers, F., Penn, D.L., & Cassisi, J.E. (2013). Psychosocial treatments for schizophrenia. *Annual Review of Clinical Psychology*, *9*, 465–497.

Nuechterlein, K. H., Green, M. F., Kern, R. S., Baade, L. E., Barch, D. M., Cohen, J. D., ... & Goldberg, T. (2008). The MATRICS Consensus Cognitive Battery, part 1: test selection, reliability, and validity. *The American journal of psychiatry*, *165*(2), 203-213.

Patterson, T. L., Moscona, S., McKibbin, C. L., Davidson, K., & Jeste, D. V. (2001). Social skills performance assessment among older patients with schizophrenia. *Schizophrenia research*, *48*(2), 351-360.

Penn, DL, Corrigan, PW, Bentall, RP, Racenstein, JM, & Newman, L. (1997). Social cognition in schizophrenia. *Psychol Bull*, 121114–132.

Perkins, D.O., Gu, H., Boteva, K., & Lieberman, J.A. (2005). Relationship between duration of untreated psychosis and outcome in first-episode schizophrenia: A critical review and meta-analysis. *The American Journal Of Psychiatry*, *162*(10), 1785-1804.

Pijnenborg, G.H.M., Withaar, F.K., Evans, J.J., Van den Bosch, R.J., Timmerman, M.E., & Brouwer, W.H. (2009). The predictive value of measures of social cognition for community functioning in schizophrenia: implications for neuropsychological assessment. *J Int Neuropsychol Soc*, 15, 239–247.

Pinkham AE, Penn DL, Green MF, Buck B, Healey K, Harvey PD (2014). The Social Cognition Psychometric Evaluation study: results of the expert survey and RAND panel. *Schizophrenia Bulletin, 40*(4): 813-823.

Pinkham AE, Penn DL, Green MF, Harvey PD (2015). Social Cognition Psychometric Evaluation: Results of the Initial Psychometric Study. *Schizophrenia Bulletin*.

Prevost, M., Brodeur, M., Onishi, K. H., Lepage, M., & Gold, I. (2015). Judging Strangers' Trustworthiness is Associated with Theory of Mind Skills. *Frontiers in Psychiatry*, *6*, 52.

Priebe, S (2007). Social Outcomes in Schizophrenia. *British Journal of Psychiatry*, *191*(50): s15-s20.

Rabinovitch, M., Béchard-Evans, L., Schmitz, N., Joober, R., & Malla, A. (2009). Early predictors of nonadherence to antipsychotic therapy in first episode psychosis. *The Canadian Journal Of Psychiatry / La Revue Canadienne De Psychiatrie*, *54*(1), 28-35.

Radua, J., Borgwardt, S., Crescini, A., Mataix-Cols, D., Meyer-Lindenberg, A., McGuire, P. K., & Fusar-Poli, P. (2012). Multimodal meta-analysis of structural and functional brain changes in first episode psychosis and the effects of antipsychotic medication. *Neuroscience And Biobehavioral Reviews*, *36*(10), 2325-2333.

Revell, E.R. Neill, J.C., Harte, M., Khan, Z., & Drake, R.J. (2015). A systematic review and meta-analysis of cognitive remediation in early schizophrenia. Schizophrenia Research, *168*, 213–222.

Romero-Ferreiro, M. V., Aguado, L., Rodriguez-Torresano, J., Palomo, T., Rodriguez-Jimenez, R., & Pedreira-Massa, J. L. (2016). Facial affect recognition in early and late-stage schizophrenia patients. *Schizophrenia Research*, 172(1-3), 177-183.

Savla G, Vella L, Armstrong C., Penn D, & Twamley E. (2013). Deficits in domains of social cognition in schizophrenia: A meta-analysis of the empirical evidence. *Schizophrenia Bulletin*, 39, 979–992.

Schmidt, S. J., Mueller, D. R., & Roder, V. (2011). Social cognition as a mediator variable between neurocognition and functional outcome in schizophrenia: empirical review and new results by structural equation modeling. *Schizophrenia Bulletin*, *37*(suppl 2), S41-S54.

Schneider, L. C., & Struening, E. L. (1983). SLOF: a behavioral rating scale for assessing the mentally ill. In *Social Work Research and Abstracts* (Vol. 19, No. 3, pp. 9-21). Oxford University Press.

Segarra, R., Ojeda, N., Peña, J., García, J., Rodriguez-Morales, A., Ruiz, I., & ... Gutiérrez, M.(2012). Longitudinal changes of insight in first episode psychosis and its relation to clinical symptoms, treatment adherence and global functioning: One-year follow-up from the Eiffel study. *European Psychiatry*, *27*(1), 43-49.

Sergi, M. J., Fiske A. P., Horan W. P., Kern, R. S., Kee, K. S., Subotnik, K. L., Nuechterlein, K. H., & Green, M.F (2009). Development of a measure of relationship perception in schizophrenia. *Psychiatry Research,* 166(1), 54–62.

Sprong, M., Schothorst, P., Vos, E., Hox, J., & van Engeland, H. (2007). Theory of mind in schizophrenia: meta-analysis. *Br J Psychiatry,* 191, 5–13.

Stouten, L. H., Veling, W., Laan, W., Helm, M., & Gaag, M. (2015). Psychosocial functioning in first-episode psychosis and associations with neurocognition, social cognition, psychotic and affective symptoms. *Early Intervention In Psychiatry.*

Thompson AD, Bartholomeusz C, Yung AR (2011). Social cognition deficits and the ultra-high risk' for psychosis population: a review of the literature. *Early Intervention in Psychiatry*, 5, 192-202.

Thompson, A., Papas, A., Bartholomeusz, C., Allott, K., Amminger, G. P., Nelson, B., & ...Yung, A. (2012). Social cognition in clinical 'at risk' for psychosis and first episode psychosis populations. *Schizophrenia Research*, *141*(2-3), 204-209.

Ursano, R. J., Bell, C., Eth, S., Friedman, M., Norwood, A., Pfefferbaum, B., … & Yager, J. (2004). Practice guideline for the treatment of patients with acute stress disorder and posttraumatic stress disorder. *Am. J. Psychiatry*, *161*(11 Suppl.):1–31.

Ventura, J., Ered, A., Gretchen-Doorly, D., Subotnik, K. L., Horan, W. P., Hellemann, G. S., & Nuechterlein, K. H. (2015). Theory of mind in the early course of schizophrenia: Stability, symptom and neurocognitive correlates, and relationship with functioning. *Psychological Medicine*, *45*(10), 2031-2043.

Weickert, T. W., Goldberg, T. E., Gold, J. M., Bigelow, L. B., Egan, M. F., & Weinberger, D. R. (2000). Cognitive impairments in patients with schizophrenia displaying preserved and compromised intellect. *Archives of General Psychiatry*, *57*(9), 907-913.

Wu, E. Q., Birnbaum, H. G., Shi, L., Ball, D. E., Kessler, R. C., Moulis, M., & Aggarwal, J. (2005). The economic burden of schizophrenia in the United States in 2002. *Journal of Clinical Psychiatry*.

Wykes, T., Reeder, C., Landau, S., Matthiasson, P., Haworth, E., Hutchinson, C., 2009. Does age matter? Effects of cognitive rehabilitation across the age span. *Schizophr. Res*. 113(2–3), 252–258.

Zaytseva, Y., Burova, V., Garakh, Z., & Gurovich, I. Y. (2013). Attributional style in first episode of schizophrenia and schizophrenia spectrum disorders with and without paranoid ideation. *Psychiatria Danubina*, *25*(Suppl 2), 329-331.

Yung, A. R., Killackey, E. J., Nelson, B., & McGorry, P. D. (2010). The impact of early intervention in schizophrenia. In W. F. Gattaz, G. Busatto, W. F. Gattaz, G. Busatto (Eds.): *Advances in schizophrenia research 2009* (pp. 299-316). New York, NY, US: Springer Science + Business Media.

TABLES

Table 1a. Participant Demographic and Clinical Characteristics (FEP)

Characteristic	FEP Patients (n=38)		FEP Controls (n = 39)	
	n	(%)	n	(%)
Male	33	86.7	32	82.1
Race				
Caucasian	29	76.3	29	74.4
African American	5	13.2	5	12.8
Asian	2	5.3	2	5.1
Other	2	5.3	3	7.7
Ethnicity				
Hispanic	2	5.3	6	15.4
Non-Hispanic	36	94.7	33	84.6
Diagnosis				
Schizophrenia	25	65.8		
Schizoaffective	6	15.8		
Psychosis NOS	7	18.4		
Medication Type				
Typical	1	2.6		
Atypical	32	84.2		
Combination	2	5.3		
Unmedicated	3	7.9		
	Mean	SD	Mean	SD
Age (years)*	23.45	3.01	23.77	3.39
Education (years)*	14.03	1.52	15.44	1.80
Maternal Education (years) *	16.21	2.27	14.85	1.99
Paternal Education (years)*	17.33	2.33	15.53	2.81
WRAT-3*	105.87	9.35	107.82	8.91
UPSA-B*	70.55	11.63	80.53	9.59
SSPA-Avg.	4.10	.39	4.68	.21
SLOF-Avg.	4.25	.46	4.61	.24
PANSS (Visit 1)				
Positive Total	17.53	4.91		
Negative Total	16.58	3.96		
General Total	36.00	5.95		
Overall Total	70.11	10.37		
PANSS (Visit 2)				
Positive Total	14.63	5.28		
Negative Total*	15.21	5.58		
General Total*	32.92	10.20		
Overall Total*	62.76	18.69		

Table 1b. Participant Demographic and Clinical Characteristics (SCOPE)

Characteristic	SCOPE Patients (n=179)		SCOPE Controls (n = 104)	
	n	(%)	n	(%)
Male	117	65.4	49	47.1
Race				
Caucasian	76	76.3	43	41.3
African American	94	13.2	55	52.9
Asian	4	5.3	4	3.9
Other	5	5.3	2	1.9
Ethnicity				
Hispanic	37	5.3	21	20.2
Non-Hispanic	142	94.7	83	79.8
Diagnosis				
Schizophrenia	96	53.6		
Schizoaffective	83	46.4		
Psychosis NOS	-	-		
Medication Type				
Typical	26	14.5		
Atypical	125	69.8		
Combination	3	1.7		
Unmedicated	-	-		
	Mean	SD	Mean	SD
Age (years)*	42.11	12.32	39.20	13.70
Education (years)*	12.70	2.14	13.43	1.66
Maternal Education (years) *	12.61	3.22	13.14	2.53
Paternal Education (years)*	13.04	3.75	13.43	2.49
WRAT-3*	93.68	15.88	95.35	13.19
UPSA-B*	69.92	14.33	-	-
SSPA-Avg.	4.22	.62	4.63	.47
SLOF-Avg.	3.92	.62	4.62	.44
PANSS (Visit 1)				
Positive Total	16.14	5.79		
Negative Total	13.72	5.29		
General Total	30.83	7.99		
Overall Total	60.69	15.51		
PANSS (Visit 2)				
Positive Total	15.37	5.07		
Negative Total*	13.51	5.12		
General Total*	29.48	7.61		
Overall Total*	58.36	13.79		

Table 2. Test-Retest Reliability and Internal Consistency

Task	Test-Retest Reliability (Pearson *r*)		Internal Consistency (Cronbach's Alpha)	
	FEP Patients (n=34)	Controls (n=36)	FEP Patients (n=38)	Controls (n=39)
AIHQ				
Hostility Bias (HB)	.529	.394	.497	.387
Aggression Bias (AB)	.238	.664	.259	.242
Blame Score (BS)	.737	.680	.857	.742
BLERT	.455	.665	.740	.411
ER-40	.496	.705	.599	.538
Eyes	.534	.708	.488	.630
Hinting	.735	.204	.685	.493
RAD	.753	.735	.683	.558
TASIT	.314	.338	.795	.691
Trust	.218	.537	.943	.821

Note: With the outlier included in the analyses, test-retest reliability for the BLERT was .490 (n=35).

Table 3a. Direct comparison of social cognition across studies, part 1

Task	SCOPE-FEP		Original SCOPE	
	Clinical M (SD)	Control M (SD)	Clinical M (SD)	Control M (SD)
AIHQ-HB	2.02 (.59)	1.89 (.49)	2.38 (.60)	1.99 (.60)
AIHQ-AB	1.74 (.23)	1.77 (.20)	1.89 (.38)	1.83 (.26)
AIHQ-BS	7.09 (2.15)	7.09 (1.47)	8.74 (2.81)	7.02 (2.31)
BLERT	17.14 (2.12)	17.59 (2.04)	13.17 (3.88)	15.75 (2.88)
ER-40	32.05 (3.54)	33.67 (3.00)	29.55 (5.40)	32.80 (3.54)
Eyes	25.16 (3.67)	26.54 (4.08)	20.15 (5.46)	23.55 (4.62)
Hinting	15.82 (2.82)	17.72 (1.78)	13.59 (3.87)	16.82 (2.05)
RAD	32.34 (4.92)	33.87 (3.95)	24.76 (5.76)	29.82 (5.16)
TASITA	51.89 (6.24)	57.31 (3.95)	44.43 (7.64)	51.48 (5.62)
Trust	.17 (.83)	.39 (.39)	-.09 (1.14)	.16 (.62)

Note$_a$: BLERT for FEP patients (M=16.74, SD=2.04) when the outlier is included in the analyses

Table 3b. Direct comparison of social cognition across studies, part 2

Task	Main Effects		Interaction
	Phase	Group	Phase x Group
AIHQ-HB	$F(1,354) = 2.85$ $p = .092$	***F*(1,354) = 12.26** **p = .021**	$F(1,354) = 3.72$ $p = .055$
AIHQ-AB	$F(1,354) = 2.58$ $p = .109$	$F(1,354) = 0.53$ $p = .819$	$F(1,354) = 1.06$ $p = .305$
AIHQ-BS	$F(1,354) = 1.97$ $p = .161$	***F*(1,354) = 7.23** **p = .008**	***F*(1,355) = 8.15** **p = .005**
BLERT	$F(1,353) = 0.78$ $p = .379$	***F*(1,353) = 11.53** **p = .001**	***F*(1,353) = 6.52** **p = .011**
ER-40	$F(1,354) = 0.46$ $p = .496$	***F*(1,354) = 15.65** **p < .001**	$F(1,354) = 1.72$ $p = .190$
Eyes	$F(1,355) = 5.27$ $p = .022$	***F*(1,355) = 15.46** **p < .001**	***F*(1,355) = 4.22** **p = .041**
Hinting	$F(1,353) = 9.97$ $p = .002$	***F*(1,353) = 39.42** **p < .001**	$F(1,353) = 3.14$ $p = .077$
RAD	***F*(1,351) = 10.07** **p = .002**	***F*(1,351) = 24.08** **p < .001**	***F*(1,351) = 7.97** **p = .005**
TASITA	***F*(1,355) = 9.25** **p = .003**	$F(1,3545) = 59.82$ $p < .001$	$F(1,355) = 1.18$ $p = .278$
Trust	***F*(1,355) = 5.08** **p = .025**	***F*(1,355) = 3.85** **p = .051**	$F(1,355) = .02$ $p = .895$

Note$_a$. Analyses were conducted controlling for the effects of age and estimated IQ.
Note$_b$: Tests indicating significant interactions and main effects are indicated in **bold**.

Table 4a: Utility as a Repeated Measure, part 1

Task	T_1		T_2		T_2-T_1 Difference	
	Mean	SD	Mean	SD	Mean	SD
	Patients (n=36)					
AIHQ-HB	1.99	0.577	1.86	0.561	-0.125	0.552
AIHQ-AB	1.73	0.226	1.77	0.268	-0.047	0.307
AIHQ-BS	7.1	2.19	7.07	2.29	-0.028	-1.626
BLERT	17.23	2.14	17.91	1.74	0.686	2.055
ER-40	32.22	3.55	33.97	2.65	1.75	3.21
Eyes	25.25	3.74	25.89	3.46	0.639	3.482
Hinting	15.83	2.87	17.08	2.13	1.25	1.948
RAD	32.64	4.88	32.19	5.52	-0.444	3.707
TASIT*	52.03	6.36	49	6.12	-3.028	7.307
Trust	0.172	0.834	0.198	0.531	0.026	0.885
	Controls (n=36)					
AIHQ-HB	1.86	0.46	1.69	0.539	-0.169	0.55
AIHQ-AB	1.77	0.195	1.74	0.222	-0.036	0.17
AIHQ-BS	7.13	1.482	6.59	1.819	-0.544	1.36
BLERT	17.61	2.032	18.67	1.805	1.056	1.58
ER-40	33.67	3.089	34.58	2.781	0.917	2.27
Eyes	26.78	4.134	27.78	3.958	1	3.098
Hinting	17.92	1.538	18	1.639	0.083	2.005
RAD	34.08	4.003	36.19	3.984	2.111	2.906
TASIT*	57.75	3.667	53.47	5.955	-4.278	5.843
Trust	0.394	0.392	0.327	0.413	-0.067	0.388

*Note: Alternate versions were used for the TASIT

Table 4b: Utility as a Repeated Measure, part 2

Task	Number at Floor/Ceiling				
	T_1	T_2	*t*	*p* value	Cohen's d_z
	Patients (n=36)				
AIHQ-HB	--	--	-1.359	0.183	0.226
AIHQ-AB	--	--	-0.924	0.362	0.153
AIHQ-BS	--	--	-0.102	0.919	0.017
BLERT	1/0	0/1	1.974	0.057	0.351
ER-40	0/0	0/0	3.271	0.002	0.545
Eyes	2/0	2/0	1.101	0.278	0.184
Hinting	1/2	0/2	3.851	<.001	0.642
RAD	1/0	2/0	0.719	0.477	0.12
TASIT*	1/0	0/1	-2.486	0.018	0.414
Trust	--	--	0.179	0.859	0.029
	Controls (n=36)				
AIHQ-HB	--	--	-1.836	0.075	0.307
AIHQ-AB	--	--	-1.255	0.218	0.212
AIHQ-BS	--	--	-2.41	0.021	0.4
BLERT	0/2	0/5	3.997	<.001	0.668
ER-40	0/0	0/0	2.42	0.021	0.404
Eyes	1/0	1/0	1.936	0.061	0.323
Hinting	0/4	0/3	0.249	0.805	0.041
RAD	0/0	0/0	4.359	<.001	0.726
TASIT*	0/0	0/2	-4.392	<.001	0.732
Trust	--	--	-1.091	0.282	0.173

Table 5a. Correlations between Candidate Measures of Social Cognition and Neurocognition, part 1

Task	AIHQ-HB	AIHQ-AB	AIHQ-BS	BLERT	ER-40
AIHQ-HB	--	0.247	.527**	-0.211	-0.116
AIHQ-AB	0.247	--	.461**	.380*	.422**
AIHQ-BS	.527**	.461**	--	0.005	0.007
BLERT	-0.211	.380*	0.005	--	.520**
ER-40	-0.116	.422**	0.007	.520**	--
Eyes	-0.166	-0.2	-0.093	0.264	0.259
Hinting	-0.196	0.267	0.035	.410*	.391*
RAD	0.015	0.096	0.034	.340*	0.311
TASIT	-0.168	0.149	0.179	0.084	0.077
Trust	-0.057	-0.15	-0.246	.443**	-0.076

Table 5b. Correlations between Candidate Measures of Social Cognition and Neurocognition, part 2

Task	Eyes	Hinting	RAD	TASIT	Trust	MCCB Composite
AIHQ-HB	-0.166	-0.196	0.015	-0.168	-0.057	-0.046
AIHQ-AB	-0.2	0.267	0.096	0.149	-0.15	-0.136
AIHQ-BS	-0.093	0.035	0.034	0.179	-0.246	-0.076
BLERT	0.264	.410*	.340*	0.174	.443**	0.065
ER-40	0.259	.391*	0.311	0.077	-0.076	0.11
Eyes	--	0.251	0.305	0.165	0.042	.327*
Hinting	0.251	--	0.281	.522**	-0.304	0.176
RAD	0.305	0.281	--	.327*	-0.064	0.253
TASIT	0.165	.522**	.327*	--	-0.277	0.111
Trust	0.042	-0.304	-0.064	-0.277	--	0.08

Table 6. Correlations between Social Cognitive and Neurocognitive Tasks and Functional Outcome Measures in Patients

	UPSA Total	SSPA Average	$SLOF_{SR}$ Average
Social Cognitive			
AIHQ-HB	-.096	-.162	-.360*
AIHQ-AB	.136	.069	-.253
AIHQ-BS	.158	.053	-.372*
BLERT	.265	.159	.138
ER-40	.337*	.435**	-.101
Eyes	.326*	.234	.407*
Hinting	.372*	.473**	.189
RAD	.456**	.344*	.020
TASIT	.475**	.179	.205
Trust	-.037	-.252	.161
Neurocognitive			
MCCB Composite	.489**	.384*	.123

* p<.05, ** p<.01

Note$_a$: There was an error with scene two for the SSPA role-play for one SCZ participant. This particular individual's data – the average for scene 1 only – were included in the above analyses.

Note$_b$: All participants completed the self-report (SR) version of the SLOF. Informants were identified for each SCZ participant, though only 25 individuals successfully completed the informant version of the measure. Neither performance on the social cognition measures nor scores on the social functioning measures were significantly related to the informant version of the SLOF.

Table 7. Regression Models Summarizing Independent and Combined Contributions of Neurocognition and Social Cognition to Outcomes

Neurocognition only								
MCCB Composite	*R^2*	*Adjusted R^2*	*F*	*p*				
UPSA total	.239	.218	11.334	.002				
SSPA average	.147	.123	6.039	.019				
$SLOF_{SR}$ average	.015	-.012	.550	.463				
Social Cognition only								
SC Tasks	*R^2*	*Adjusted R^2*	*F*	*p*	*b**	*t*	*p*	*sr^2*
UPSA total	.392	.297	4.12	.005				
ER-40					.198	1.241	.224	.029
Eyes					.143	.963	.343	.018
Hinting					.007	.041	.968	.000
RAD					.233	1.486	.147	.042
TASIT					.357	2.099	.044	.084
SSPA average	.323	.262	5.26	.004				
ER-40					.254	1.588	.122	.052
Hinting					.325	2.054	.048	.086
RAD					.172	1.124	.269	.026
$SLOF_{SR}$ average	.298	.236	4.82	.007				
AIHQ-HB					-.169	-.991	.329	.020
AIHQ-BS					-.250	-1.478	.149	.045
Eyes					.356	2.442	.020	.123

Neurocognition and Social Cognition

	UPSA-B		SSPA		$SLOF_{SR}$	
	*b**	*sr^2*	*b**	*sr^2*	*b**	*sr^2*
Block 1 - Neurocognition						
MCCB Composite	.373**	.120**	.274	.069	-.023	.000
Block 2 – Social Cognition						
AIHQ-HB	--	--	--	--	-.168	.020
AIHQ-AB	--	--	--	--	--	--
AIHQ-BS	--	--	--	--	-.251	.045
ER-40	.216	.035	.257	.053	--	--
Eyes	.043	.001	--	--	.363*	.115*
Hinting	-.030	.001	.294	.070	--	--
RAD	.169	.022	.106	.009	--	--
TASIT	.371*	.091*	--	--	--	--
Overall Model						
Adjusted R^2	.417*		.316*		.214**	
R^2 Change	.199*		.193*		.214**	

* $p<.05$, ** $p<.01$

Table 8. Practicality and Tolerability

	Practicality (Administration Time in Minutes)				Tolerability (Participant Ratings)			
Task	Patients (n=38)		Controls (n=39)		Patients (n=38)		Controls (n=39)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
AIHQ	6.98	1.35	6.16	1.16	5.41	1.22	5.60	0.95
BLERT	7.72	1.47	7.14	0.76	5.21	1.53	5.23	1.29
ER-40	3.46	0.83	2.60	0.57	5.39	1.31	5.62	0.99
Eyes	7.20	1.97	5.14	1.76	5.09	1.52	5.23	1.14
Hinting	7.95	1.66	6.85	1.34	4.83	1.36	5.37	1.33
RAD	16.80	3.16	14.53	3.38	4.29	1.39	4.96	1.13
TASIT	20.40	2.86	18.80	2.29	4.92	1.46	5.39	1.01
Trust	4.53	1.39	3.61	1.04	5.42	1.46	5.56	1.05

Note: These ratings are from the first administration of each SC task only.

Table 9. Comparison of our Findings with the Preliminary Findings from SCOPE

Task	Test-retest	Internal Consistency	Utility as Repeated Measure	Rel. to Functioning	Group Differences	Admin. Time < 10 min.	Rec. for Use
				SCOPE-FEP			
AIHQ-HB	**X**	**X**	✓	✓	**X**	✓	**X**
AIHQ-AB	**X**	**X**	✓	**X**	**X**	✓	**X**
AIHQ-BS	✓	✓	✓	✓	**X**	✓	**X**
BLERT	**X**	**X**	✓	**X**	**X**	✓	**X**
ER-40	**X**	**X**	**X**	✓	**X**	✓	**X**
Eyes	**X**	**X**	✓	✓	**X**	✓	**X**
Hinting	✓	**X**	**X**	✓	✓	✓	✓
RAD	✓	**X**	✓	✓	**X**	**X**	✓
TASIT	**X**	✓	**X**	✓	✓	**X**	**X**
Trust	**X**	✓	✓	**X**	**X**	✓	**X**
				Original SCOPE			
AIHQ-HB	**X**	✓	✓	**X**	✓	✓	**X**
AIHQ-AB	**X**	**X**	✓	**X**	**X**	✓	**X**
AIHQ-BS	✓	**X**	✓	**X**	✓	✓	**X**
BLERT	✓	**X**	✓	✓	✓	✓	✓
ER-40	✓	✓	✓	✓	✓	✓	✓
Eyes	✓	**X**	✓	✓	✓	✓	✓
Hinting	✓	**X**	✓	✓	✓	✓	✓
RAD	✓	**X**	**X**	✓	✓	**X**	**X**
TASIT	✓	✓	✓	✓	✓	**X**	✓
Trust	✓	✓	✓	**X**	✓	✓	**X**