



## Validity of remote administration of the MATRICS Consensus Cognitive Battery for individuals with severe mental illness

Madisen T. Russell<sup>a</sup>, Kensie M. Funsch<sup>a</sup>, Cassi R. Springfield<sup>a</sup>, Robert A. Ackerman<sup>a</sup>,  
Colin A. Depp<sup>b</sup>, Philip D. Harvey<sup>c,d</sup>, Raeanne C. Moore<sup>b</sup>, Amy E. Pinkham<sup>a,\*</sup>

<sup>a</sup> Department of Psychology, University of Texas at Dallas, 800 W. Campbell Rd., Richardson, TX 75080, United States of America

<sup>b</sup> Department of Psychiatry, University of California San Diego, 9500 Gilman Dr., La Jolla, CA, 92093, United States of America

<sup>c</sup> Department of Psychiatry and Behavioral Sciences, University of Miami Miller School of Medicine, 1120 NW 14th Street, Miami, FL 33136, United States of America

<sup>d</sup> Research Service, Bruce W. Carter VA Medical Center, Miami, FL 33136, United States of America

### ARTICLE INFO

#### Keywords:

Cognitive functioning  
Remote assessment  
Bipolar disorder  
Schizophrenia  
Telehealth

### ABSTRACT

The MATRICS Consensus Cognitive Battery (MCCB) is a gold-standard tool for assessing cognitive functioning in individuals with severe mental illness. This study is an initial examination of the validity of remote administration of 4 MCCB tests measuring processing speed (Trail Making Test: Part A, Animal Fluency), working memory (Letter-Number Span), and verbal learning and memory (Hopkins Verbal Learning Test-Revised). We conducted analyses on individuals with bipolar disorder (BD) and schizophrenia-spectrum disorders (SCZ), as well as healthy volunteers, who were assessed in-person (BD = 80, SCZ = 116, HV = 14) vs. remotely (BD = 93, SCZ = 43, HV = 30) to determine if there were significant differences in performance based on administration format. Additional analyses tested whether remote and in-person assessment performance was similarly correlated with symptom severity, cognitive and social cognitive performance, and functional outcomes. Individuals with BD performed significantly better than those with SCZ on all MCCB subtests across administration format. Animal Fluency did not differ by administration format, but remote participants performed significantly worse on Trail Making and HVLTR. On the Letter-Number Span task, individuals with bipolar disorder performed significantly better when participating remotely. Finally, patterns of correlations with related constructs were largely similar between administration formats. Thus, results suggest that remote administration of some of the MCCB subtests may be a valid alternative to in-person testing, but more research is necessary to determine why some tasks were affected by administration format.

### 1. Introduction

Cognitive impairment is a central feature of schizophrenia spectrum disorders (Bowie and Harvey, 2006). Individuals with schizophrenia demonstrate cognitive deficits across numerous domains, including attention, verbal learning and memory, processing speed, working memory, and executive functioning (e.g., Nuechterlein et al., 2004; Kahn and Keefe, 2013; Bowie and Harvey, 2006; Bora et al., 2009) with average weighted effect sizes (Hedge's *g*) ranging from 0.43 to 1.55 (Schaefer et al., 2013). Most patients with schizophrenia demonstrate cognitive impairment to some extent, though the breadth and severity of cognitive dysfunction varies (Bowie and Harvey, 2006). Within an individual, the level of cognitive impairment appears to be stable across time and fluctuations in clinical status (Harvey et al., 1999; Gold, 2004).

Similarly, bipolar disorder is also associated with cognitive deficits across the same domains (e.g., Cardenas et al., 2016; Bora and Ozerdem, 2017) with average weighted effect sizes (Hedge's *g*) ranging from 0.42 to 0.96 (Torres et al., 2007; Mann-Wrobel et al., 2011). Though cognitive impairment in bipolar disorder is relatively less severe than impairments observed in schizophrenia, cognitive profiles of both disorders are very similar (e.g., Bortolato et al., 2015; Krabbendam et al., 2005; Lynham et al., 2018; Bora and Pantelis, 2015; Reichenberg et al., 2008). Recent work examining cognitive impairment across the bipolar-schizophrenia spectrum suggests that cognitive dysfunction increased in severity from bipolar disorder to schizoaffective bipolar type, to schizophrenia and schizoaffective depressive type, with no differences in severity of cognitive impairment between schizophrenia and schizoaffective depressive type (Lynham et al., 2018). Overall, the

\* Corresponding author.

E-mail address: [amy.pinkham@utdallas.edu](mailto:amy.pinkham@utdallas.edu) (A.E. Pinkham).

<https://doi.org/10.1016/j.scog.2021.100226>

Received 28 September 2021; Received in revised form 17 November 2021; Accepted 17 November 2021

Available online 2 December 2021

2215-0013/© 2021 The Authors.

Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

trajectory of cognitive dysfunction in bipolar disorder appears somewhat similar to schizophrenia, with impairment beginning early in both disorders and remaining relatively stable over time after the diagnosis of the disorder (e.g., Bora and Ozerdem, 2017; Bora and Pantelis, 2015).

Importantly, cognitive dysfunction contributes significantly to functional disability in both schizophrenia and bipolar disorder. In schizophrenia, cognitive impairment is associated with poorer community living skills, deficits in problem-solving, and difficulty maintaining employment (Bryson and Bell, 2003; Green et al., 2000). Estimates derived from reviews of the literature suggest that neurocognitive dysfunction explains between 20% and 60% of the variance in functional outcomes of individuals with schizophrenia (Green et al., 2000; Fett et al., 2011). Similarly, cognitive disability accounts for a significant proportion of variation in functioning in bipolar disorder, with estimates consistent with those identified in schizophrenia (Depp et al., 2012). Although there is greater functional disability in schizophrenia relative to bipolar disorder, neurocognitive dysfunction still predicts poorer work skills, poorer community living skills, and difficulties in interpersonal behavior in bipolar disorder, with evidence suggesting that the structure of the correlational relationships are essentially identical (Bowie et al., 2010; Mausbach et al., 2010).

Given the cognitive deficits and associated functional outcomes seen in both bipolar disorder and schizophrenia spectrum disorders, it is important to have a standardized way to assess cognition in these groups. The MATRICS Consensus Cognitive Battery (MCCB; Nuechterlein et al., 2008) was developed to standardize assessment of cognitive impairment in schizophrenia and is typically considered the gold standard in the field. It contains ten tasks covering seven cognitive domains: processing speed, verbal learning, working memory, visual learning, reasoning/problem solving, social cognition, and attention. Although there have been suggestions regarding modifying the MCCB for bipolar disorder (Yatham et al., 2010), studies demonstrate that the MCCB is sensitive to cognitive impairment in both schizophrenia and bipolar disorder (Bo et al., 2017; Burdick et al., 2011; Kern et al., 2011; Lystad et al., 2014). Abbreviated forms of the MCCB (e.g., Pinkham et al., 2018) and similar neurocognitive assessments (Keefe et al., 2006) capture schizophrenia-related impairment while still showing expected correlations with functional outcomes.

In situations where in-person testing may not be feasible (e.g., due to health concerns, lack of transportation, limited funds, etc.), it is crucial to have a reliable, remote battery to assess cognitive functioning in individuals with severe mental illness. To our knowledge, no remote version of the MCCB has currently been developed, but web-based and smartphone app assessments designed to mirror widely used assessments like the MCCB show strong correlations with in-person measures in schizophrenia spectrum, bipolar, and healthy control populations (Biagianni et al., 2019; Domen et al., 2019; Miskowiak et al., 2021). Though they still require validation for at-home administration, these assessments appear to be comparable, albeit not identical, alternatives to traditional assessment. However, limited technology literacy and lack of access to technology and internet could hinder implementation of internet- and app-based assessments in certain populations. Telephone-based assessment offers a potentially more accessible solution but has received less attention within psychiatric populations. Notably, Berns et al. (2004) compared performance on an in-person and telephone-based cognitive battery in outpatients with schizophrenia. They found no difference between administration modes on tasks that were conceptually simple or that gradually increased in complexity, such as Letter Number Span (Gold, 1997). Tasks that were complex and demanding from the outset, however, showed poorer performance when administered by phone, such as the California Verbal Learning Test (CVLT; Delis et al., 1987).

At-home telephone assessment has been more thoroughly studied in non-psychiatric cognitively impaired populations. Telephone versions of two MCCB tasks, the Hopkins Verbal Learning Test-Revised (Brandt and Benedict, 2001) and a category fluency task, demonstrated strong

correlations with in-person assessments and good discrimination between cognitively impaired and healthy participants (e.g., Bunker et al., 2016; Lachman et al., 2014). Other tasks not contained within the MCCB, but which assess overlapping cognitive domains (e.g., verbal learning, memory, processing speed) have similarly shown good agreement between telephone and in-person administration (Jagtap et al., 2021; Lachman et al., 2014; Rapp et al., 2012).

Given these data suggesting that remote administration of cognitive assessments like the MCCB may be a feasible and valid alternative to in-person testing, the current paper presents results from an initial assessment of the validity of telephone-based administration of select MCCB subtests in individuals with schizophrenia/schizoaffective disorder and bipolar disorder. Task performance was compared between individuals who completed in-person assessments vs. those who completed them remotely, as well as between diagnoses. Correlations between task performance and symptoms and functional outcomes were also compared between administration formats. Based on previous findings, we anticipated that individuals with bipolar disorder would perform better than individuals with schizophrenia spectrum disorders. Similarly, given previous findings that telephone based assessments are comparable to in-person assessments, we also predicted minimal effects of administration format on task performance and similar patterns of correlations between most MCCB tasks and related constructs regardless of administration format. For HVLTL however, and based on Berns et al. (2004), we anticipated that performance could be poorer for remote administration in the SCZ group.

## 2. Methods

### 2.1. Participants

Participants were adults between the ages of 18 and 60 with schizophrenia/ schizoaffective disorder (SCZ), bipolar disorder I or II (BP), or non-psychiatric healthy controls (HV). Psychiatric diagnoses were confirmed via the Mini International Neuropsychiatric Interview (MINI) (Sheehan et al., 1998) and Structured Clinical Interview for DSM Disorders- Psychosis Module (SCID) (First et al., 2015). Individuals were required to be proficient in English, as well as to have had no psychiatric hospitalizations for at least 6 weeks, no significant medication regimen changes for a minimum of 6 weeks, and no dose changes >20% for a minimum of 2 weeks. Additionally, participants could not have (1) presence or history of medical or neurological disorders that may affect brain function (e.g., stroke, epilepsy), (2) presence or history of neurodegenerative disorder (e.g., dementia, Parkinson's Disease), (3) history of unconsciousness for a period greater than 15 min, (4) significant impairment of visual (e.g., blindness, glaucoma, vision uncorrectable to 20/40) or hearing (e.g., hearing loss) abilities, (5) presence or history of pervasive developmental disorder (e.g., autism) or intellectual disability (defined as IQ <70), or (6) current diagnosis of substance use disorder.

Data were collected across three sites between December 2018 and June 2021: The University of Texas at Dallas, University of Miami, and University of California, San Diego, resulting in a total of 376 participants. Data were collected as part of a larger study, during which COVID-19 related restrictions on in-person data collection necessitated a transition to remote assessment. Participants were separated into groups based on MCCB administration format and mental health diagnosis: 166 completed the MCCB subtests remotely (93 BD, 43 SCZ, 30 HV) and 210 completed the MCCB subtests in-person (80 BD, 116 SCZ, 14 HV).

### 2.2. Measures

#### 2.2.1. MATRICS Consensus Cognitive Battery (MCCB; Nuechterlein et al., 2008)

Participants completed four tests from the MCCB, two measuring processing speed (Trail Making Test: Part A, Category Fluency: Animal Naming), one assessing verbal working memory (Letter-Number Span),

and one assessing verbal learning and memory (Hopkin's Verbal Learning Test-Revised). The Trail Making Test (TMT): Part A (range 0–300 s) is a timed paper-and-pencil task in which participants draw a single line to consecutively connect numbered circles placed irregularly on a sheet of paper. The Category Fluency: Animal Naming test is an oral test in which participants name as many animals as they can in a one-minute period. The Letter-Number Span test (range 0–24) is an orally administered test in which the tester reads a string of numbers and letters, and the participant mentally reorders them (numbers consecutively, then letters alphabetically) and repeats them back. The HVLt-R (range 0–36) is an orally administered test in which the researcher reads aloud a list of 12 words from three different categories and the participant is asked to recall as many words as possible after each of three learning trials.

### 2.2.2. Social & cognitive measures

Measures assessing global cognitive, social cognitive, and real-world functioning were completed separately by an informant and the research coordinator. Informants were high-contact individuals who knew the participant well and who themselves did not have any psychiatric diagnoses (e.g., first-degree relative, significant other, close friend). All informant reports were collected via telephone. Research coordinators generated ratings using an “all-sources” approach consistent with Harvey et al. (2019) that integrated information gathered from interviews with the patients, informants, and their own experiences with the participants.

The Specific Levels of Functioning Scale (SLOF; Schneider and Streuening, 1983) is a 30-item survey assessing participants' functioning and behavior across 4 domains: interpersonal relationships, social acceptability, activities of community living, and work skills. Informants responded to items using a 5-point Likert scale, with higher mean values representing better functioning in each domain.

The Observable Social Cognition Rating Scale (OSCARS; Healey et al., 2015) is an 8-item self-report or interviewer assessment of ability across social cognitive domains (i.e., theory of mind, emotion perception, cognitive rigidity, jumping to conclusions, attributional style), yielding a total score ranging from 8 to 56, with higher scores indicating greater impairment.

The Cognitive Assessment Interview (CAI; Ventura et al., 2010) assesses subjective cognitive functioning across 6 domains (10-items): (1) working memory, (2) attention/vigilance, (3) verbal learning and memory, (4) reasoning and problem-solving, (5) speed of processing, and (6) social cognition. The CAI was administered to informants as an oral semi-structured interview, and ratings were made by the researcher according to participant/informant responses. In addition to a total score comprised of the sum of all items (range 7–70), a global assessment of function score is also given (range 0–100). Both indices were used here, with higher scores indicating worse cognitive functioning on the summed score and better cognitive functioning on the global assessment of function score.

### 2.2.3. Symptom assessments

Severity of positive, negative, and general symptoms was assessed with the Positive and Negative Syndrome Scale (PANSS; Kay et al., 1987). Mood symptoms were further assessed with the Montgomery-Asperg Depression Rating Scale (MADRS; Montgomery and Asberg, 1979) and the Young Mania Rating Scale (YMRS; Young et al., 1978). For all measures, higher scores indicate greater severity.

### 2.2.4. Premorbid intellectual functioning

Estimated premorbid IQ was assessed using the Wide Range of Achievement Test III (WRAT-III; Snelbaker et al., 2001) Reading subtest.

## 2.3. Procedures

All participants provided documented informed consent, and IRBs at

the University of Texas at Dallas, University of California San Diego, and University of Miami approved the study. In-person visits took place in labs on campus, while remote visits were done via telephone. Research staff had a bachelor's degree or higher, and were trained over the course of several weeks, within and across sites, to administer and score all assessments in-person and remotely. After establishing reliability (ICCs > 0.80), regular consensus meetings were held to ensure acceptable reliability between raters over time.

### 2.3.1. Remote visit task modifications

All tasks and interviews were completed primarily via telephone and required minimal modification. Those tasks that are typically administered orally (i.e., Animal Fluency, Letter-Number Span, HVLt-R) were implemented as is. The forms needed to complete the MCCB Trail Making Test were mailed to participants in advance of their appointments in separate, sealed envelopes with instructions to only open these materials when prompted and observed by the examiner. A supplemental video call via smartphone or tablet was used during the Trail Making Test so that researchers could accurately gauge participants' time to completion. Participants were texted or emailed a link to view the WRAT-III stimuli. PANSS ratings for 4 items that required prolonged visual behavioral observations (i.e., Blunted Affect, Tension, Mannerisms and Posturing, and Motor Retardation) were omitted from both in-person and remote participants' total scores.

Prior to task administration, participants were instructed to move to a quiet environment without distractions (e.g., away from other individuals, silencing/powering down extraneous devices) and researchers ensured that the participant could hear them well. Participants were also asked to refrain from using any performance aids, such as writing down stimulus items, or seeking help from others.

## 2.4. Statistical analyses

Groups were first split by diagnosis (BD, SCZ, HV), and demographic differences between administration format groups (remote vs. in-person administration) were assessed using independent sample *t*-tests or Chi-Square tests ( $\chi^2$ ) as appropriate. Extreme outliers ( $\pm 3$  SDs) on each task were excluded from analyses task-by-task, resulting in slight N differences between tasks (see Table 2). Because Trail Making scores are completion times, they were most likely to be outliers, and thus excluded than scores from other tasks. The numbers of participants performing at levels consistent with floor/ceiling effects on each of the MCCB subtests were also assessed to evaluate score distributions in each administration format.

Separate two-way analysis of covariance (ANCOVA) tests were then conducted to identify statistically significant effects of diagnosis (BD, SCZ) and administration format on MCCB test performance, controlling for PANSS symptom ratings (positive, negative). An additional independent-samples *t*-test was utilized to examine the effects of administration format among healthy participants. PANSS ratings were converted to averages for each participant, to account for the difference in number of items rated between administration types. To determine differences in the strength of associations between related constructs as a function of administration format, we analyzed comparisons between Pearson's *r* correlations by performing Fisher's *r* to *z* transformation, then calculating observed *z* values ( $z_{\text{observed}}$ ).

## 3. Results

### 3.1. Demographics

All groups were similar on age, gender, race, ethnicity, years of education, and estimated IQ (see Table 1). Compared to remote participants, in-person BD and SCZ groups had significantly higher ratings of positive symptoms ( $t_{\text{BD}}(171) = 3.946, p < .001$ ;  $t_{\text{SCZ}}(157) = 3.324, p = .001$ ), and in-person individuals with BD also had higher ratings of

**Table 1**  
Participant demographics.

	BD (n = 173)			SCZ (n = 159)			HV (n = 44)		
	N/M (%/SD)			N/M (%/SD)			N/M (%/SD)		
	In-person (n = 80)	Remote (n = 93)	t/x <sup>2</sup>	In-person (n = 116)	Remote (n = 43)	t/x <sup>2</sup>	In-person (n = 14)	Remote (n = 30)	t/x <sup>2</sup>
Gender									
Female	52 (65%)	64 (68.8%)	<i>p</i> = .594	64 (55.2%)	17 (39.5%)	<i>p</i> = .069	9 (64.3%)	15 (50%)	<i>p</i> = .322
Male	28 (35%)	29 (31.2%)		52 (44.8%)	25 (58.1%)		5 (35.7%)	15 (50%)	
Other	0	0		0	1 (2.3%)		0	0	
Age	40.5 (11.7)	37.8 (11.0)	<i>p</i> = .119	41.7 (10.6)	40.4 (12.0)	<i>p</i> = .494	35.9 (9.7)	37.0 (13)	<i>p</i> = .761
Race									
Caucasian	44 (55%)	60 (64.5%)	<i>p</i> = .265	38 (32.8%)	17 (39.5%)	<i>p</i> = .521	10 (71.4%)	15 (50%)	<i>p</i> = .239
African American	21 (26.3%)	15 (16.1%)		61 (52.6%)	23 (53.5%)		2 (14.3%)	11 (36.7%)	
American Indian/Alaskan Native	1 (1.3%)	0		1 (0.9%)	0		0	1 (3.3%)	
Asian	4 (5%)	9 (9.7%)		2 (1.7%)	1 (2.3%)		1 (7.1%)	3 (10%)	
Native Hawaiian/Other Pacific Islander	1 (1.3%)	0		1 (0.9%)	1 (2.3%)		0	0	
Other	9 (11.3%)	9 (9.7%)		13 (11.2%)	1 (2.3%)		1 (7.1%)	0	
Years of education	14.3 (2.6)	14.7 (2.2)	<i>p</i> = .303	12.6 (2.3)	12.9 (2.4)	<i>p</i> = .352	15 (1.1)	13.5 (1.8)	<i>p</i> = .199
IQ (WRAT-3)	101.4 (11.8)	103.8 (11.0)	<i>p</i> = .170	96.2 (12.3)	96.3 (12.3)	<i>p</i> = .973	104.1 (8.4)	101.1 (11.5)	<i>p</i> = .471
Ethnicity									
Hispanic	22 (27.5%)	25 (27.2%)	<i>p</i> = .962	30 (25.9%)	6 (14%)	<i>p</i> = .111	3 (21.4%)	8 (26.7%)	<i>p</i> = .752
Non-Hispanic	58 (72.5%)	67 (72.8%)		86 (74.1%)	37 (86%)		11 (78.6%)	22 (73.3%)	
Psychopathology									
PANSS- positive (mean)	1.9 (0.75)	1.5 (0.55)	<i>p</i> < .001**	2.4 (0.65)	2.0 (0.66)	<i>p</i> = .001**	–	–	
PANSS- negative (mean)	1.6 (0.45)	1.4 (0.46)	<i>p</i> = .004**	2.0 (0.59)	1.8 (0.69)	<i>p</i> = .206*	–	–	
PANSS- general (mean)	2.1 (0.54)	1.9 (0.41)	<i>p</i> = .056	2.0 (0.50)	2.0 (0.38)	<i>p</i> = .923	–	–	
MADRS	13.1 (10.8)	13.0 (10.3)	<i>p</i> = .939	9.7 (10.7)	11.7 (10.2)	<i>p</i> = .288	–	–	
YMRS	3.3 (4.6)	2.9 (4.3)	<i>p</i> = .545	0.9 (3.1)	1.8 (4.6)	<i>p</i> = .186	–	–	
Site									
UT Dallas	27 (33.8%)	46 (49.5%)		45 (38.8%)	24 (55.8%)		14 (100%)	30 (100%)	
U Miami	14 (17.5%)	16 (17.2%)		43 (37.1%)	9 (20.9%)		0	0	
UCSD	39 (48.8%)	31 (33.3%)		28 (24.1%)	10 (23.3%)		0	0	

Note. PANSS = Positive and Negative Syndrome Scale; MADRS = Montgomery-Asperg Depression Rating Scale; YMRS = Young Mania Rating Scale.

\* *p* < .05.  
\*\* *p* < .01.

negative symptoms ( $t(171) = 2.945, p = .004$ ).

### 3.2. Score distributions

Outliers represented less than 5% of the overall sample and were distributed relatively evenly across groups (see Table 2). No participants in either administration format performed at ceiling or floor on any of the tasks.

**Table 2**  
In-person and remote MCCB test scores and effect sizes.

	BD					SCZ					HV				
	In-person		Remote		Cohen's d	In-person		Remote		Cohen's d	In-person		Remote		Cohen's d
	N	M (SD)	N	M (SD)		N	M (SD)	N	M (SD)		N	M (SD)	N	M (SD)	
TMT-A	79	28.86 (10.98)	88	33.70 (14.90)	0.37*	114	36.07 (16.41)	35	46.33 (28.09)	0.45**	13	30.50 (10.59)	29	38.20 (13.62)	0.63
LNS	80	12.93 (3.63)	93	14.99 (3.75)	0.56**	116	11.80 (4.02)	43	11.86 (4.36)	0.01	14	15.00 (4.06)	30	15.93 (2.85)	0.27
Animal fluency	80	23.45 (5.84)	92	24.32 (6.33)	0.14	115	19.93 (5.79)	43	20.81 (6.43)	0.14	14	25.57 (5.61)	30	22.87 (4.53)	0.53
HVLT-R	79	23.99 (6.17)	91	22.79 (5.35)	0.21	114	20.19 (5.61)	43	19.00 (6.51)	0.20	14	27.86 (4.09)	30	24.87 (4.85)	0.67

Note. TMT-A = Trail Making Test- Part A (completion times; higher scores indicate worse performance); LNS = Letter-Number Span (range 0–24; higher scores indicate better performance); Animal Fluency (range 0–48; higher scores indicate better performance); HVLT-R = Hopkins Verbal Learning Test- Revised (range 0–36; higher scores indicate better performance).

Cohen's  $d = (M_2 - M_1) / SD_{pooled}$  where  $SD_{pooled} = \sqrt{((SD_1^2 + SD_2^2) / 2)}$ ; small effect, 0.20; medium effect, 0.50; large effect, 0.80.

Six scores were excluded from TMT-A (1 BD in-person, 1 BD remote, 2 SCZ in-person, 1 SCZ remote, 1 HV in-person) and 2 scores were excluded from HVLT-R (1 BD in-person, 1 SCZ in-person).

Due to various extraneous circumstances, twelve individuals did not complete the TMT-A (4 BD remote, 7 SCZ remote, 1 HV remote), two did not complete Animal Fluency (1 BD remote, 1 SCZ in-person), and one individual did not complete the HVLT-R (1 SCZ in-person).

\* *p* < .05.  
\*\* *p* < .01.



324) = 12.985,  $p < .001$ ), HVLTR ( $F(1,321) = 12.056, p < .001$ ), showing that individuals with bipolar disorder performed significantly better than individuals with schizophrenia/schizoaffective disorders across all tasks. There was a significant main effect of administration format on Trail Making ( $F(1, 310) = 22.393, p < .001$ ) and HVLTR ( $F(1, 321) = 6.499, p = .007$ ), with remote participants performing worse on both tasks. No other tasks showed significant main effects of format. The only task that had a significant interaction between diagnosis and administration format was Letter-Number Span ( $F(1, 326) = 4.487, p = .05$ ), indicating that individuals with bipolar disorder performed significantly better on this task when it was administered remotely.

There were no statistically significant differences in MCCB task performance by administration format in the healthy volunteer group (all  $p$  values  $> .05$ ).

### 3.4. Relationship to symptom severity

Across diagnostic groups and administration formats, higher severity of negative symptoms significantly correlated with poorer performance on several MCCB tests. In BD, positive symptoms, depression (MADRS), and mania (YMRS) did not correlate with performance on any MCCB tests (Table 3). In the SCZ group, increased positive symptoms significantly correlated with higher scores on Animal Fluency, regardless of administration format, and increased depressive symptoms were positively correlated with HVLTR performance in the remote group (Table 4).

### 3.5. Relationship to functional outcomes

Across administration types and diagnostic groups, both informant- and RA-ratings of SLOF, OSCARS, and CAI were significantly associated with performance on several MCCB tests at different correlation strengths. As expected, correlations with MCCB tasks were strongest for the CAI, which assesses cognitive functioning (Tables 3 & 4).

### 3.6. Differences in strengths of associations

There were relatively few differences in correlation strengths across

administration formats (see Supplemental Table 1). For the BD group, Trail Making showed the highest number of discrepancies based on remote vs. in-person administration, with 4 pairs of correlations (out of 19) showing significant differences. Letter-Number Span had no differences, whereas Animal Fluency and HVLTR each had one (see Table 3). For the SCZ group, Letter-Number Span and HVLTR had 4 and 3 discrepancies, respectively, between administration formats. Trail Making and Animal Fluency each had only one (see Table 4). Discrepancies were not concentrated in any particular domains; however, within the SCZ group, most discrepancies occurred for positive and negative symptoms, with stronger correlations in the remote group as compared to in-person.

## 4. Discussion

With advances in technology and potential limitations to in-person testing, adaptation of cognitive functioning assessments for remote administration may be a viable option for assessing cognitive abilities in individuals with severe mental illness. This study provides an initial assessment of the validity of remote administration of select MCCB subtests (Trail Making, Letter-Number Span, Animal Fluency, and HVLTR) in individuals with schizophrenia-spectrum disorders and bipolar disorder.

As anticipated, the bipolar group performed significantly better than the schizophrenia-spectrum group on all MCCB tasks, regardless of administration format, supporting previous research findings that individuals with bipolar disorder have higher levels of cognitive functioning than individuals with schizophrenia-spectrum disorders (Bortolato et al., 2015; Krabbendam et al., 2005; Lynham et al., 2018; Bora and Pantelis, 2015). Additionally, this finding provides some validation for remote telephone administration of the MCCB given that group differences were evident in both formats.

Further supporting the validity of remote assessment, we found that, across diagnostic groups and administration formats, MCCB task performance was significantly correlated with symptom severity (especially negative symptoms), social functioning, and overall cognitive functioning, which is in line with previous research findings (e.g., August et al., 2012; Harvey et al., 2006). While the strength of some correlations varied between administration formats, only 8.89% of these strength

**Table 3**  
Task correlations for in-person and remote administration in BD.

	TMT-A		LNS		Animal fluency		HVLTR	
	In-person	Remote	In-person	Remote	In-person	Remote	In-person	Remote
Age	0.36**	0.20	-0.31**	-0.09	-0.13	-0.11	-0.35**	-0.29**
YOE	<b>0.09</b>	<b>-0.19</b>	0.23*	0.17	0.35**	0.27*	<b>0.37**</b>	<b>0.10</b>
PANSS positive	0.12	-0.04	-0.01	-0.12	-0.04	0.05	-0.17	-0.06
PANSS negative	0.07	0.31**	-0.29**	-0.10	-0.37**	-0.27*	<b>-0.39**</b>	<b>-0.12</b>
PANSS general	0.14	-0.10	-0.14	-0.08	-0.26*	-0.06	-0.16	-0.04
PANSS total	0.15	0.10	-0.17	-0.15	-0.26*	-0.13	-0.30**	-0.11
MADRS	0.08	0.03	0.04	0.02	-0.16	-0.06	-0.05	0.13
YMRS	<b>0.12</b>	<b>-0.14</b>	-0.08	0.01	-0.02	0.11	-0.10	-0.05
LNS	-0.19	-0.09	-	-	-	-	-	-
Animal fluency	-0.27*	-0.35**	0.47**	0.36**	-	-	-	-
HVLTR	-0.24*	-0.17	0.44**	0.58**	<b>0.57**</b>	<b>0.32**</b>	-	-
SLOF- RA	-0.20	-0.24*	0.23*	0.08	0.29**	0.08	0.30**	0.08
OSCARs- RA	0.13	0.26*	-0.23*	-0.24*	-0.14	-0.12	-0.26*	-0.33**
CAI overall- RA	0.23*	0.25*	-0.16	-0.38**	-0.25*	-0.21	-0.25*	-0.32**
CAI GAF- RA	-0.11	-0.23*	0.22	0.40**	0.25*	0.11	0.30**	0.38**
SLOF- informant	-0.17	-0.36**	-0.13	0.11	0.08	0.09	-0.06	0.08
OSCARs- informant	0.12	0.37**	-0.05	-0.16	-0.07	-0.22	-0.04	-0.23*
CAI overall- informant	<b>0.08</b>	<b>0.41**</b>	0.08	-0.16	-0.02	-0.20	0.01	-0.14
CAI GAF- informant	<b>-0.03</b>	<b>-0.28*</b>	-0.12	0.16	0.01	0.19	0.002	0.16

Note. Bold text indicates correlations that differ significantly in strength between in-person and remote formats. TMT-A = Trail Making Test- Part A; LNS = Letter-Number Span; HVLTR = Hopkins Verbal Learning Test- Revised; YOE = years of education; PANSS = Positive and Negative Syndrome Scale; MADRS = Montgomery-Asperg Depression Rating Scale; YMRS = Young Mania Rating Scale; SLOF = Specific Levels of Functioning; OSCARS = Observable Social Cognition: A Rating Scale; CAI = Cognitive Assessment Interview; GAF = Global Assessment of Functioning.

\*  $p < .05$ .

\*\*  $p < .01$ .

**Table 4**  
Task correlations for in-person and remote administration in SCZ.

	TMT-A		LNS		Animal fluency		HVLt-R	
	In-person	Remote	In-person	Remote	In-person	Remote	In-person	Remote
Age	0.14	-0.01	-0.16	-0.07	<b>-0.03</b>	<b>-0.34*</b>	-0.05	-0.03
YOE	-0.14	-0.11	0.25**	0.31*	0.16	0.16	0.37**	0.17
PANSS positive	0.18	-0.11	<b>-0.14</b>	<b>0.23</b>	0.24*	0.31*	<b>-0.09</b>	<b>0.22</b>
PANSS negative	<b>0.25**</b>	<b>0.57**</b>	<b>-0.25**</b>	<b>-0.55**</b>	<b>-0.36**</b>	<b>-0.52**</b>	<b>-0.09</b>	<b>-0.50**</b>
PANSS general	0.01	0.09	-0.19*	-0.04	0.17	-0.04	-0.17	-0.02
PANSS total	0.22*	0.38*	<b>-0.28**</b>	-0.23	0.02	-0.16	0.01	-0.20
MADRS	-0.06	-0.28	-0.02	0.19	0.07	0.22	0.03	0.30*
YMRS	-0.04	0.03	0.01	-0.02	0.04	0.15	0.04	0.28
LNS	<b>-0.33**</b>	<b>-0.55**</b>	-	-	-	-	-	-
Animal fluency	<b>-0.35**</b>	<b>-0.35*</b>	0.29**	0.49**	-	-	-	-
HVLt-R	<b>-0.41**</b>	<b>-0.41*</b>	0.51**	0.47**	0.39**	0.56**	-	-
SLOF- RA	-0.19*	-0.04	0.28**	0.23	0.09	0.09	0.16	0.26
OSCARS- RA	0.25**	-0.01	<b>-0.29**</b>	-0.06	-0.09	0.04	-0.16	-0.22
CAI overall- RA	0.30**	0.40*	<b>-0.21*</b>	<b>-0.49**</b>	<b>-0.23*</b>	-0.31	-0.16	<b>-0.37*</b>
CAI GAF- RA	<b>-0.28**</b>	<b>-0.37*</b>	0.27**	0.41**	0.23*	0.18	0.15	0.30
SLOF- informant	-0.15	-0.10	0.25*	0.05	0.25*	0.001	0.001	0.07
OSCARS- informant	0.09	0.34	-0.20	0.03	<b>-0.24*</b>	0.09	0.03	-0.25
CAI overall- informant	0.23*	0.06	<b>-0.08</b>	<b>-0.45*</b>	<b>-0.24*</b>	-0.11	0.06	-0.16
CAI GAF- informant	<b>-0.22*</b>	-0.15	0.08	0.33	0.29**	0.05	<b>-0.16</b>	<b>0.29</b>

Note. Bold text indicates correlations that differ significantly in strength between in-person and remote formats. TMT-A = Trail Making Test- Part A; LNS = Letter-Number Span; HVLt-R = Hopkins Verbal Learning Test- Revised; YOE = years of education; PANSS = Positive and Negative Syndrome Scale; MADRS = Montgomery-Asperg Depression Rating Scale; YMRS = Young Mania Rating Scale; SLOF = Specific Levels of Functioning; OSCARS = Observable Social Cognition: A Rating Scale; CAI = Cognitive Assessment Interview; GAF = Global Assessment of Functioning.

\*  $p < .05$ .

\*\*  $p < .01$ .

differences were significant in the BD group, and 11.11% were significantly different in the SCZ group, suggesting comparable patterns of correlation for the majority of tasks. While it is not possible to draw definitive conclusions from the current data, it is possible that having a social versus non-social environment during testing, as well as symptom severity, and task attention/engagement may explain some of the correlation strength differences. Future research should look at factors that may moderate remote versus in-person cognitive performance's relationship with symptom severity, as well as social and non-social functioning.

In terms of specific subtests, administration format did not appear to affect performance on the Animal Fluency task. This test has relatively short and simple instructions and does not require any back-and-forth between the administrator and participant, which may explain why we did not see any significant difference in performance between administration types. Thus, Animal Fluency can be validly administered via telephone.

However, performance on both HVLt-R and Trail Making were worse when administered remotely. As noted previously, tasks that are complex and demanding from the outset, like HVLt, may be more difficult when administered by phone versus in-person, and similar findings have been reported for the CVLT (Berns et al., 2004). Slower completion of remotely administered Trail Making may be related to reduced control over participants' testing environments, as well as technological difficulties (e.g., difficulty setting up devices for video call, potential lag in video call making corrections processes take more time, assessor not able to point directly at participants' papers during mistake corrections, etc.). Future studies administering these two tasks may consider conducting thorough prescreening to ensure strong audio and video connections or additional practice trials to ensure participant understanding.

Individuals with bipolar disorder performed better on the Letter-Number Span task when administered remotely versus in-person, but the reasons for this are unclear. We did not see the same pattern in individuals with schizophrenia, concurrent with Berns et al.'s (2004) findings. While premorbid IQ was slightly higher in the remote BD group, accounting for variability related to IQ had only a minimal effect on the results, increasing the  $p$ -value very minimally ( $p = .057$ ). IQ

differences are therefore unlikely to account for the interaction effect. Because administrators were not on video with participants during this task, it is also possible that participants could have been cheating; however, because BD remote participants did worse on HVLt-R, for which they could have also written down the items, this seems unlikely.

Some limitations require consideration. First, only between-subject comparisons were assessed. Definitive attempts to examine the validity of remote assessments would require within-person comparisons between formats. Second, our sample of healthy individuals was relatively small, as was the SCZ remote group. Third, while the current analyses addressed sensitivity to group differences, relationships to functional outcomes, and floor/ceiling effects, a full psychometric analysis that allows examination of test-retest reliability and utility as repeated measures is still needed. Fourth, the COVID-19 pandemic provided impetus for the adaptation of our measures to remote administration, and remote data collection occurred exclusively during the pandemic. Therefore, it is possible that differences in administration formats may be confounded with presence of the pandemic. Finally, while the prospect of remote assessment has many feasible benefits, it is important to note that this format may also be less accessible than in-person testing to different demographic groups (e.g., those with reduced access to video calling, etc.). Overall, while not definitive, our results suggest remote telephone-based administration of some MCCB tests may be a feasible and valid method for assessing cognitive functioning in individuals with bipolar disorder or schizophrenia spectrum disorders.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.scog.2021.100226>.

#### Funding source

This work was supported by the National Institute of Mental Health (grant number R01 MH112620 to A.P.).

#### CRedit authorship contribution statement

**Madisen T. Russell:** Formal analysis; Investigation; Resources; Data curation; Writing- Original draft; Writing- Review & Editing;

Visualization **Kensie M. Funsch**: Investigation; Resources; Writing-Original draft **Cassi R. Springfield**: Resources; Writing-Original draft **Robert A. Ackerman**: Formal analysis; Writing- Review & Editing **Colin A. Depp**: Conceptualization; Methodology; Resources; Writing- Review & Editing; Supervision; Project administration **Philip D. Harvey**: Conceptualization; Methodology; Resources; Writing- Review & Editing; Supervision; Project administration **Raeanne C. Moore**: Conceptualization; Methodology; Writing- Review & Editing; Supervision **Amy E. Pinkham**: Conceptualization; Methodology; Resources; Formal analysis; Writing- Review & Editing; 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, 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. Moore is a co-founder of KeyWise, Inc. and a consultant for NeuroUX. All other authors report no conflicts of interest.

## References

- August, S.M., Kiwanuka, J.N., McMahon, R.P., Gold, J.M., 2012. The MATRICS Consensus Cognitive Battery (MCCB): Clinical and cognitive correlates. *Schizophr. Res.* 134 (1), 76–82. <https://doi.org/10.1016/j.schres.2011.10.015>.
- Berns, S., Davis-Conway, S., Jaeger, J., 2004. Telephone administration of neuropsychological tests can facilitate studies in schizophrenia. *Schizophr. Res.* 71, 505–506. <https://doi.org/10.1016/j.schres.2004.03.023>.
- Biagianti, B., Fisher, M., Brandrett, B., et al., 2019. Development and testing of a web-based battery to remotely assess cognitive health in individuals with schizophrenia. *Schizophr. Res.* 208, 250–257. <https://doi.org/10.1016/j.schres.2019.01.047>.
- Bo, Q., Mao, Z., Li, X., Wang, Z., Wang, C., Ma, X., 2017. Use of the MATRICS consensus cognitive battery (MCCB) to evaluate cognitive deficits in bipolar disorder: a systematic review and meta-analysis. *PLoS ONE* 12 (4), 1–13. <https://doi.org/10.1371/journal.pone.0176212>.
- Bora, E., Ozerdem, A., 2017. Meta-analysis of longitudinal studies of cognition in bipolar disorder: comparison with healthy controls and schizophrenia. *Psychol. Med.* 47, 2753–2766. <https://doi.org/10.1017/S0033291717001490>.
- Bora, E., Pantelis, C., 2015. Meta-analysis of cognitive impairment in first-episode bipolar disorder: comparison with first-episode schizophrenia and healthy controls. *Schizophr. Bull.* 41, 1095–1104.
- Bora, E., Yucel, M., Pantelis, C., 2009. Cognitive functioning in schizophrenia, schizoaffective disorder and affective psychoses: meta-analytic study. *Br. J. Psychiatry* 195, 475–482. <https://doi.org/10.1192/bjp.bp.108.055731>.
- Bortolato, B., Miskowiak, K., Köhler, C., Vieta, E., Carvalho, A.F., 2015. Cognitive dysfunction in bipolar disorder and schizophrenia: a systematic review of meta-analyses. *Neuropsychiatr. Dis. Treat.* 3111–3125 <https://doi.org/10.2147/ndt.s76700>.
- Bowie, C.R., Harvey, P.D., 2006. Cognitive deficits and functional outcome in schizophrenia. *Neuropsychiatr. Dis. Treat.* 2 (4), 531–536. <https://doi.org/10.2147/ndt.2006.2.4.531>.
- Bowie, C.R., Depp, C., McGrath, J.A., et al., 2010. Prediction of real-world functional disability in chronic mental disorders: a comparison of schizophrenia and bipolar disorder. *Am. J. Psychiatry* 167 (9), 1116–1124. <https://doi.org/10.1176/appi.ajp.2010.09101406>.
- Brandt, J., Benedict, R.H.B., 2001. Hopkins verbal learning test- revised. *Psychological Assessment Resources*.
- Bryson, G., Bell, M.D., 2003. Initial and final work performance in schizophrenia: cognitive and symptom predictors. *J. Nerv. Ment. Dis.* 191 (2), 87–92. <https://doi.org/10.1097/01.NMD.0000050937.06332.3C>.
- Bunker, L., Hshieh, T.T., Wong, B., et al., 2016. The SAGES telephone neuropsychological battery: correlation with in-person measures. *Int. J. Geriatr. Psychiatry* 32, 991–999. <https://doi.org/10.1002/gps.4558>.
- Burdick, K.E., Goldberg, T.E., Cornblatt, B.A., et al., 2011. The MATRICS consensus cognitive battery in patients with bipolar I disorder. *Neuropsychopharmacology* 36, 1587–1592. <https://doi.org/10.1038/npp.2011.36>.
- Cardenas, S.A., Kassem, L., Brotman, M.A., Leibenluft, E., McMahon, F.J., 2016. Neurocognitive functioning in euthymic patients with bipolar disorder and unaffected relatives: a review of the literature. *Neurosci. Biobehav. Rev.* 69, 193–215.
- Delis, D.C., Kramer, J.H., Kaplan, E., Ober, B.A., 1987. *The California Verbal Learning Test*. Psychological Corporation, New York.
- Depp, C.A., Mausbach, B.T., Harmell, A.L., Savla, G.N., Bowie, C.R., Harvey, P.D., Patterson, T.L., 2012. Meta-analysis of the association between cognitive abilities and everyday functioning in bipolar disorder. *Bipolar Disord.* 14 (3), 217–226. <https://doi.org/10.1111/j.1399-5618.2012.01011.x>.
- Domen, A.C., Weijer, S.C.F., Jaspers, M.W., Denys, D., Nieman, D.H., 2019. The validation of a new online cognitive assessment tool: the mycognition quotient. *Int. J. Methods Psychiatr. Res.* 28 (3), 1–10. <https://doi.org/10.1002/mpr.1775>.
- Fett, A.-K.J., Viechtbauer, W., Dominguez, M.G., Penn, D.L., van Os, J., Krabbendam, L., 2011. The relationship between neurocognition and social cognition with functional outcomes in schizophrenia: a meta-analysis. *Neurosci. Biobehav. Rev.* 35 (3), 573–588. <https://doi.org/10.1016/j.neubiorev.2010.07.001>.
- First, M.B., Williams, J.B.W., Karg, R.S., Spitzer, R.L., 2015. *Structured Clinical Interview for DSM-5—Research Version (SCID-5 for DSM-5, Research Version; SCID-5-RV)*. American Psychiatric Association, Arlington VA.
- Gold, J.M., 1997. Auditory working memory and Wisconsin card sorting test performance in schizophrenia. *Arch. Gen. Psychiatry* 54 (2), 159–165. <https://doi.org/10.1001/archpsyc.1997.01830140071013>.
- Gold, J.M., 2004. Cognitive deficits as treatment targets in schizophrenia. *Schizophr. Res.* 72 (1), 21–28. <https://doi.org/10.1016/j.schres.2004.09.008>.
- Green, M.F., Kern, R.S., Braff, D.L., Mintz, J., 2000. Neurocognitive deficits and functional outcome in schizophrenia: are we measuring the "right stuff"? *Schizophr. Bull.* 26 (1), 119–136. <https://doi.org/10.1093/oxfordjournals.schbul.a033430>.
- Harvey, P.D., Silverman, J.M., Mohs, R.C., et al., 1999. Cognitive decline in late-life schizophrenia: a longitudinal study of geriatric chronically hospitalized patients. *Biol. Psychiatry* 45, 32–40.
- Harvey, P.D., Koren, D., Reichenberg, A., Bowie, C.R., 2006. Negative symptoms and cognitive deficits: What is the nature of their relationship? *Schizophrenia Bulletin* 32 (2), 250–258. <https://doi.org/10.1093/schbul/sb011>.
- Harvey, P.D., Deckler, E., Jones, M.T., Jarskog, L.F., Penn, D.L., Pinkham, A.E., 2019. Autism symptoms, depression, and active social avoidance in schizophrenia: association with self-reports and informant assessments of everyday functioning. *J. Psychiatr. Res.* 115, 36–42. <https://doi.org/10.1016/j.jpsychires.2019.05.010>.
- Healey, K.M., Combs, D.R., Gibson, C.M., Keefe, R.S., Roberts, D.L., Penn, D.L., 2015. Observable social cognition—A rating scale: an interview-based assessment for schizophrenia. *Cognitive Neuropsychiatry* 20 (3), 198–221. <https://doi.org/10.1080/13546805.2014.999915>.
- Jagtap, S., Dawson, D.R., Vandermorris, S., et al., 2021. Known-groups and convergent validity of the telephone rey auditory verbal learning test total learning scores for distinguishing between older adults with amnesic cognitive impairment and subjective cognitive decline. *Arch. Clin. Neuropsychol.* 36, 626–631. <https://doi.org/10.1093/arclin/aca085>.
- Kahn, R.S., Keefe, R.S., 2013. Schizophrenia is a cognitive illness: time for a change in focus. *JAMA Psychiatry* 70 (10), 1107–1112. <https://doi.org/10.1001/jamapsychiatry.2013.155>.
- Kay, S.R., Fiszbein, A., Opler, L.A., 1987. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr. Bull.* 13 (2), 261–276. <https://doi.org/10.1093/schbul/13.2.261>.
- Keefe, R.S., Bilder, R.M., Harvey, P.D., et al., 2006. Baseline neurocognitive deficits in the CATIE schizophrenia trial. *Neuropsychopharmacology* 31, 2033–2046. <https://doi.org/10.1038/sj.npp.1301072>.
- Kern, R.S., Gold, J.M., Dickinson, D., et al., 2011. The MCCB impairment profile for schizophrenia outpatients: results from the MATRICS psychometric and standardization study. *Schizophr. Res.* 126, 124–131. <https://doi.org/10.1016/j.schres.2010.11.008>.
- Krabbendam, L., Arts, B., van Os, J., Aleman, A., 2005. Cognitive functioning in patients with schizophrenia and bipolar disorder: a quantitative review. *Schizophr. Res.* 80, 137–149.
- Lachman, M.E., Agrigoroaei, S., Tun, P.A., Weaver, S.L., 2014. Monitoring cognitive functioning: psychometric properties of the brief test of adult cognition by telephone. *Assessment* 21 (4), 404–417. <https://doi.org/10.1177/1073191113508807>.
- Lynham, A.J., Hubbard, L., Tansey, K.E., et al., 2018. Examining cognition across the bipolar/schizophrenia diagnostic spectrum. *J. Psychiatry Neurosci.* 43 (4), 245–253. <https://doi.org/10.1503/jpn.170076>.
- Lystad, J.U., Falkum, E., Mohn, C., et al., 2014. The MATRICS consensus cognitive battery (MCCB): performance and functional correlates. *Psychiatry Res.* 220, 1094–1101. <https://doi.org/10.1016/j.psychres.2014.08.060>.
- Mann-Wrobel, M.C., Carreno, J.T., Dickinson, D., 2011. Meta-analysis of neuropsychological functioning in euthymic bipolar disorder: an update and investigation of moderator variables. *Bipolar Disord.* 13, 334–342.
- Mausbach, B.T., Harvey, P.D., Pulver, A.E., et al., 2010. Relationship of the brief UCSD performance-based skills assessment (UPSA-B) to multiple indicators of functioning in people with schizophrenia and bipolar disorder. *Bipolar Disord.* 12 (1), 45–55. <https://doi.org/10.1111/j.1399-5618.2009.00787.x>.
- Miskowiak, K.W., Jespersen, A.E., Obenhausen, K., et al., 2021. Internet-based cognitive assessment tool: sensitivity and validity of a new online cognition screening tool for patients with bipolar disorder. *J. Affect. Disord.* 289, 125–134. <https://doi.org/10.1016/j.jad.2021.04.039>.
- Montgomery, S.A., Asberg, M., 1979. A new depression scale designed to be sensitive to change. *Br. J. Psychiatry J. Ment. Sci.* 134, 382–389. <https://doi.org/10.1192/bjp.134.4.382>.
- Nuechterlein, K.H., Barch, D.M., Gold, J.M., Goldberg, T.E., Green, M.F., Heaton, R.K., 2004. Identification of separable cognitive factors in schizophrenia. *Schizophr. Res.* 72 (1), 29–39. <https://doi.org/10.1016/j.schres.2004.09.007>.
- Nuechterlein, K.H., Green, M.F., Kern, R.S., et al., 2008. The MATRICS consensus cognitive battery, part 1: test selection, reliability, and validity. *Am. J. Psychiatr.* 165 (2), 203–213. <https://doi.org/10.1176/appi.ajp.2007.07010042>.

- Pinkham, A.E., Harvey, P.D., Penn, D.L., 2018. Social cognition psychometric evaluation: results of the final validation study. *Schizophr. Bull.* 44 (4), 737–748. <https://doi.org/10.1093/schbul/sbx117>.
- Rapp, S.R., Legault, C., Espeland, M.A., et al., 2012. Validation of a cognitive assessment battery administered over the telephone. *J. Am. Geriatr. Soc.* 60, 1616–1623. <https://doi.org/10.1111/j.1532-5415.2012.04111.x>.
- Reichenberg, A., Harvey, P.D., Bowie, C.R., et al., 2008. Neuropsychological function and dysfunction in schizophrenia and psychotic affective disorders. *Schizophr. Bull.* 35 (5), 1022–1029. <https://doi.org/10.1093/schbul/sbn044>.
- Schaefer, J., Giangrande, E., Weinberger, D.R., Dickinson, D., 2013. The global cognitive impairment in schizophrenia: consistent over decades and around the world. *Schizophr. Res.* 150, 42–50. <https://doi.org/10.1016/j.schres.2013.07.009>.
- Schneider, L.C., Streuening, E.L., 1983. SLOF: a behavioral rating scale for assessing the mentally ill. *Soc. Work Res. Abstr.* 19 (3), 9–21. <https://doi.org/10.1093/swra/19.3.9>.
- Sheehan, D.V., Lecrubier, Y., Sheehan, K.H., Amorim, P., Janavs, J., Weiller, E., Dunbar, G.C., 1998. The Mini-international neuropsychiatric interview (MINI): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J. Clin. Psychiatry* 59 (20), 22–33.
- Snelbaker, A.J., Wilkinson, G.S., Robertson, G.J., Glutting, J.J., 2001. Wide range achievement test 3 (wrat3). In: Dorfman, W.I., Hersen, M. (Eds.), *Understanding Psychological Assessment. Perspectives on Individual Differences*. Springer, Boston, MA. [https://doi.org/10.1007/978-1-4615-1185-4\\_13](https://doi.org/10.1007/978-1-4615-1185-4_13).
- Torres, I.J., Boudreau, V.G., Yatham, L.N., 2007. Neuropsychological functioning in euthymic bipolar disorder: a meta-analysis. *Acta Psychiatr. Scand.* 116, 17–26.
- Ventura, J., Reise, S.P., Keefe, R.S., 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 (1–3), 24–31. <https://doi.org/10.1016/j.schres.2010.04.016>.
- Yatham, L.N., Torres, I.J., Malhi, G.S., et al., 2010. The International Society for Bipolar Disorders-Battery for assessment of neurocognition (ISBD-BANC). *Bipolar Disord.* 12 (4), 351–363. <https://doi.org/10.1111/j.1399-5618.2010.00830.x>.
- Young, R.C., Biggs, J.T., Ziegler, V.E., Meyer, D.A., 1978. A rating scale for mania: reliability, validity and sensitivity. *Br. J. Psychiatry J. Ment. Sci.* 133, 429–435. <https://doi.org/10.1192/bjps.133.5.429>.