



## Brief battery of the Social Cognition Psychometric Evaluation study (BB-SCOPE): Development and validation in schizophrenia spectrum disorders

Tate F. Halverson<sup>a,b,\*</sup>, Amy E. Pinkham<sup>c,d</sup>, Philip D. Harvey<sup>e,f</sup>, David L. Penn<sup>g,h</sup>

<sup>a</sup> Durham Veterans Affairs Health Care System, Durham, NC, USA

<sup>b</sup> VA Mid-Atlantic Mental Illness Research, Education and Clinical Center, Durham, NC, USA

<sup>c</sup> School of Behavioral and Brain Sciences, University of Texas at Dallas, Richardson, TX, USA

<sup>d</sup> Department of Psychiatry, University of Texas Southwestern Medical School, Dallas, TX, USA

<sup>e</sup> Department of Psychiatry and Behavioral Sciences, University of Miami Miller School of Medicine, Miami, FL, USA

<sup>f</sup> Research Services, Miami Veterans Affairs Health Care System, Miami, FL, USA

<sup>g</sup> Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

<sup>h</sup> School of Behavioural and Health Sciences, Australian Catholic University, Melbourne, VIC, Australia

### ARTICLE INFO

#### Keywords:

Schizophrenia  
Schizoaffective disorder  
Social cognition  
Measurement

### ABSTRACT

The aim of the present study was to develop an abbreviated social cognition (SC) battery for individuals with schizophrenia spectrum disorders (SSD) to reduce the heterogeneity of and increase the frequency of assessment of SC impairment. To this end, the present study utilized Item Response Theory to develop brief versions of SC tasks administered to individuals with SSD ( $n = 386$ ) and individuals without a psychiatric diagnosis ( $n = 292$ ) during the Social Cognition Psychometric Evaluation (SCOPE) Study. Seven brief measures of SC were evaluated (i.e., Ambiguous Intentions and Hostility Questionnaire [AIHQ], Bell Lysaker Emotion Recognition Task [BLERT], Penn Emotion Recognition Task, Reading the Mind in the Eyes Task, Hinting Task, Intentionality Bias Task, Relationships Across Domains Task), and the existing brief version of The Awareness of Social Inference Test was reviewed. Psychometric properties for each brief SC measure were evaluated and compared to the original measures. Based on psychometric properties and relationships with other measures of SC, neurocognition, and functioning, two brief tasks (AIHQ, BLERT) and the full-length Hinting task were recommended for inclusion in a brief battery of SC tasks from the SCOPE Study (BB-SCOPE). The resulting BB-SCOPE is efficient, with an estimated administration time of 15 min, and comprehensively assesses three domains of SC (i.e., attributional bias, emotion processing, theory of mind) to identify severe SC impairment. Scoring of BB-SCOPE is also straightforward and includes a recommended cut-point of 60 for identifying SC impairment.

Social cognition (SC) comprises how individuals think about themselves, others, social situations, and social interactions, and impairments in SC are well-documented in schizophrenia spectrum disorders (SSD; Hajdúk et al., 2018; Penn et al., 1997). SC impairments correlate with current symptoms, predict poorer outcome trajectories, and demonstrate reliable and unique relationships to functional outcomes, highlighting the role of SC as a compelling treatment target (Green et al., 2010; Halverson et al., 2019; Harvey and Penn, 2010; Velthorst et al., 2017).

While many individuals with SSD exhibit SC impairment, research suggests 25% of individuals with SSD do not show impairment in SC

(Hajdúk et al., 2018). Therefore, accurate assessment of SC in SSD is important to identify individuals that may optimally benefit from treatments targeting these impairments. To this end, the multi-phase Social Cognition and Psychometric Evaluation (SCOPE) Study was designed to identify existing SC tasks that optimally assess SC in SSD. SC tasks were classified as “acceptable as is”, “acceptable with modifications”, or “not recommended” based on psychometric properties and expert consensus (Pinkham et al., 2014, 2016, 2018). The SCOPE Study evaluated 11 tasks of SC and identified six SC tasks rated as “acceptable” assessing the following domains of SC: *emotion processing* – Penn Emotion Recognition Task (ER40), Bell Lysaker Emotion Recognition

\* Corresponding author. Durham Veterans Affairs Health Care System 3022 Croasdaile Drive, Durham, NC, 27705, USA.

E-mail address: [tate.halverson@va.gov](mailto:tate.halverson@va.gov) (T.F. Halverson).

<https://doi.org/10.1016/j.jpsychires.2022.03.017>

Received 28 November 2021; Received in revised form 28 February 2022; Accepted 9 March 2022

Available online 22 March 2022

0022-3956/Published by Elsevier Ltd.

Task (BLERT); *mental state attribution* – Eyes Task, The Awareness of Social Inferences Test – Part III (TASIT-III), Hinting Task; *attributional style* – Intentionality Bias Task (IBT). Five tasks were not recommended (i.e., Ambiguous Intentions Hostility Questionnaire [AIHQ], Trustworthiness Task, Relationships Across Domains Task [RAD], Social Attribution Test – Multiple Choice [SAT-MC], Mini Profile of Nonverbal Sensitivity [MiniPONS]), including all tasks of *social perception* (i.e., RAD, SAT-MC, MiniPONS).

Altogether, administration of tasks with an acceptable rating from the SCOPE Study (i.e., ER40, BLERT, Eyes, TASIT, Hinting, IBT) is estimated to have a total mean administration time around 1 h (Pinkham et al., 2018). An administration length of one hour is less than ideal for individuals in an acute illness phase (e.g., inpatient setting) or for individuals attending a standard 50-min community care outpatient appointment. Given the relationship of SC with functional impairment and the importance of early intervention, a brief battery of SC is imperative for efficient treatment planning to obtain optimal functional improvements.

Similar challenges (i.e., long administration times, heterogeneity in measurement) existed in the assessment of neurocognition (NC) in SSD before the development of the Brief Assessment of Cognition in Schizophrenia (BACS; Keefe et al., 2004). The BACS is a comprehensive assessment of key domains of NC in SSD with strong relationships to functional outcomes specifically developed as a brief battery for use in clinical trials. With a total administration time of 35 min, the BACS is feasible to administer, demonstrates good psychometric properties, and is strongly correlated with standard batteries of NC with longer administration times.

An analogous approach to brief assessment of SC in SSD is needed to prompt similar dissemination of a standard battery to identify SC impairment. A more clinic-friendly battery may reduce heterogeneity of SC measurement in research and increase SC assessment in clinical practice. Whereas the BACS was developed de novo, the SCOPE battery offers a useful starting point to create a brief battery of SC for several reasons. First, SCOPE measures were identified according to expert survey and finalized based on thorough psychometric evaluation, suggesting this is one of the most comprehensive evaluated batteries possible. Second, with over 650 individuals across all phases, the SCOPE Study is one of the largest samples to examine task performance in both SSD and individuals without a psychiatric diagnosis. The SCOPE Study also carefully assessed symptoms, NC, and functioning, allowing for comprehensive validity investigation.

Item Response Theory (IRT) is a compelling approach for developing brief versions of measures (Bock, 1997; Thissen and Orlando, 2001) due to an emphasis on individual item performance (DeVellis, 2017). For each item, IRT estimates a parameter for item difficulty and item discrimination. Estimation of these parameters and graphical presentation allow for identification of items that optimally discriminate different levels of a latent trait (e.g., emotion perception) as well as ensure the entire trait continuum is covered (e.g., low to high levels of emotion perception). There are several examples of using an IRT approach for development and validation of brief measures (Bortolotti et al., 2013; Petrillo et al., 2015), including in SSD (Ventura et al., 2010).

The aims of the present study are 1) develop brief versions of all tasks from the SCOPE Study with “acceptable” ratings; 2) examine brief versions of tasks with previous “not recommended” ratings where concerns may be addressed by shortening the task (e.g., long administration times such as in the RAD Task); 3) examine psychometric properties of brief versions of all tasks (i.e., group differences, internal reliability, utility as a repeated measure, and relationships with functional outcomes); 4) make recommendations for a comprehensive brief battery of SCOPE (BB-SCOPE) to efficiently assess SC impairment in SSD.

## 1. Method

### 1.1. Participants

Participants were individuals with SSD ( $n = 386$ ) and healthy controls (HC;  $n = 292$ ) recruited from Southern Methodist University ( $n = 165$ ), University of Miami Miller School of Medicine ( $n = 227$ ), University of Texas at Dallas ( $n = 135$ ), and University of North Carolina at Chapel Hill ( $n = 151$ ). Institutional Review Boards at all institutions approved study procedures.

Participants during phase three (SSD  $n = 179$ , HC  $n = 104$ ) and phase five (SSD  $n = 158$ , HC  $n = 153$ ) completed a baseline visit and a retest assessment two to four weeks later. Participants provided informed consent and completed SC and NC tasks as well as functional outcome measures (counterbalanced order). For individuals with SSD, symptom severity was assessed using the Positive and Negative Syndrome Scale (PANSS; Kay et al., 1987), rater ICCs  $>0.80$ . Participants during Phase Four (SSD  $n = 49$ , HC  $n = 35$ ) completed a single study visit of SC tasks rated as “acceptable with modifications” from phase three. Visit procedures were similar to phases three and five with the exception of no retest visit. Detailed study methods and procedures are published elsewhere (Cornacchio et al., 2017; Pinkham et al., 2016, 2018).

### 1.2. Measures

#### 1.2.1. Social cognition measures

**Attributional Style/Bias.** The *Ambiguous Intentions and Hostility Questionnaire* (AIHQ; Combs et al., 2007) presents five vignettes and asks participants to rate the extent to which the person from the vignette performed an action on purpose, how angry the action would make them feel, and how much they would blame the other person. Ratings for each vignette are summed and then averaged (i.e., sum of vignette total scores divided by five) to create a Blame Index (range 3–16). The original AIHQ includes open-ended responses scored by trained raters, however, research suggests scoring only the Likert items may improve AIHQ performance, and thus only the Blame Index will be examined in the present study (Buck et al., 2017).

The *Intentionality Bias Task* (IBT; Rosset, 2008) consists of 24 short sentences describing simple actions. Participants indicate whether these actions occurred “on purpose” or “by accident.” Intentionality bias is calculated as the percentage of trials indicated as intentional, with higher scores indicating greater intentionality bias (range 0–1).

Tasks of attributional bias tend to show stronger relationships to paranoia and suspiciousness than other domains of SC, and therefore correlations with symptoms were also examined (Buck et al., 2017; Pinkham et al., 2016).

**Emotion Perception and Processing.** The *Bell Lysaker Emotion Recognition Task* (BLERT; Bryson et al., 1997) includes 21 videos of an actor displaying different emotions. Participants select which affect state (e.g., happiness, sadness) was most prominently displayed. Correct responses are summed for a total score (range 0–21).

The *Penn Emotion Recognition Test* (ER40; Kohler et al., 2003) includes 40 photographs of faces expressing four basic emotions as well as neutral expressions. Participants choose the correct emotion from five choices. An accuracy score is calculated by summing correctly identified emotions (range 0–40).

**Mental State Attribution.** The *Awareness of Social Inferences Test Part III* (TASIT; McDonald et al., 2003) measures the ability to detect lies or sarcasm and has two forms (i.e., form A and form B) for counterbalanced administration across retesting. Participants watch videos of social interactions and answer questions regarding the intentions, beliefs, and meanings of the speakers and their interactions. A short version of the TASIT, the TASIT-S (Honan et al., 2016) was developed after initial data collection and will be re-examined in the present study. The TASIT-S includes only one form and includes items from both form A and form B so alternate form reliability will not be examined since

forms A and B were counterbalanced across administrations in the SCOPE Study (total range of correctly answered questions 0–36).

The *Reading the Mind in the Eyes Test* (Eyes; Baron-Cohen et al., 2001) consists of 36 black-and-white photographs of eye regions expressing different thoughts and feelings. Participants select the thought/feeling portrayed in the photograph from a list of four options. Correct items are summed for a total score (range 0–36).

The *Hinting Task* (Corcoran et al., 1995) includes short passages of interactions between two characters ending with an indirect statement. Participants are prompted to explain the intention of this statement. Responses are coded as “2” (correctly described intention), “1” (correctly described intention after additional information is provided), or “0” (did not provide accurate description of intention) with a total score range 0–20, see Klein et al. (2020) for recommended scoring criteria.

**Social Perception.** The *Relationship Across Domains Test* (RAD; Sergi et al., 2009) presents interactions consistent with four relational models. Participants respond “yes” or “no” if they think a future behavior is likely based on the vignette. A total correct score is calculated (range 0–45). The RAD was not recommended after phase three due to long administration time (i.e., 15 min), however a brief version with careful item selection may address long administration time.

### 1.2.2. Neurocognition measures

A subset of the MATRICS Consensus Cognitive Battery (MCCB; Nuechterlein et al., 2008) assessed NC. MCCB tasks administered included: Trail Making Test – Part A, BACS – Symbol Coding, Category Fluency, Animal Naming, Letter-Number Span, and the Hopkins Verbal Learning Test-Revised.

### 1.2.3. Functional outcomes

The UCSD Performance-Based Skills Assessment – Brief (UPSA-B; Maudsley et al., 2007) assessed functional capacity. Social competence was assessed with the Social Skills Performance Assessment (SSPA; Patterson et al., 2001). Real-world functioning was assessed by informants (SSD only) as well as self-reports using the Specific Levels of Functioning Scale (SLOF; Schneider and Struening, 1983).

## 1.3. Data analytic plan

Measures were assessed for multidimensionality by comparing unidimensional and multidimensional IRT models using the *mirt* package for R (Chalmers, 2012) with Bayesian Information Criterion (BIC) values emphasized since this criterion rewards parsimonious models (Kuha, 2004). For measures where multidimensional models were recommended, unidimensional IRT models were fit to each dimension separately to preserve original task factors. IRT model fitting and test statistics were computed for all tasks (i.e., one-, two-, or three-parameter models for dichotomous responses and graded response, graded partial credit, and nominal models for polytomous responses). Once an appropriate model fit was established, individual item difficulty ( $b$ ) and discrimination ( $a$ ) statistics were reviewed. Item information values ( $\theta$ ) were evaluated based on performance at one standard deviation below average ability since a brief battery is intended to identify individuals with impaired SC. Candidate items were compiled into brief versions of each SC task.

Psychometric properties of each brief measure were examined: internal consistency (coefficient omegas), test-retest (Pearson's  $r$  correlation coefficient), utility as a repeated measure (paired-samples  $t$ -test comparing administration timepoints), floor/ceiling effects (proportion of participants performing at chance level/achieving perfect scores), and group differences ( $t$ -test comparing healthy control performance with SSD sample). Relationships with NC task performance and functional outcome measures were also assessed (Pearson's  $r$ ). Pearson's  $r$  values greater than .60 were considered acceptable (Akoglu, 2018; Kraemer et al., 2012). Relationships with indicators of functioning were assessed

through a series of regressions with SC measures as predictors and functional outcome measures as dependent variables. Regression models including both NC and brief SC measures as predictors of functional outcomes tested incremental validity. Relationships between original and brief SC measures were also examined (Pearson's  $r$ ). While some approaches to develop brief batteries emphasize a single efficient test, SC is comprised of meaningful domains (Buck et al., 2016a,b; Riedel et al., 2021). Therefore, development of BB-SCOPE emphasized representation of SC measures across SC domains with good psychometric properties and unique relationships with functional outcomes. Receiver operating characteristic (ROC) analyses identified an optimal BB-SCOPE total score based on Youden's Index maximizing sensitivity and specificity (Robin et al., 2011; Youden, 1950) for identifying SC impairment. Choosing an optimal cut-score balanced identification of individuals one standard deviation below HC performance and high area under the curve (AUC) values for identifying social competence and functional capacity.

## 2. Results

Participant demographic and clinical characteristics are presented in Table 1. Performance on SC tasks, NC tasks, and functional outcomes are presented in Table 2. SSD individuals completed fewer years of education and had lower WRAT-3 scores compared with HC and performed significantly lower on all indices of SC, NC, and functional outcomes.

### 2.1. Development of brief social cognition tasks

Assessment of SC task dimensionality, IRT fit statistics, and optimal model fit are presented in Supplementary Materials. Items retained for brief versions are presented in Supplementary Tables 1–8. The number of items retained for brief versions was guided by a preference to select efficient items balanced with number of items needed for acceptable psychometric properties. Efforts were made to retain original factors as intended by the original task authors. Fig. 1 presents IRT analysis of original and brief SC tasks.

**Table 1**  
Sample characteristics by diagnostic group.

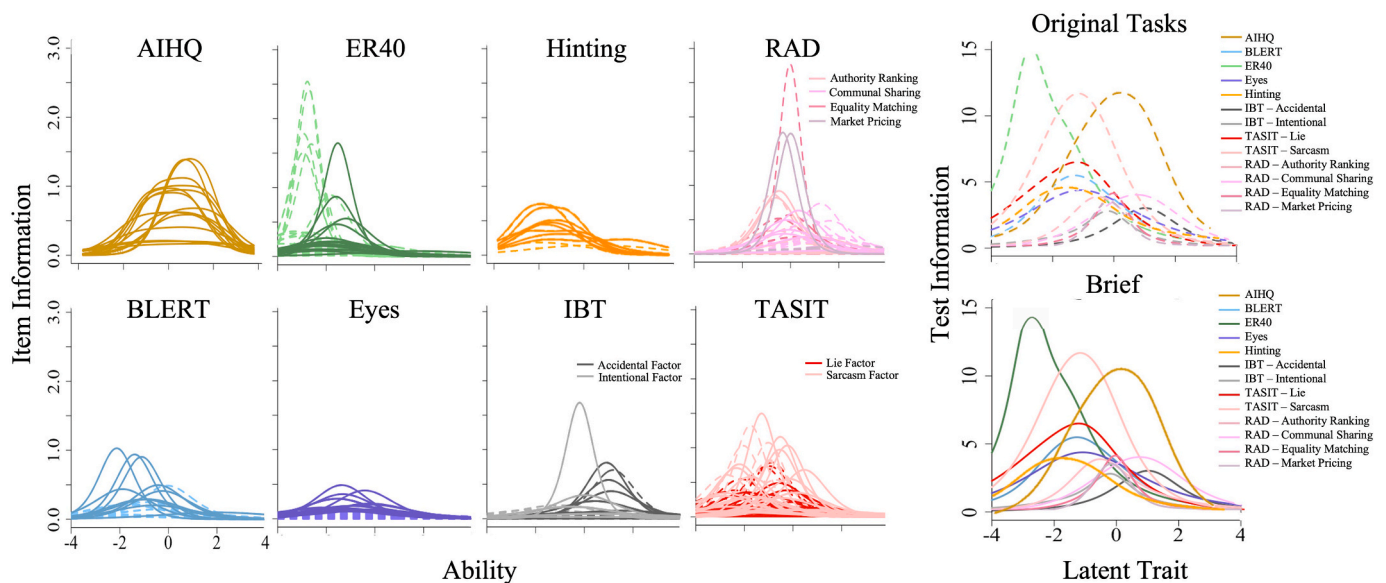
	SSD (n = 386)	HC (n = 292)	p-value <sup>a</sup>
Age, years	41.5 ± 11.9	41.2 ± 12.7	.74
Male, % (n)	66.0 (254)	57.2 (167)	.02
Education, years	12.9 ± 2.4	13.9 ± 1.9	<.01
WRAT-3 Standard Score	94.6 ± 15.0	98.8 ± 12.3	<.01
Race/Ethnicity <sup>b</sup> , % (n)			
White	48.4 (187)	46.6 (136)	.63
Black	45.3 (175)	46.6 (136)	.75
Other	6.2 (24)	6.8 (20)	.74
Hispanic/Latinx	17.6 (68)	19.5 (57)	.64
Diagnosis, % (n)			
Schizophrenia	51.8 (200)		
Schizoaffective	47.2 (182)		
Psychosis NOS	1.0 (4)		
Medication Type <sup>c</sup> , % (n)			
Typical	12.7 (49)		
Atypical	74.6 (288)		
Combination	4.9 (19)		
No Antipsychotic	7.8 (30)		
PANSS			
Positive	16.3 ± 5.3		
Negative	14.1 ± 5.4		
General	31.9 ± 8.0		
Total	62.2 ± 14.8		

Note: <sup>a</sup>Chi-squared for categorical variables (sex, race/ethnicity),  $t$ -test for continuous variables; <sup>b</sup>individuals were able to identify more than one race/ethnicity; <sup>c</sup>Medication information was unavailable for 7 participants. WRAT - Wide Range Achievement Test – 3rd Edition, PANSS = Positive and Negative Syndrome Scale, SSD = schizophrenia spectrum disorder, HC = healthy control; all values presented are  $M \pm SD$  unless otherwise noted.

**Table 2**  
Social cognition tasks and functional outcome measures by diagnostic group.

	SSD	HC	p-value
<b>Social Cognition Tasks</b>			
AIHQ <sup>a</sup>	8.7 ± 2.9	7.0 ± 2.4	<.01
BLERT	13.6 ± 4.1	15.8 ± 2.7	<.01
ER40	30.5 ± 5.1	33.0 ± 3.4	<.01
Eyes	21.0 ± 5.5	24.3 ± 4.5	<.01
Hinting	13.4 ± 3.8	16.1 ± 2.5	<.01
IBT <sup>b</sup>	44.5 ± 17.9	40.4 ± 14.6	.03
RAD <sup>a</sup>	24.8 ± 5.8	29.8 ± 5.2	<.01
TASIT <sup>c</sup>	44.6 ± 7.6	51.1 ± 6.3	<.01
<b>Neurocognition Tasks</b>			
Animal Naming	19.2 ± 5.7	22.6 ± 5.9	<.01
HVLT	20.8 ± 5.6	25.1 ± 4.6	<.01
Letter Number Sequence	11.9 ± 4.1	14.6 ± 3.8	<.01
Symbol Coding	42.9 ± 11.6	53.2 ± 12.5	<.01
Trails A	40.4 ± 18.3	31.5 ± 10.9	<.01
<b>Functional Outcome Measures</b>			
<u>SSPA Average</u>			
SSPA1	4.1 ± 0.5	4.6 ± 0.4	<.01
SSPA2	4.2 ± 0.6	4.6 ± 0.4	<.01
<u>SLOF Self-Report Average</u>			
Interpersonal Relationships	4.0 ± 0.6	4.6 ± 0.4	<.01
Social Acceptability	3.6 ± 0.9	4.1 ± 0.7	<.01
Activities of Community Living	4.5 ± 0.5	4.7 ± 0.4	<.01
Work Skills	4.4 ± 0.8	4.8 ± 0.5	<.01
<u>SLOF Informant Average<sup>d</sup></u>			
Interpersonal Relationships	4.1 ± 0.6	4.7 ± 0.5	<.01
Social Acceptability	3.4 ± 0.9	4.4 ± 0.6	<.01
Activities of Community Living	4.4 ± 0.8	4.8 ± 0.5	<.01
Work Skills	4.1 ± 0.8	4.7 ± 0.5	<.01
UPSA-B Total <sup>d</sup>	4.0 ± 0.6	4.7 ± 0.5	<.01
UPSA-B Total <sup>d</sup>	70.3 ± 14.2		

Note: <sup>a</sup>only collected during initial phase; <sup>b</sup>IBT only collected during final phase; <sup>c</sup>task version counterbalanced across visits, values presented are for Form A; <sup>d</sup>only administered to SSD group; AIHQ = Ambiguous Intentions and Hostility Questionnaire Blame Index, BLERT = Bell Lysaker Emotion Recognition Task, ER40 = Penn Emotion Recognition Task, Eyes = Reading the Mind in the Eyes Task, Hinting = Hinting Task, IBT = Intentionality Bias Task, RAD = Relationships Across Domains Task; TASIT = The Awareness of Social Inference Test, HVLT = Hopkins Verbal Learning Test, SSPA = Social Skills Performance Assessment, SLOF = Specific Levels of Functioning, UPSA-B = UCSD Performance-Based Skills Assessment - Brief, HC = healthy control, SSD = schizophrenia spectrum disorder; all values presented are  $M \pm SD$ .



**Fig. 1.** Item and Test Characteristic Curves.

Note: Dashed lines represent original task items, solid lines represent items retained for brief versions of tasks; Test information curves presented separately by task factor for multidimensional tasks (i.e., IBT, TASIT, RAD); AIHQ = Ambiguous Intentions and Hostility Questionnaire, BLERT = Bell Lysaker Emotion Recognition Task, ER40 = Penn Emotion Recognition Task, Eyes = Reading the Mind in the Eyes Task, Hinting = Hinting Task, IBT = Intentionality Bias Task, RAD = Relationship Across Domains Task, TASIT = The Awareness of Social Inference Test.



**Table 3**  
Performance on brief social cognition tasks by diagnostic group.

	Brief Task Characteristics			Group Differences			Relationship with Original Task	
	Items	Range	SSD	HC	t-test	Cohen's d	p	Pearson's r
<b>Attributional Bias</b>								
AIHQ - B	15	3–16	8.7 ± 2.9	7.0 ± 2.4	t(303.6) = 7.1	0.64	<.01	–
IBT - B	14	0–100	44.4 ± 17.2	40.6 ± 14.8	t(278) = 1.9	0.23	.05	.91
<b>Emotion Processing</b>								
BLERT - B	10	0–10	6.4 ± 2.3	7.5 ± 1.7	t(674.1) = 7.4	0.55	<.01	.91
ER40 - B	18	0–18	13.4 ± 3.2	15.0 ± 2.1	t(666.5) = 7.8	0.64	<.01	.92
<b>Social Perception</b>								
RAD - B	21	0–21	11.8 ± 3.5	14.7 ± 2.8	t(253.9) = 7.6	0.79	<.01	.91
<b>Theory of Mind</b>								
Eyes - B	18	0–18	11.2 ± 3.6	13.3 ± 2.8	t(674.8) = 8.4	0.64	<.01	.91
Hinting - B	8	0–16	11.1 ± 3.2	13.2 ± 2.3	t(672.6) = 9.9	0.76	<.01	.96
TASIT-S	36	0–36	24.0 ± 4.7	27.5 ± 4.1	t(345.9) = 10.4	0.80	<.01	.93

Note: AIHQ-B = Ambiguous Intentions and Hostility Questionnaire Blame Index- Brief, BLERT-B = Bell Lysaker Emotion Recognition Task - Brief, ER40-B = Penn Emotion Recognition Task - Brief, Eyes-B = Reading the Mind in the Eyes Task - Brief, Hinting-B = Hinting Task-Brief, IBT-B = Intentionality Bias Task - Brief, RAD-B = Relationships Across Domains Task - Brief; TASIT-S= The Awareness of Social Inference Test Part III – short version, SSD = schizophrenia spectrum disorder, HC = healthy control; all values presented are M ± SD. AIHQ-B includes only Likert items from AIHQ (i.e., not rater-scored items) so no comparison (i.e., Pearson's r) with original task presented.

**Table 4**  
Comparison of brief social cognition tasks by diagnostic group.

Brief Social Cognition Task	HC	Internal Consistency	Test-Retest	Repeated Measures Utility			Relationship with Original Task
	SSD			Repeated t-test	Cohen's d	% Floor/Chance	
	n	McDonald's ω	Pearson's r				
<b>Attributional Bias</b>							
AIHQ - B	138	.86	.76**	t(126) = 4.14**	0.37	-	-
	228	.87	.67**	t(213) = 1.75**	0.12	-	-
IBT - B	150	.57	.50**	t(135) = 3.04**	0.26	-	.90**
	130	.65	.58**	t(116) = 2.90**	0.27	-	.92**
<b>Emotion Processing</b>							
BLERT - B	292	.53	.61**	t(274) = 2.65**	0.16	0	8.22
	386	.71	.69**	t(361) = 4.86**	0.26	2.59	5.96
ER40 - B	292	.55	.60**	t(274) = 0.98	0.06	0	4.79
	386	.72	.71**	t(360) = 4.13**	0.22	0.26	1.04
<b>Social Perception</b>							
RAD - B	104	.63	.66**	t(97) = 0.35	0.04	9.62	0
	175	.76	.66**	t(165) = 1.97	0.15	40.57	0.57
<b>Theory of Mind</b>							
Eyes - B	292	.62	.71**	t(274) = 0.49,	0.03	0.34	3.08
	386	.73	.76**	t(361) = 0.75	0.04	3.63	1.55
Hinting - B	288	.65	.61**	t(273) = 3.02**	0.18	0	12.15
	384	.68	.65**	t(359) = 3.37**	0.18	0.26	2.08
TASIT - S	286	.64	-	-	-	1.40	1.05
	378	.74	-	-	-	7.14	0

Note: \*p<.01; Floor and ceiling effects not reported for measures of attributional bias (i.e., AIHQ-B and IBT-B) since scores reflect bias rather than correct or incorrect responses, TASIT-S test-retest and repeated t-test not reported since TASIT-S is based on TASIT Form A which was only administered to each participant once; HC = healthy control group; SSD = schizophrenia spectrum disorder group; AIHQ-B = Ambiguous Intentions and Hostility Questionnaire Blame Index- Brief, BLERT-B = Bell Lysaker Emotion Recognition Task - Brief, ER40-B = Penn Emotion Recognition Task - Brief, Eyes-B = Reading the Mind in the Eyes Task - Brief, Hinting-B = Hinting Task-Brief, IBT-B = Intentionality Bias Task - Brief, RAD-B = Relationships Across Domains Task - Brief; TASIT-S= The Awareness of Social Inference Test Short Form.

2.2. Evaluation of brief social cognition tasks

HC outperformed SSD on all brief SC tasks except for the brief IBT (IBT-B; see Table 3). Brief SC tasks exhibited better psychometric properties within SSD compared with HC (see Table 4). In general, brief SC tasks retained similar, albeit slightly reduced, psychometric properties and demonstrated strong relationships with original tasks (i.e., Pearson's rs .87 - .96, ps < .05; see Table 5). The brief ER40 (ER40-B) and the brief RAD (RAD-B) demonstrated reduced relationships with indicators of functioning compared with original task performance while the brief Eyes (Eyes-B) demonstrated improved incremental validity and stronger relationships with indicators of functioning.

2.3. Internal consistency

Most brief SC tasks retained good internal consistency (i.e., McDonald's coefficient omegas [ω] = .71 - .88) with higher internal consistency observed in SSD compared with HC, as also seen in SCOPE. The IBT-B demonstrated acceptable internal consistency (ω = .63), similar to the original IBT (ω = .68).

2.4. Test-retest reliability

Test-retest reliability was acceptable (rs .66-.76), except for the IBT-B which demonstrated poor test-retest reliability (r = .55), similar to the original IBT (r = .58). Test-retest reliability for brief tasks was slightly higher within HC compared to SSD.

**Table 5**  
Comparison of original and brief social cognition tasks.

SC Task	Items	Original	Internal	Test-	Repeated Measures Utility			Convergent	Divergent	Relationships with	Incremental	Time
		Brief	Consistency	Retest	Repeated	Cohen's	%	Validity	Validity	Indicators of	Validity	Minutes
		ω	r	t-test	D	Floor/ Chance	% Ceiling	Proportion of SC with <i>r</i> s <i>p</i> < .05	Proportion of NC with <i>r</i> s <i>p</i> < .05	Functional Outcomes Predicted	Functional Outcomes Predicted (after NC)	
Attributional Bias												
AIHQ	15	-	-	-	-	-	-	-	-	-	-	-
	<b>15</b>	<b>.88</b>	<b>.73</b> **	<i>t</i> (340) = <b>3.5</b> **	<b>0.19</b>	-	-	<b>2/6</b>	<b>0/5</b>	<b>SLOF-Inf, SSPA</b>	<b>SSPA</b>	<b>6.35</b>
IBT	24	.68	.58**	<i>t</i> (245) = 3.9**	0.25	-	-	3/5	2/5	UPSA-B	UPSA-B	5.43
	<b>14</b>	<b>.63</b>	<b>.55</b> **	<i>t</i> (252) = <b>4.2</b> **	0.26	-	-	1/5	1/5	SSPA, UPSA-B	UPSA-B	3.16
Emotion Processing												
BLERT	21	.76	.75**	<i>t</i> (636) = 7.4**	0.29	0	1.18	7/7	5/5	SLOF-Inf	SLOF-Inf	7.09
	<b>10</b>	<b>.69</b>	<b>.68</b> **	<i>t</i> (636) = <b>5.5</b> **	<b>0.22</b>	<b>1.48</b>	<b>6.93</b>	<b>6/7</b>	<b>5/5</b>	<b>SLOF-Inf</b>	<b>SLOF-Inf</b>	<b>3.37</b>
ER40	40	.81	.74**	<i>t</i> (625) = 4.1**	0.16	0.15	0.15	7/7	5/5	SSPA	-	3.21
	<b>18</b>	<b>.71</b>	<b>.70</b> **	<i>t</i> (635) = <b>3.9</b> **	0.15	0.15	2.65	5/7	5/5	-	-	1.44
Social Perception												
RAD	45	.78	.79**	<i>t</i> (263) = 3.6**	0.22	31.9	0	6/6	5/5	UPSA-B	-	15.84
	<b>21</b>	<b>.76</b>	<b>.71</b> **	<i>t</i> (263) = <b>1.9</b>	0.11	29.03	0.36	5/6	5/5	-	-	7.39
Theory of Mind												
Eyes	36	.74	.79**	<i>t</i> (636) = 0.3	0.01	1.18	0	7/7	5/5	-	-	6.56
	<b>18</b>	<b>.72</b>	<b>.76</b> **	<i>t</i> (636) = <b>0.3</b>	0.01	2.21	2.21	7/7	5/5	UPSA-B	UPSA-B	3.28
Hinting	<b>10</b>	<b>.72</b>	<b>.68</b> **	<i>t</i> (633) = <b>4.8</b> **	<b>0.19</b>	<b>0.15</b>	<b>2.83</b>	<b>7/7</b>	<b>5/5</b>	<b>SSPA, UPSA-B</b>	<b>SSPA, UPSA-B</b>	<b>6.13</b>
	8	.71	.66**	<i>t</i> (633) = 4.5**	0.18	0.15	6.40	5/7	5/5	SSPA, UPSA-B	SLOF-SR, SSPA, UPSA-B	4.90
TASIT	64	.82	.64**	<i>t</i> (634) = 5.3**	0.21	3.23	0	6/7	5/5	SLOF-Inf, UPSA-B	-	17.92
	<b>36</b>	<b>.71</b>	-	-	-	4.46	0.46	5/7	5/5	SLOF-Inf, SSPA, UPSA-B	-	10.08

Note: \*\**p* < .01; Bolded values indicate tasks selected for the final brief battery of social cognition (BB-SCOPE); Floor and ceiling effects not reported for measures of attributional bias (i.e., AIHQ-B and IBT-B) since scores reflect bias rather than correct or incorrect responses, TASIT-S test-retest and repeated t-test not reported since TASIT-S is based on TASIT Form A which was only administered to each participant once; AIHQ-B = Ambiguous Intentions and Hostility Questionnaire Blame Index- Brief, BLERT-B = Bell Lysaker Emotion Recognition Task - Brief, ER40-B = Penn Emotion Recognition Task - Brief, Eyes-B = Reading the Mind in the Eyes Task - Brief, Hinting-B = Hinting Task - Brief, IBT-B = Intentionality Bias Task - Brief, RAD-B = Relationships Across Domains Task - Brief; TASIT-S = The Awareness of Social Inference Test Short Form.

2.5. Utility as a repeated measure

Similar performance was observed for brief measures compared with original measures. Brief tasks demonstrated significant differences in performance between testing visits (repeated t-test *ps* < .05) except for the RAD-B and the Eyes-B. Effect sizes for repeated performance were small (i.e., Cohen's *ds* 0.04 - 0.27 in SSD) and similar to original tasks. Effect sizes for repeated performance were generally larger in SSD compared with HC except for the brief Hinting task (Hinting-B) and brief AIHQ (AIHQ-B) which demonstrated smaller or identical effect sizes in SSD compared with HC.

Floor/chance and ceiling affects were mostly limited and comparable between brief and original tasks except for the RAD-B (29.05% performance at chance). Pronounced chance performance effects observed were similar to the original RAD (31.90% performance at chance). Overall, SSD had slightly higher floor/chance and lower ceiling performance compared with HC.

2.6. Convergent validity

Brief SC tasks demonstrated significant relationships with other brief SC measures (see Table 6).

2.7. Divergent validity

Divergent validity of brief SC tasks was less impressive with most brief SC tasks demonstrating significant relationships with NC tasks in SSD, similar to original SC tasks (see Table 6). Measures of attributional bias demonstrated stronger divergent validity.

2.8. Relationships with indicators of functioning

Correlations with functional outcomes in SSD are presented in Table 6. Brief SC tasks demonstrated correlations with functional outcomes similar to original versions. Brief SC tasks also demonstrated

unique relationships with functional outcomes in a series of linear regressions with the exception of the ER40-B and the RAD-B (see Table 7). Follow-up analyses examined relationships between brief SC tasks and high-quality informants (i.e., high-contact family members or friends) with similar results with the exception of no significant relationships observed between the informant SLOF and the AIHQ-B.

2.9. Incremental validity

The AIHQ-B, IBT-B, BLERT-B, Eyes-B, and Hinting-B demonstrated incremental validity after controlling for NC task performance (see Table 7). Altogether, brief SC tasks explained an additional 1–12% of variance in functional outcomes. These results were similar to incremental validity observed in original SC tasks with the exception that Eyes-B demonstrated incremental validity with the UPSA-B, whereas the original Eyes did not demonstrate incremental validity.

2.10. Final BB-SCOPE recommendations

Three tasks were identified for inclusion in BB-SCOPE based on parsimonious representation of one task per SC domain, good psychometric properties, and unique relationships with functional outcomes: the AIHQ-B, the BLERT-B, and the original Hinting task. The Hinting-B only achieved adequate psychometric properties with inclusion of 80% of original items. With a reduced length of just 1 min, a recommendation is made to administer the full-length Hinting Task which can also be considered brief.

2.11. Utility of a BB-SCOPE total score

Total scores from the AIHQ-B, BLERT-B, and the Hinting task were divided by total respective possible scores, multiplied to weight each task equally, and then summed for a BB-SCOPE total score ranging from 0 to 100 (see Appendix A for example scoring and scoring template). The SSD group (*M* = 61.29, *SD* = 12.49) scored significantly lower than the

**Table 6**  
Social cognition, neurocognition, and functional outcome correlations in schizophrenia spectrum disorders.

	Attributional Bias		Emotion Processing		Social Perception	Theory of Mind			
	AIHQ-B	IBT-B	BLERT -B	ER40-B	RAD-B	Eyes-B	Hinting-B	TASIT-S	
<b>Brief Social Cognition Tasks</b>									
<b>Convergent Validity</b>	AIHQ	-	-	-.15*	-.07	-.08	-.14*	.04	-.04
	IBT	-	-	-.19	-.30	-	-.18*	-.12	-.03
	BLERT	-.18**	-.16	-	.59**	.39**	.58**	.35**	.47**
	ER40	-.07	-.27**	.64**	-	.36**	.56**	.31**	.39**
	RAD	-.07	-	.47**	.45**	-	.51**	.31**	.48**
	Eyes	-.15*	-.16	.61**	.56**	.58**	-	.41**	.45**
	Hinting	.04	-.14	.38**	.36**	.35**	.39**	-	.39**
	TASIT	-.05	-.04	.54**	.48**	.61**	.57**	.40**	-
<b>Divergent Validity</b>	Animal Naming	-.13	-.04	.34**	.29**	.25**	.37**	.22**	.34**
	HVLT-R	-.03	-.19*	.43**	.38**	.51**	.40**	.30**	.44**
	Letter Number Span	-.10	-.12	.45**	.40**	.55**	.48**	.35**	.43**
	Symbol Coding	-.02	-.12	.38**	.35**	.39**	.42**	.25**	.43**
	Trails A	.04	-.08	-.27**	-.24**	-.24**	-.28**	-.17**	-.28**
<b>Relationships with Indicators of Functioning</b>	<u>SLOF Informant Total</u>	-.13	-.15	.23**	.14*	.17*	.13*	.15*	.19**
	Activities of Community Living	-.15	-.14	.17*	.13*	.15	.10	.13*	.11
	Interpersonal Relationships	-.05	-.05	.19**	.10	.08	.09	.12*	.14*
	Social Acceptability	-.16*	-.13	.11	.04	.03	.10	.01	.01
	Work Skills	.02	-.14	.24**	.14*	.22**	.13*	.17**	.25**
	<u>SLOF Self-Report Total</u>	-	.03	.12	.13	-	.05	-.08	-.05
	Activities of Community Living	-	-.05	.22*	.28**	-	.25**	.08	.12
	Interpersonal Relationships	-	.08	-.01	-.03	-	-.11	-.13	-.22**
	Social Acceptability	-	-.08	.02	-.01	-	-.06	-.13	.01
	Work Skills	-	.15	.02	.02	-	-.07	-.12	-.10
	<u>SSPA Average</u>	.12	-.27**	.35**	.33**	.23**	.32**	.41**	.35**
	SSPA 1	.04	.19*	.30**	.29**	.11	.28**	.32**	.29**
	SSPA 2	.17*	-.28**	.31**	.31**	.32**	.29**	.40**	.33**
	<u>UPSA-B Total</u>	.03	-.31**	.36**	.34**	.40**	.43**	.43**	.39**

Note: Correlations between original social cognition tasks presented in gray, SLOF Self-Report not administered during same phases at AIHQ and RAD task administration; \* $p < .05$ , \*\* $p < .01$ . AIHQ = Ambiguous Intentions and Hostility Questionnaire, BLERT = Bell Lysaker Emotion Recognition Task, ER40 = Penn Emotion Recognition Task, Eyes = Reading the Mind in the Eyes Task, Hinting = Hinting Task, IBT = Intentionality Bias Task, RAD = Relationship Across Domains Task, TASIT = The Awareness of Social Inference Test, -B and -S after task names indicate brief versions of tasks, HVLT = Hopkins Verbal Learning Test, SSPA = Social Skills Performance Assessment, SLOF = Specific Levels of Functioning, UPSA-B = UCSD Performance-Based Skills Assessment - Brief, SSD = schizophrenia spectrum disorder.

HC group ( $M = 67.13$ ,  $SD = 7.82$ ;  $t(361.5) = 5.49$ ,  $p < .01$ ). ROC analyses identified the utility of BB-SCOPE total scores predicting impaired functioning defined as one standard deviation below HC performance and a cut score of 60 for the UPSA-B based on previous literature (Mausbach et al., 2007) since this measure was not administered to the HC group (see Table 8). Balancing optimal cutoff scores across indices and a score less than one standard deviation below HC performance, a cutoff score of 60 was recommended to identify SC impairment associated with impairments in functional capacity and social competence (see Supplementary Fig. 1 for distribution of SSD and HC BB-SCOPE total scores).

### 3. Discussion

The aim of the present study was to develop a brief battery to identify SC impairment in SSD to reduce heterogeneity of SC measurement and facilitate more widespread assessment. To this end, BB-SCOPE consists of three tasks (AIHQ-B, BLERT-B, original Hinting) for a total administration time of 15 min to assess three SC domains of attributional bias, emotion perception and processing, and theory of mind. The BB-SCOPE is also easy to score (see Appendix A) with a recommended cutoff score of 60 for identifying marked SC impairment associated with impairments in functional capacity and social competence.

The AIHQ-B was selected to assess attributional bias due to good psychometric properties and utility in uniquely predicting social competence beyond NC performance with good divergent validity. The AIHQ-B was selected over the IBT-B based on better psychometric properties and high level of missingness (close to 10%) observed in the IBT-B due to limited response times. Recommendation of the AIHQ-B replicates previous work demonstrating improved performance of the

AIHQ when using only the Likert items and a unique relationship with suspiciousness/persecution (Buck et al., 2016, 2017).

The BLERT-B and ER40-B tasks demonstrated comparable psychometric properties, but the BLERT-B was selected to assess emotion processing due to utility in uniquely predicting real-world functioning beyond NC performance. The ER40-B did not demonstrate incremental validity. The BLERT-B did not demonstrate divergent validity; however, recent findings suggest NC and SC may be less distinct than previously thought, and therefore relationships observed between the BLERT-B and NC may accurately reflect shared variance (Deckler et al., 2018).

The original Hinting task was selected over the Eyes-B and TASIT-S to assess theory of mind based on stronger relationships with indicators of functioning and incremental validity. The TASIT-S did not demonstrate incremental validity. Additionally, prior work demonstrates that performance on the Eyes task is influenced by social class and culture, and that this task may be better conceptualized as an assessment of cognitive performance or emotion recognition than SC (Deckler et al., 2018; Dodell-Feder et al., 2020; Kittel et al., 2021; Oakley et al., 2016).

No measures of social perception were recommended for inclusion in BB-SCOPE. The RAD-B demonstrated acceptable consistency and test-retest reliability but pronounced chance effects and limited validity.

#### 3.1. Limitations

One limitation is the use of HC and SSD samples to estimate item parameters and psychometric properties. Although final recommendations for BB-SCOPE were primarily based on performance within SSD, the decision to include HC was done to maximize ability range (e.g., individuals high and low in SC ability) to better estimate performance of

**Table 7**  
Regressions predicting functional outcomes in schizophrenia spectrum disorders.

	Model Predictors	Predicting Functional Outcomes				Incremental Validity Predicting Functional Outcomes Beyond NC Performance				
		SLOF Informant	SLOF Self-Report	SSPA	UPSA-B	SLOF Informant	SLOF Self-Report	SSPA	UPSA-B	
All Phases (n = 316)	<u>Block: NC Tasks</u>									
	Adjusted R <sup>2</sup>	–	–	–	–	<b>.06</b>	<.01	<b>.16</b>	<b>.30</b>	
	<u>Block: SC Tasks<sup>a</sup></u>									
	Adjusted R <sup>2</sup>	<b>.06</b>	<b>.02</b>	<b>.24</b>	<b>.28</b>	$\Delta < .01$	$\Delta < .01$	$\Delta .08$	$\Delta .06$	
	<u>SC Brief Tasks (<math>\beta</math>)</u>									
	BLERT-B	<b>.19</b>	.20	.11	.06	<b>.19</b>	.14	.06	-.01	
	ER40-B	-.02	.13	.10	.05	-.06	.15	.06	.02	
	Eyes-B	-.06	-.06	.04	<b>.17</b>	-.07	-.06	.01	<b>.12</b>	
Phases Three & Four (n = 163)	<u>Block: SC Tasks<sup>b</sup></u>									
	Adjusted R <sup>2</sup>	<b>.11</b>	–	<b>.20</b>	<b>.30</b>	$\Delta .04$	–	$\Delta .09$	$\Delta .08$	
	<u>SC Brief Tasks (<math>\beta</math>)</u>									
	AIHQ-B	-.16	–	.17	.05	-.14	–	.21	.06	
	RAD-B	-.01	–	.01	.15	-.01	–	-.10	.01	
	Phase Five (n = 123)	<u>Block: SC Tasks<sup>c</sup></u>								
		Adjusted R <sup>2</sup>	<b>.11</b>	<b>.01</b>	<b>.30</b>	<b>.31</b>	$\Delta .05$	$\Delta < .01$	$\Delta .12$	$\Delta .07$
		<u>SC Brief Tasks (<math>\beta</math>)</u>								
IBT-B		-.10	.10	-.17	-.25	-.06	.15	-.13	-.20	

Note: <sup>a</sup>Model includes tasks from all phases (i.e., BLERT, ER40, Eyes, Hinting, TASIT), <sup>b</sup>Model includes tasks administered during all phases with RAD and AIHQ-B, <sup>c</sup>Model includes tasks administered during all phases with IBT-B;  $\Delta$  indicates change in adjusted R<sup>2</sup> after accounting for NC block; VIF values for all predictors across models <2.5 indicating acceptable multicollinearity, bold values indicate p < .05; SLOF Self-Report not administered during phases Three and Four; AIHQ-B = Ambiguous Intentions and Hostility Questionnaire Blame Index, BLERT-B = Bell Lysaker Emotion Recognition Task, ER40-B = Penn Emotion Recognition Task, Eyes = Reading the Mind in the Eyes Task, Hinting = Hinting Task, IBT-B = Intentionality Bias Task, RAD = Relationship Across Domains Task, TASIT-S = The Awareness of Social Inference Test, SSPA = Social Skills Performance Assessment, SLOF = Specific Levels of Functioning, UPSA-B = UCSD Performance-Based Skills Assessment - Brief.

**Table 8**  
Utility of BB-SCOPE total score predicting impairment across functional indices.

Predicted Outcome	Optimal Cutoff Score <sup>b</sup>	AUC [95% CI]	Sensitivity	Specificity	60 Cutoff Score	Sensitivity	Specificity
SSPA Total Score < 1SD below average HC	58	.69 [.62, .76]	.63	.65	60	.60	.67
SLOF Self-Report total score <1 SD below average HC	66	.62 [.47, .78]	.18	.84	60	.37	.78
SLOF Informant-Report total score <1 SD below average HC <sup>a</sup>	51	.53 [.44, .61]	.50	.51	60	.39	.60
UPSA-B total score <60	61	.75 [.68, .83]	.80	.59	60	.80	.59

Note: A total BB-SCOPE score of 59.31 is consistent with performance less than one standard deviation below the average score of healthy control participants, <sup>a</sup>the same criteria for impaired SLOF Self-Report was used (i.e., cutoff score of 4.2) since the SLOF informant-report was only administered to informants of SSD participants, <sup>b</sup>optimal cutoff score determined using Youden’s Index; SSPA = Social Skills Performance Assessment, SLOF = Specific Levels of Functioning, UPSA-B = UCSD Performance-Based Skills Assessment – Brief.

items as they relate to the full spectrum of SC ability. Previous work demonstrates that the majority of individuals with SSD, but not all, experience impairment in SSD (Hajdúk et al., 2018). Therefore, inclusion of a HC group accounts for task performance estimates of individuals that may not experience SC impairment.

Additionally, BB-SCOPE tests share considerably more variance with performance-based measures of competence (SSPA, UPSA-B) than with ratings of real-world functioning. This was also the case in SCOPE and abbreviating the battery has not increased the correlation between SC

measures and rating scale indices of everyday outcomes. Thus, the broad limitation that traditional assessment of SC impairments manifest somewhat limited relationships with real-world functioning applies here as well. While BB-SCOPE demonstrates utility for assessing SC impairment, there are informant and self-report measures of SC such as the Observable Social Cognition Rating Scale (OSCARS; Halverson et al., 2020; Healey et al., 2015), similar to the Schizophrenia Cognition Rating Scale (SCoRS; Keefe et al., 2006) for NC, which demonstrate stronger relationships with indicators of functioning (Jones et al., 2019;



Silberstein et al., 2018). However, in the SCOPE studies, obtaining high quality clinician informant reports was possible for fewer than half of the participants, cautioning reliance on informant reports, and self-reports of cognitive abilities can be biased in SSD participants (Burton et al., 2016; Lysaker et al., 2013).

Additionally, more nuanced approaches to SC assessment such as the inclusion of introspective accuracy (self-evaluation of one's own ability compared to observed ability, (Harvey and Pinkham, 2015) and overconfidence (overestimation of task performance captured by confidence ratings, Silberstein and Harvey, 2019) have been shown to improve correlational relationships with functioning, performance on other tasks, and improved differentiation of SSD from HC individuals (Badal et al., 2021; Perez et al., 2020). While BB-SCOPE offers an efficient and straightforward approach to establishing marked SC impairment, other approaches should be considered, especially when there is a focus on SC ability or change over time (rather than identification of impairment) and relationships with functional outcomes.

A final limitation is that the estimates and psychometric properties of brief SC tasks are based on secondary analysis of the SCOPE Study. While the estimated time of BB-SCOPE is 15 min, BB-SCOPE needs to be validated with additional data collection. Future research administering BB-SCOPE can provide information about observed battery duration and agreement with estimated psychometric properties.

### 3.2. Future directions

While BB-SCOPE was developed to improve dissemination of SC assessment, there is ample opportunity for the improvement of brief assessment of SC. Importantly, there remains a gap in the measurement of social perception with few tasks demonstrating acceptable psychometric properties and unique relationships with functioning. Another future direction is the development of brief SC tasks with demonstrated sensitivity to change appropriate for use in clinical trials, rather than the focus of BB-SCOPE on identification of marked SC impairment (Vaskinn and Horan, 2020). The BB-SCOPE may be best utilized as a short screen to identify severe SC in SSD and to better understand if remedial interventions are broadly effective in this group.

A final future direction for consideration is the underlying structure of SC in SSD. The present study conceptualized SC as a four-dimensional construct (i.e., attribution bias, emotion processing, social perception, and theory of mind) identified by experts in the field of SC in SSD (Pinkham et al., 2014), although more recent research indicates simpler structures of SC may also be appropriate (Buck et al., 2016a; Mike et al., 2019; Riedel et al., 2021). Future work exploring the underlying factor structure of SC may identify more parsimonious and efficient SC batteries to assess SC in SSD.

## 4. Conclusion

The BB-SCOPE is a brief battery for identifying severe SC impairment which includes three tasks (i.e., AIHQ-B, BLERT-B, Hinting Task – full length) assessing three domains of SC (i.e., attribution bias, emotion processing, and theory of mind). BB-SCOPE is an efficient battery with a duration of 15 min, good psychometric properties, and good criterion and predictive validity. Development of BB-SCOPE is meant to improve dissemination of SC assessment as well as decrease heterogeneity in identification of SC impairment. To this end, BB-SCOPE now meets most criteria outlined by the National Institute of Mental Health when selecting the Computerized Neurocognitive Battery as a measure of neurocognition for the PhenX Toolkit for psychosis: comprehensive, easily administered, scalable in a variety of settings, and established psychometric properties in SSD (Öngür et al., 2020). The BB-SCOPE is also an efficient battery for use in clinical practice to identify individuals with marked SC impairment that may optimally benefit from psychosocial interventions targeting SC and related constructs to improve functioning.

## Funding

National Institute of Mental Health At National Institute of Health (R01 MH093432 to PHD, DLP, AEP). Dr. Halverson is funded by the Department of Veterans Affairs Office of Academic Affiliations Mid-Atlantic Mental Illness Research, Education, and Clinical Center.

## Credit author statement

Tate Halverson: Conceptualization, Formal analysis, Writing – original draft, Visualization; Amy Pinkham: Conceptualization, Methodology, Investigation, Resources, Writing – reviewing and editing, Funding acquisition, Project administration; Philip Harvey: Conceptualization, Methodology, Investigation, Resources, Writing – reviewing and editing, Funding acquisition, Project administration; David Penn: Conceptualization, Methodology, Investigation, Writing – review and editing, Resources, Supervision, Funding acquisition, Project administration.

## Declaration of competing interest

Dr. Harvey has received consulting fees or travel reimbursements from Alkermes, Bio Excel, Boehringer Ingelheim, Karuna Pharma, Merck Pharma, Minerva Pharma, SK Pharma, and Sunovion Pharma during the past year. He receives royalties from the Brief Assessment of Cognition in Schizophrenia (Owned by Verasci, Inc. and contained in the MCCB). He is chief scientific officer of i-Function, Inc. Dr. Halverson, Dr. Pinkham, and Dr. Penn have no conflicts of interest to disclose.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jpsychires.2022.03.017>.

## References

- Akoglu, H., 2018. User's guide to correlation coefficients. *Turk. J. Emerg. Med.* 18, 91–93. <https://doi.org/10.1016/j.tjem.2018.08.001>.
- Badal, V.D., Depp, C.A., Hitchcock, P.F., Penn, D.L., Harvey, P.D., Pinkham, A.E., 2021. Computational methods for integrative evaluation of confidence, accuracy, and reaction time in facial affect recognition in schizophrenia. *Schizophr. Res. Cogn.* 25, 100196. <https://doi.org/10.1016/j.scog.2021.100196>.
- Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., Plumb, I., Hill, J., Plumb, I., Wheelwright, S., 2001. The "Reading the Mind in the Eyes" Test Revised Version: a study with normal adults, and adults with asperger syndrome or high-functioning autism. *JCPP (J. Child Psychol. Psychiatry)* 42, 241–251. <https://doi.org/10.1111/1469-7610.00715>.
- Bock, R.D., 1997. A brief history of item theory response. *Educ. Meas.* 16, 21–33. <https://doi.org/10.1111/j.1745-3992.1997.tb00605.x>.
- Bortolotti, S.L.V., Tezza, R., de Andrade, D.F., Borna, A.C., de Sousa Júnior, A.F., 2013. Relevance and advantages of using the item response theory. *Qual. Quantity* 47, 2341–2360. <https://doi.org/10.1007/s11135-012-9684-5>.
- Bryson, G., Bell, M., Lysaker, P., 1997. Affect recognition in schizophrenia: a function of global impairment or a specific cognitive deficit. *Psychiatr. Res.* 71, 105–113. [https://doi.org/10.1016/S0165-1781\(97\)00050-4](https://doi.org/10.1016/S0165-1781(97)00050-4).
- Buck, B., Kern, R.S., Marder, S.R., Penn, D.L., Healey, K.M., Green, M.F., Lee, J., Reise, S.P., Horan, W.P., Iwanski, C., Healey, K.M., Green, M.F., Horan, W.P., Kern, R.S., Lee, J., Marder, S.R., Reise, S.P., Penn, D.L., 2017. Improving measurement of attributional style in schizophrenia: A psychometric evaluation of the Ambiguous Intentions Hostility Questionnaire (AIHQ). *J. Psychiatr. Res.* 89, 48–54. <https://doi.org/10.1016/j.jpsychires.2017.01.004>.
- Buck, B.E., Healey, K.M., Gagen, E.C., Roberts, D.L., Penn, D.L., 2016a. Social cognition in schizophrenia: factor structure, clinical and functional correlates. *J. Ment. Health* 25, 330–337. <https://doi.org/10.3109/09638237.2015.1124397>.
- Buck, B.E., Pinkham, A.E., Harvey, P.D., Penn, D.L., 2016b. Revisiting the validity of measures of social cognitive bias in schizophrenia: additional results from the social cognition psychometric evaluation (SCOPE) study. *Br. J. Clin. Psychol.* 55, 441–454. <https://doi.org/10.1007/s11065-015-9294-9> (Functional).
- Burton, C.Z., Harvey, P.D., Patterson, T.L., Twamley, E.W., 2016. Neurocognitive insight and objective cognitive functioning in schizophrenia. *Schizophr. Res.* 171, 131–136. <https://doi.org/10.1016/j.schres.2016.01.021>.
- Chalmers, R.P., 2012. mirt: a multidimensional item response theory package for the R environment. *J. Stat. Software* 48, 1–29. <https://doi.org/10.18637/jss.v048.i06>.
- 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. *Cognit. Neuropsychiatry* 12, 128–143. <https://doi.org/10.1080/13546800600787854>.
- Corcoran, R., Mercer, G., Frith, C.D., 1995. Schizophrenia, symptomatology and social inference: investigating “theory of mind” in people with schizophrenia. *Schizophr. Res.* 17, 5–13. [https://doi.org/10.1016/0920-9964\(95\)0024-g](https://doi.org/10.1016/0920-9964(95)0024-g).
- Cornacchio, D., Pinkham, A.E., Penn, D.L., Harvey, P.D., 2017. Self-assessment of social cognition in individuals with schizophrenia: appraising task difficulty and allocation of effort. *Schizophr. Res.* 179, 85–90. <https://doi.org/10.1016/j.schres.2016.09.033>.
- Deckler, E., Hodgins, G.E., Pinkham, A.E., Penn, D.L., Harvey, P.D., 2018. Social cognition and neurocognition in schizophrenia and healthy controls: intercorrelations of performance and effects of manipulations aimed at increasing task difficulty. *Front. Psychiatr.* 9, 1–10. <https://doi.org/10.3389/fpsy.2018.00356>.
- DeVellis, R.F., 2017. *Scale Development: Theory and Applications*, fourth ed. Sage, Thousand Oaks, CA.
- Dodell-Feder, D., Ressler, K.J., Germine, L.T., 2020. Social cognition or social class and culture? On the interpretation of differences in social cognitive performance. *Psychol. Med.* 50, 133–145. <https://doi.org/10.1017/S003329171800404X>.
- Green, M.F., Horan, W.P., Lee, J., 2010. Social cognition in schizophrenia. *Curr. Dir. Psychol. Sci.* 19, 243–248. <https://doi.org/10.1038/nrn4005>.
- Hajdúk, M., Harvey, P.D., Penn, D.L., Pinkham, A.E., 2018. Social cognitive impairments in individuals with schizophrenia vary in severity. *J. Psychiatr. Res.* 104, 65–71. <https://doi.org/10.1016/j.jpsychires.2018.06.017>.
- Halverson, T.F., Hajdúk, M., Pinkham, A.E., Harvey, P.D., Jarskog, L.F., Nye, L., Penn, D.L., 2020. Psychometric properties of the observable social cognition rating scale (OSCARs): self-report and informant-rated social cognitive abilities in schizophrenia. *Psychiatr. Res.* 286, 112891 <https://doi.org/10.1016/j.psychres.2020.112891>.
- Halverson, T.F., Orleans-Pobee, M., Merritt, C., Sheeran, P., Fett, A.-K., Penn, D.L., 2019. Pathways to functional outcomes in schizophrenia spectrum disorders: meta-analysis of social cognitive and neurocognitive predictors. *Neurosci. Biobehav. Rev.* 105, 212–219. <https://doi.org/10.1016/j.neubiorev.2019.07.020>.
- Harvey, P.D., Penn, D., 2010. Social cognition: the key factor predicting social outcome in people with schizophrenia? *Psychiatry* 7, 41–44.
- Harvey, P.D., Pinkham, A., 2015. Impaired self-assessment in schizophrenia: why patients misjudge their cognition and functioning. *Curr. Psychiatr.* 14, 53–59.
- Healey, K.M., Combs, D.R., Gibson, C.M., Keefe, R.S.E., Roberts, D.L., Penn, D.L., 2015. Observable social cognition - a rating scale: an interview-based assessment for schizophrenia. *Cognit. Neuropsychiatry* 20, 198–221. <https://doi.org/10.1080/13546805.2014.999915>.
- Honan, C.A., McDonald, S., Sufani, C., Hine, D.W., Kumfor, F., 2016. The awareness of social inference test: development of a shortened version for use in adults with acquired brain injury. *Clin. Neuropsychol.* 30, 243–264. <https://doi.org/10.1080/13854046.2015.1136691>.
- Jones, M.T., Deckler, E., Laurrari, C., Jarskog, L.F., Penn, D.L., Pinkham, A.E., Harvey, P.D., 2019. Confidence, performance, and accuracy of self-assessment of social cognition: a comparison of schizophrenia patients and healthy controls. *Schizophr. Res. Cogn.* <https://doi.org/10.1016/j.scog.2019.01.002>, 0–1.
- Kay, S.R., Fiszbein, A., Opler, L.A., 1987. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr. Bull.* 13, 261–276. <https://doi.org/10.1016/j.schres.2003.09.011>.
- Keefe, R.S.E., Goldberg, T.E., Harvey, P.D., Gold, J.M., Poe, M.P., Coughenour, L., 2004. The Brief Assessment of Cognition in Schizophrenia: reliability, sensitivity, and comparison with a standard neurocognitive battery. *Schizophr. Res.* 68, 283–297. <https://doi.org/10.1016/j.schres.2003.09.011>.
- Keefe, R.S.E., Poe, M., Walker, T.M., Kang, J.W., Harvey, P.D., 2006. The Schizophrenia Cognition Rating Scale: an interview-based assessment and its relationship to cognition, real-world functioning, and functional capacity. *Am. J. Psychiatr.* 163, 426–432. <https://doi.org/10.1176/appi.ajp.163.3.426>.
- Kittel, A.F.D., Olderbak, S., Wilhelm, O., 2021. Sty in the Mind’s Eye: A Meta-Analytic Investigation of the Nomological Network and Internal Consistency of the “Reading the Mind in the Eyes” Test. *Assessment*. <https://doi.org/10.1177/1073191121996469> (ahead of print).
- Klein, H.S., Springfield, C.R., Bass, E., Ludwig, K., Penn, D.L., Harvey, P.D., Pinkham, A.E., 2020. Measuring mentalizing: a comparison of scoring methods for the hinting task. *Int. J. Methods Psychiatr. Res.* 29, 1–11. <https://doi.org/10.1002/mpr.1827>.
- 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. Psychiatr.* 160, 1768–1774. <https://doi.org/10.1176/appi.ajp.160.10.1768>.
- Kraemer, H.C., Kupfer, D.J., Clarke, D.E., Narrow, W.E., Regier, D.A., 2012. DSM-5: how reliable is reliable enough? *Am. J. Psychiatr.* 169, 13–15. <https://doi.org/10.1176/appi.ajp.2011.11010050>.
- Kuha, J., 2004. AIC and BIC: comparisons of assumptions and performance. *Socio. Methods Res.* 33, 188–229. <https://doi.org/10.1177/0049124103262065>.
- Lysaker, P.H., Vohs, J., Ohayon, J., Kukla, M., Wierwille, J., Dimaggio, G., 2013. Depression and insight in schizophrenia: comparisons of levels of deficits in social cognition and metacognition and internalized stigma across three profiles. *Schizophr. Res.* 148, 18–23. <https://doi.org/10.1016/j.schres.2013.05.025>.
- 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. *Schizophr. Bull.* 33, 1364–1372. <https://doi.org/10.1093/schbul/sbm014>.
- 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. <https://doi.org/10.1097/00001199-200305000-00001>.
- Mike, L., Guimond, S., Kelly, S., Thermenos, H., Mesholam-Gately, R., Eack, S., Keshavan, M., 2019. Social cognition in early course of schizophrenia: exploratory factor analysis. *Psychiatr. Res.* 272, 737–743. <https://doi.org/10.1016/j.psychres.2018.12.152>.
- Nuechterlein, K.H., Green, M.F., Kern, R.S., Baade, L.E., Barch, D.M., Cohen, J.D., Essock, S., Fenton, W.S., Frese 3rd, F.J., Gold, J.M., Goldberg, T., Heaton, R.K., Keefe, R.S.E., Kraemer, H., Mesholam-Gately, R., Seidman, L.J., Stover, E., Weinberger, D.R., Young, A.S., Zalcman, S., Marder, S.R., 2008. The MATRICS Consensus Cognitive Battery, part 1: test selection, reliability, and validity. *Am. J. Psychiatr.* 165, 203–213. <https://doi.org/10.1176/appi.ajp.2007.07010042>.
- Oakley, B.F.M., Brewer, R., Bird, G., Catmur, C., 2016. Theory of mind is not theory of emotion: a cautionary note on the Reading the Mind in the Eyes Test. *J. Abnorm. Psychol.* 125, 818. <https://doi.org/10.1037/abn0000182>.
- Öngür, D., Carter, C.S., Gur, R.E., Perkins, D., Sawa, A., Seidman, L.J., Tamminga, C., Huggins, W., Hamilton, C., 2020. Common data elements for national institute of mental health-funded translational early psychosis research. *Biol. Psychiatr. Cogn. Neurosci. Neuroimaging* 5, 10–22. <https://doi.org/10.1016/j.bpsc.2019.06.009>.
- Patterson, T.L., Moscona, S., McKibbin, C.L., Davidson, K., Jeste, D.V., 2001. Social skills performance assessment among older patients with schizophrenia. *Schizophr. Res.* 48, 351–360. [https://doi.org/10.1016/s0920-9964\(00\)00109-2](https://doi.org/10.1016/s0920-9964(00)00109-2).
- Penn, D.L., Corrigan, P.W., Bentall, R.P., Racenstein, J.M., Newman, L., 1997. Social cognition in schizophrenia. *Psychol. Bull.* 121, 114–132. <https://doi.org/10.1037/0033-2909.121.1.114>.
- Perez, M.M., Tercero, B.A., Penn, D.L., Pinkham, A.E., Harvey, P.D., 2020. Overconfidence in social cognitive decision making: correlations with social cognitive and neurocognitive performance in participants with schizophrenia and healthy individuals. *Schizophr. Res.* 224, 51–57. <https://doi.org/10.1016/j.schres.2020.10.005>.
- Petrillo, J., Cano, S.J., McLeod, L.D., Coon, C.D., 2015. Using classical test theory, item response theory, and rasch measurement theory to evaluate patient-reported outcome measures: a comparison of worked examples. *Value Health* 18, 25–34. <https://doi.org/10.1016/j.jval.2014.10.005>.
- Pinkham, A.E., Harvey, P.D., Penn, D.L., 2018. Social cognition psychometric evaluation: results of the final validation study. *Schizophr. Bull.* 44, 737–748. <https://doi.org/10.1093/schbul/sbx117>.
- Pinkham, A.E., Harvey, P.D., Penn, D.L., 2016a. Paranoid individuals with schizophrenia show greater social cognitive bias and worse social functioning than non-paranoid individuals with schizophrenia. *Schizophr. Res. Cogn.* 3, 33–38. <https://doi.org/10.1016/j.scog.2015.11.002>.
- Pinkham, A.E., Penn, D.L., Green, M.F., Buck, B., Healey, K., Harvey, P.D., 2014. The social cognition psychometric evaluation study: results of the expert survey and RAND Panel. *Schizophr. Bull.* 40, 813–823. <https://doi.org/10.1093/schbul/sbt081>.
- Pinkham, A.E., Penn, D.L., Green, M.F., Harvey, P.D., 2016b. Social cognition psychometric evaluation: results of the initial psychometric study. *Schizophr. Bull.* 42, 494–504. <https://doi.org/10.1093/schbul/sbv056>.
- Riedel, P., Horan, W.P., Lee, J., Helleman, G.S., Green, M.F., 2021. The factor structure of social cognition in schizophrenia: a focus on replication with confirmatory factor analysis and machine learning. *Clin. Psychol. Sci.* 9, 38–52. <https://doi.org/10.1177/2167702620951527>.
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.C., Müller, M., 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinf.* 12, 77. <https://doi.org/10.1109/ACCESS.2018.2805869>.
- Rosset, E., 2008. It’s no accident: our bias for intentional explanations. *Cognition* 108, 771–780. <https://doi.org/10.1016/j.cognition.2008.07.001>.
- Schneider, L.C., Struening, E.L., 1983. SLOF: a behavioral rating scale for assessing the mentally ill. *Soc. Work. Res. Abstr.* 19, 9–21.
- 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. *Psychiatr. Res.* 166, 54–62. <https://doi.org/10.1016/j.psychres.2008.03.010>.
- Silberstein, J., Harvey, P.D., 2019. Cognition, social cognition, and self-assessment in schizophrenia: prediction of different elements of everyday functional outcomes. *CNS Spectr.* 1–6. <https://doi.org/10.1017/S1092852918001414>.
- Silberstein, J.M., Pinkham, A.E., Penn, D.L., Harvey, P.D., 2018. Self-assessment of social cognitive ability in schizophrenia: association with social cognitive test performance, informant assessments of social cognitive ability, and everyday outcomes. *Schizophr. Res.* 199, 75–82. <https://doi.org/10.1016/j.schres.2018.04.015>.
- Thissen, D., Orlando, M., 2001. *Item Response Theory for Items Scored in Two Categories*.
- Vaskinn, A., Horan, W.P., 2020. Social cognition and schizophrenia: unresolved issues and new challenges in a maturing field of research. *Schizophr. Bull.* 46, 464–470. <https://doi.org/10.1093/schbul/sbaa034>.
- Velthorst, E., Fett, A.K.J., Reichenberg, A., Perlman, G., Van Os, J., Bromet, E.J., Kotov, R., 2017. The 20-year longitudinal trajectories of social functioning in individuals with psychotic disorders. *Am. J. Psychiatr.* 174, 1075–1085. <https://doi.org/10.1176/appi.ajp.2016.15111419>.
- Ventura, J., Reise, S.P., Keefe, R.S.E., Baade, L.E., Gold, J.M., Green, M.F., Kern, R.S., Mesholam-Gately, R., Nuechterlein, K.H., Seidman, L.J., Bilder, R.M., 2010. The Cognitive Assessment Interview (CAI): development and validation of an empirically derived, brief interview-based measure of cognition. *Schizophr. Res.* 121, 24–31. <https://doi.org/10.1016/j.schres.2010.04.016>.
- Youden, W.J., 1950. Index for rating diagnostic tests. *Cancer* 3, 32–35.