Title

The characteristics of cognitive neuroscience tests in a schizophrenia cognition clinical trial: Psychometric properties and correlations with standard measures

Authors

Kraus, Michael S
Gold, James M
Barch, Deanna M
et al.

Publication Date

2020-03-01

DOI

10.1016/j.scog.2019.100161

Peer reviewed

Schizophrenia Research: Cognition 19 (2020) 100161

Contents lists available at ScienceDirect

Schizophrenia Research: Cognition

journal homepage: www.elsevier.com/locate/scog

Research Paper

The characteristics of cognitive neuroscience tests in a schizophrenia cognition clinical trial: Psychometric properties and correlations with standard measures

Michael S. Kraus[a,*], James M. Gold[b], Deanna M. Barch[c], Trina M. Walker[a], Charlotte A. Chun[d], Robert W. Buchanan[b], John G. Csernansky[e], Donald C. Goff[f], Michael F. Green[g,h], L. Fredrik Jarskog[i], Daniel C. Javitt[j], David Kimhy[k], Jeffrey A. Lieberman[l], Joseph P. McEvoy[a], Raquelle I. Mesholam-Gately[m], Larry J. Seidman[f,m], M. Patricia Ball[b], Robert S. Kern[g,h], Robert P. McMahon[b], James Robinson[j], Stephen R. Marder[g,h], Richard S.E. Keefe[a]

[a] Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, NC, United States of America
[b] Maryland Psychiatric Research Center, Department of Psychiatry, University of Maryland School of Medicine, Baltimore, MD, United States of America
[c] Department of Psychiatry, Washington University in St. Louis School of Medicine, St. Louis, MO, United States of America
[d] Temple University, Philadelphia, PA, United States of America
[e] Department of Psychiatry, Northwestern Feinberg School of Medicine, Chicago, IL, United States of America
[f] Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School, United States of America
[g] UCLA Semel Institute for Neuroscience and Human Behavior, United States of America
[h] VA VISN 22 Mental Illness Research, Education, and Clinical Center, Los Angeles, CA, United States of America
[i] North Carolina Psychiatric Research Center, Department of Psychiatry, University of North Carolina at Chapel Hill, United States of America
[j] Department of Psychiatry, Nathan Kline Institute for Psychiatric Research, New York University School of Medicine, New York, NY, United States of America
[k] Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, United States of America
[l] Department of Psychiatry, New York State Psychiatric Institute and College of Physicians and Surgeons, Columbia University, United States of America
[m] Department of Psychiatry, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, United States of America

A B S T R A C T

In comparison to batteries of standard neuropsychological tests, cognitive neuroscience tests may offer a more specific assessment of discrete neurobiological processes that may be aberrant in schizophrenia. However, more information regarding psychometric properties and correlations with standard neuropsychological tests and functional measures is warranted to establish their validity as treatment outcome measures. The N-back and AX-Continuous Performance Task (AX-CPT) are two promising cognitive neuroscience tests designed to measure specific components of working memory and contextual processing respectively. In the current study, we report the psychometric properties of multiple outcome measures from these two tests as well as their correlations with standard neuropsychological measures and functional capacity measures. The results suggest that while the AX-CPT and N-back display favorable psychometric properties, they do not exhibit greater sensitivity or specificity with functional measures than standard neurocognitive tests.

1. Introduction

Cognitive impairments are a core feature of schizophrenia and a major determinant of poor functional outcome (Green, 1996; Harvey et al., 1998). Over the past decade interest has grown in treating cognitive impairments through cognitive remediation (Best and Bowie, 2017; Kurtz et al., 2007; McGurk et al., 2007; Medalia et al., 2000; Vinogradov, 2019), aerobic exercise (Firth et al., 2017; Kimhy et al., 2015), and pharmacologic approaches (Davidson and Keefe, 1995; Hyman and Fenton, 2003). As part of the Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) initiative, the MATRICS Consensus Cognitive Battery (MCCB) was developed for assessing cognitive change in clinical treatment studies. In choosing the tests for the MCCB, an emphasis was placed on their psychometric properties, standardization and ease of administration for multi-site clinical trials (Nuechterlein et al., 2008) and thus most of the ten tests eventually chosen were selected from existing standard neuropsychological (NP) tests. However, because of the history of clinical NP test development, with a focus on broad sensitivity to impairment, the standard NP tests chosen for the MCCB are likely to be limited in their sensitivity to specific cognitive functions mediated by discrete neurobiological processes.

* Corresponding author.
E-mail address: michael.kraus@duke.edu (M.S. Kraus).

https://doi.org/10.1016/j.scog.2019.100161
Received 2 November 2018; Received in revised form 24 July 2019; Accepted 25 August 2019
Available online 31 October 2019

Cognitive neuroscience methods with known linkages to specific brain systems provide a logical alternative or supplementary assessment strategy for identifying specific cognitive impairments to be targeted in schizophrenia treatment trials. However, their utility in a multi-site clinical trial remains unproven.

Two promising cognitive neuroscience tasks, the AX-Continuous Performance Task (AX-CPT) and the N-back, were designed to measure specific components of working memory and contextual processing. The AX-CPT measures contextual processing, i.e., the ability to maintain a representation of context in order to mediate an appropriate behavioral response (ServanSchreiber et al., 1996). In the AX-CPT, a series of letters are presented to the subject who is instructed to hit a "target" button only when they see an "X" that was immediately preceded by an "A". A meta-analysis of behavioral studies using the AX-CPT and its nonverbal analog, the Dot Pattern Expectancy Task (DPX), demonstrated a specific deficit in contextual processing in patients with schizophrenia and groups at risk for schizophrenia-spectrum psychopathology (Chun et al., in press). The N-back is a working memory test in which letters are presented sequentially and participants indicate whether the current stimulus matches the letter presented a specified interval back (Cohen et al., 1997). A meta-analysis of functional neuroimaging studies found that during N-back performance, patients with schizophrenia showed both hypo- and hyper-activation in specific regions theoretically implicated in working memory processes (Glahn et al., 2005).

The current study, Treatment Units for Research on Neurocognition and Schizophrenia (TURNS) Merck (MK)-0777, employed the N-back and AX-CPT. These tasks were chosen because a pilot study found beneficial effects of the compound on performance of these two tests (Lewis et al., 2008). Both of these tasks have shown promise in characterizing the cognitive deficits associated with schizophrenia. However, questions remain regarding the utility of these tests in a clinical trial setting. The psychometric characteristics of the N-back have been relatively poorly examined in comparison with those of the AX-CPT, and the psychometric properties of neither test have been characterized in a multi-site treatment study. Additionally only a handful of studies have assessed the relationships between these tasks and aspects of functional outcome such as performance in clinical NP tests and functional capacity measures. Those studies have presented mixed findings: several AX-CPT studies found that overall sensitivity to context (d-prime-context index) predicted performance-based and informant-rated measures of functioning but had no association with clinician-rated or self-reported levels of functioning (Gold et al., 2012; Owoso et al., 2013; Sheffield et al., 2014). The N-back has shown associations with performance-based measures of financial, communication, and total functioning, as well as clinician-rated measures of occupational functioning following intensive cognitive training (Lees et al., 2015; Subramaniam et al., 2014). Overall, however, the relationship of the AX-CPT and N-back with functional capacity remains unclear, and may depend on which functional measures are used. In sum, the literature to date has primarily examined the correlates of these measures in non-treatment studies but their utility in large-scale treatment outcome trials has not yet been firmly established.

Before these tests are likely to be widely adopted as treatment outcome measures, they must demonstrate utility in multi-site treatment trials. In the current manuscript, we analyze data collected during the course of a multi-site treatment study to characterize the psychometric properties of the AX-CPT and N-back as well as the relationship of performance on these two tests to performance on a battery of standard neuropsychological tests (the MCCB) and measures of functional capacity (Schizophrenia Cognition Rating Scale (SCoRS) (Keefe et al., 2006) and UCSD Performance-Based Skills Assessment (UPSA) (Patterson et al., 2001)). We report here that the AX-CPT and N-back did not demonstrate superior psychometric properties or stronger associations with functional capacity in comparison to standard neuropsychological measures.

2. Materials and methods

2.1. Subjects

Six of the seven NIMH TURNS network sites implemented the 4-week, placebo-controlled, parallel group, double-blind study, with the primary purpose of testing the effect of adjunctive MK-0777 administration on cognition in individuals with schizophrenia. Inpatients or outpatients aged 18 to 60, who met DSM-IV-TR criteria for schizophrenia, were selected for study entry. Participants were diagnosed based on information from the Structured Clinical Interview for DSM-IV (First et al., 1997), direct assessment, family informants, and past medical records. Participants were required to be clinically stable, in the non-acute phase of their illness, and to meet the following inclusion criteria (Buchanan et al., 2005): a) treatment with no more than two second generation antipsychotic medications, other than clozapine, for the previous two months, with no dose change in the month prior to study entry; b) Brief Psychiatric Rating Scale (BPRS) (Overall and Gorham, 1962) hallucinatory behavior and unusual thought content item scores ≤ 5; c) BPRS conceptual disorganization item ≤ 4; d) Simpson-Angus Scale (Hawley et al., 2003) total score ≤ 6; and e) Calgary Depression Scale (Addington et al., 1990) total score ≤ 10 (Buchanan et al., 2005; Buchanan et al., 2011b).

Participants were required to: 1) validly complete the MCCB (Kern et al., 2008; Nuechterlein et al., 2008); 2) to score at least one standard deviation (SD) below absolute maximum on one or more of the following tests: Letter-Number Span; Hopkins Verbal Learning Test, and Identical Pairs Continuous Performance Test (CPT-IP) as a means to avoid ceiling effects; and 3) have a Wechsler Test of Adult Reading (Wechsler, 2001) raw score ≥ 6.

Participants were excluded if they had a DSM-IV diagnosis of alcohol or substance abuse (other than nicotine) within the last month, alcohol or substance dependence (other than nicotine) within the last 6 months, or mental retardation; had a history of significant head injury/trauma or clinically significant medical or neurological disease; were treated with drugs known to act at the $GABA_A$ receptor or to inhibit CYP3A4; had a history of severe benzodiazepine withdrawal; participated in a clinical trial of investigational medication within 60 days; or a history of posterior subcapsular cataracts or age-inconsistent nuclear or cortical cataracts, or uveitis. Women of childbearing age were included if using adequate birth control.

The participating institution IRBs approved the study protocol and informed consent procedures. Written informed consent was obtained from all participants after study procedures had been fully explained and prior to study participation. Participant ability to provide valid informed consent was documented using study specific procedures.

2.2. Assessments

Cognition was assessed with both the MCCB (Nuechterlein et al., 2008) and with the SCoRS (Keefe et al., 2006). The UPSA-2 (Patterson et al., 2001), was used to assess functional capacity.

In addition to the MCCB and SCoRS, cognition was assessed by two tests that have been found to activate prefrontal cortical function, the AX-CPT (Cohen et al., 1999) and N-back (Cohen et al., 1997), as a pilot study found that patients receiving MK-077 showed improvement on these measures (Lewis et al., 2008). Of note, this version of the AX-CPT did not include any of the optimizations or administration improvements reported in Henderson et al. (Henderson et al., 2012). Stimuli were white, 60 point Arial letters on a black background. Subjects responded by pressing one of two buttons on a response pad: left button for non-targets and right button for targets. Stimuli were comprised of a series of letters explicitly grouped into cue-probe pairings by underlining of the cue letters. Subjects were instructed to respond to each letter as quickly as possible. Targets were defined as an X that is immediately preceded by an A (AX trials). A letter in any other sequence

was defined as a non-target (AY, BX and BY trials, where "B" represents any non-A cue and "Y" represents any non-X probe). Both cues and probes were presented for 500 ms. The delay between the cue and the probe was 5500 ms, while the inter-trial interval was 1500 ms. The test consisted of four blocks of trials, each composed of: 30 AX trials (78.9%), 3 AY trials (7.9%), 3 BX trials (7.9%) and 2 BY trials (5.3%). This trial type distribution produces a high expectation of an X probe following the presentation of an A cue. Intact use of this expectation is associated with a high rate of AY errors (and slow responding on correct AY trials), while impaired use of the context provided by the A or not-A cue is associated with increased false alarms on BX trials.

The N-back is a working memory test in which memory load is manipulated by changing the size of the interval across blocks. In this study, 0-back, 1-back, and 2-back condition were presented, in that order. In the 0-back condition, the subject was asked to press the "target" button upon the presentation of a pre-specified target letter ("Z"). In the 1-back condition, the subject was asked to press the "target" button to the second of two consecutive, identical letters. In the 2-back condition, the subject was asked to press the "target" button if a letter was identical to that presented two trials earlier. In all conditions, letters were presented for 500 ms with a 2000 ms delay between the end of one letter presentation and the beginning of the next. In addition to target trials (33.3% of trials), non-target trials consisted of repeated distractor trials (16.65% of trials) in which the current letter was a repeat of a letter near the target interval (e.g., 1 or 3 letters back during the 2-back condition) and novel trials (50% of trials) in which letters were not repeated within the "repeated distractor" window.

2.3. Study design

Participants who met inclusion criteria entered a 2-week placebo lead-in Evaluation Phase during which they underwent baseline cognitive, functional capacity, symptom and safety assessments. Participants who continued to meet inclusion criteria entered the 4-week, double-blind Treatment Phase and were randomized to MK-0777 3 mg BID; MK-0777 8 mg BID; or placebo BID. At week 4 of the Treatment Phase, patients received repeat administration of cognitive, functional capacity, symptom and safety assessments. The treatment data have been presented in a previous manuscript (Buchanan et al., 2011a).

2.4. Data quality review

Throughout the trial, all neurocognitive data were reviewed to ensure adherence to testing protocols and to assess test validity. This process was particularly relevant to the AX-CPT and N-back tests as several unique issues, including failure to understand the task instructions, inattentiveness to the test and equipment problems, could impinge upon the collection of valid data in these tests. For the AX-CPT and N-back, final decisions regarding data validity were agreed upon by a core of neuropsychologists (Deanna Barch, Jim Gold and Mike Kraus) most experienced with these tests. All decisions were made blind to subject group and time point. Although decisions were made based on the whole picture the data presented and considering notes left by the tester regarding the sessions in question, the group agreed on some general guidelines. AX-CPT tests were flagged for review if any of the following were true: the participant failed to respond to 20% or more of trials, missed all BX trials, missed > 25% of BY trials or failed to score significantly above chance on AX trials. N-back tests were flagged for review if any of the following were true: the participant failed to respond to 20% or more of trials, the subject missed all repeat distractor trials or answered very few target trials correctly (this last criterion varied across conditions). Because the N-back was divided into 0, 1 and 2-back blocks, decisions on test validity were made on a block-by-block basis.

2.5. Statistical analyses

For the AX-CPT and N-back tests, the loglinear method was used to calculate d-prime (Stanislaw and Todorov, 1999). For the AX-CPT, only BX trials were used to calculate the false alarm rate, thus establishing a "d-prime-context" score (Cohen et al., 1999). Although we designed the AX-CPT with a low percentage of BX trials to produce a high expectation of AX target trials, we had sufficient variance on BX trials to produce a valid d-prime measure. For the N-back, trials with novel and repeated distractors were pooled in calculating the false alarm rate.

As previously reported (Buchanan et al., 2011b), treatment in this study had no significant effect beyond placebo on any measure, and all treatment groups had very similar mean scores and SDs (SD). Therefore, all analyses were completed from data pooled across all three treatment groups. The test-retest reliabilities of the AX-CPT and N-back outcome variables were calculated as intra-class correlation coefficients using the REML method with SAS PROC VARCOMP. Week, treatment, and treatment x week were treated as fixed effects in the variance component model, to remove systematic practice or drug effects from estimates of random within-subject error. Learning effects were investigated by calculation of Cohen's d (mean change/SD change) and the statistical significance of these effects was assessed using paired t-tests. Floor and ceiling effects were calculated as the percentage of subjects within one SD (estimated as the square root of the sum of between- and within-subjects variance components using SAS PROC Varcomp) of worst possible performance and perfect performance respectively. Similarly, we also calculated the percentage of subjects "near chance" as those within one SD from chance performance. Pearson correlation coefficients were calculated to investigate the relationships at baseline across all subjects of AX-CPT and N-back outcomes with standard neurocognitive measures, an interview-based measure of cognition, functional capacity and symptom measures.

3. Results

3.1. Subject numbers, demographics and missing data rates

54 subjects completed the overall study. However, 2 of these did not complete the AX-CPT or N-back at either time point. Thus the subject pool for this study consisted of 52 patients who completed AX-CPT and N-back. Baseline characteristics of this sample are presented in Table 1. Of the 104 AX-CPT tests across both time points, 6 (5.77%) were deemed invalid. Of the 104 N-back tests, 1 (0.96%) was judged to have invalid 1-back conditions and 1 (0.96%) was judged to have invalid 2-back conditions.

3.2. Psychometric properties

Of all the AX-CPT variables, the highest reliability was found for d-prime (0.72), while the ICCs of individual trial types were lower (0.37 to 0.64) (see Table 2). N-back d-prime reliability was considerably better for 1-back (0.66) and 2-back (0.62) than for 0-back (0.40). Learning effects were investigated by calculating Cohen's d. Of all the AX-CPT and N-back variables, only the 2-back d-prime exhibited a significant learning effect (Cohen's $d = 0.55$, $p = .05$). Of the 2-back reaction time measures, target hit and repeat correct rejection reaction times demonstrated similar reliability to the 2-back d-prime measure, while the novel correct rejection reaction time reliability was very low. None of the 2-back reaction time measures exhibited a learning effect.

Because no effect of treatment was found, the analysis of floor and ceiling effects was completed in the entire sample. These measures were calculated using all completed tests regardless of their judged validity, thus allowing an accurate assessment of patients scoring near floor and chance levels. The AX-CPT exhibited a substantial ceiling effect at both baseline (23.3% of d-prime-context scores were within 1 SD of the maximum possible score) and four weeks (26.9%) (Table 3).

Table 1
Baseline characteristics of study participants.

Characteristic	%
Female gender	32.8
White	46.6
Black	39.7
Other	13.8
Hispanic	15.5
Married	10.3
Divorced/separated	22.4
Never married	67.2
Antipsychotics	
Aripiprazole	31.5
Olanzapine	22.2
Quetiapine	11.1
Risperidone/Paliperidone	33.3
Ziprasidone	3.7
Polypharmacy[a]	3.7

Characteristic	Mean	S.D.
Age, years	42.6	9.5
Education, years	13.3	2.8
WTAR score	29.5	11.9
BPRS total score	27.6	6.0
BPRS psychosis	6.9	3.1
SANS total score	18.7	12.5
CDRS total score	1.5	2.0
UPSA summary score	91.6	15.3
ScoRS interviewer global rating	4.4	2.3

[a] One patient each treated with aripiprazole plus olanzapine or aripiprazole plus ziprasidone.

Table 2
AX-CPT and N-back reliability and learning effects.

Measure	ICC	Change over 4 weeks				
		Mean	S.D.	d	t	*P*-value
AX-CPT						
D-prime	0.72	−0.03	1.06	−0.03	−0.11	0.92
AX hits (%)	0.37	3.47	15.15	0.26	0.89	0.39
AY correct rejections (%)	0.50	−1.80	8.43	−0.21	−0.83	0.42
BX correct rejections (%)	0.64	−4.93	26.29	−0.19	−0.73	0.48
N-back d-prime						
0-Back (all distractors)	0.40	−0.04	0.49	−0.08	−0.32	0.76
1-Back (all distractors)	0.66	0.03	0.53	0.06	0.26	0.80
2-Back (all distractors)	0.62	0.23	0.45	0.55	2.15	0.05
2-Back correct response times (RT)						
Target hit RT	0.61	41.22	183.39	0.24	0.95	0.35
Novel correct rejection RT	0.04	−135.94	440.80	−0.31	−1.31	0.21
Repeated correct rejection RT	0.55	−20.39	228.22	−0.09	−0.38	0.71

Table 3
Floor and ceiling effects for AX-CPT and N-back measures.

Measure	Week	% Near floor	% Near chance	% Near ceiling
AX-CPT d-prime	0	0.0	23.3	23.3
	4	0.0	30.8	26.9
N-back d-prime				
0-back (all distractors)	0	0.0	0.0	47.5
	4	0.0	0.0	50.9
1-back (all distractors)	0	0.0	1.6	13.1
	4	0.0	0.0	13.2
2-back (all distractors)	0	0.0	16.4	0.0
	4	0.0	9.4	0.0

Additionally, a fairly large percentage of subjects scored near chance at both baseline (23.3%) and four weeks (30.8%). Ceiling effects for the N-back varied across conditions, as expected, with a large percentage of subjects within 1 SD of ceiling on the 0-back (47.5% at baseline, 50.9% at 4 weeks), a small percentage on the 1-back (13.1% at baseline, 13.2% at 4 weeks) and none on the 2-back. Subjects only performed within 1 SD of chance on the 1-back at baseline (1.6%) and the 2-back (16.4% at baseline, 9.4% at 4 weeks). No floor effects were found for either task.

Reliability was stronger for AX-CPT critical (BX) trials and overall sensitivity to context than for other trials. N-back trials requiring working memory (1-back and 2-back) showed greater reliability than those primarily tapping sustained attention (0-back); however, none of these d-prime variables achieved the minimum level of reliability of the CNTRICS criteria (ICC = 0.70) (Barch et al., 2008). Learning effects were minimal.

3.3. Relationships to other cognitive domains

For the AX-CPT, d-prime-context showed significant positive correlations with all cognitive domain and test scores from the MCCB (Table 4). The strongest correlations were with the domains of processing speed (0.57), visual memory (0.50) working memory (0.44) and attention/vigilance (0.42). AX and BX accuracy showed similar—but slightly weaker—correlations with MCCB variables.

Of the N-back variables, the 2-back d-prime correlated most strongly and consistently with cognitive outcomes (Table 4), only failing to be statistically significant for social cognition. The 1-back d-prime also correlated with most cognitive measures, only failing to be statistically significant for T-scores in the domains of social cognition and reasoning/problem solving and the test of category fluency. The 0-back d-prime was less strongly correlated with cognitive variables, and was only statistically significant for verbal memory and visual memory. 2-back reaction time variables (Supplementary Table 1) exhibited far fewer significant correlations with cognitive variables than did 2-back d-prime. In sum, the AX-CPT and working memory trials of the N-back were associated with most cognitive variables on the MCCB, thereby demonstrating convergent validity.

3.4. Relationship to UPSA and SCoRS

The AX-CPT d-prime-context showed significant positive correlations with UPSA summary and subscale scores, but was not correlated with the SCoRS global interviewer score (Table 5). Of the N-back variables, 2-back d-prime was most strongly correlated with UPSA variables, with all but Comprehension/Planning being statistically significant (Table 5). By means of comparison, the CPT-IP from the MCCB only correlated significantly with the Comprehension/Planning ($r = 0.31$, $p < .05$) and Financial ($r = 0.32$, $p < .05$) domains of the UPSA. 2-back reaction times were not as consistently correlated with UPSA scores as was 2-back d-prime. The 1-back d-prime was positively correlated with the UPSA summary score, but none of the correlations with the UPSA subscales reached significance. The 0-back was not significantly correlated with any functional outcomes. No N-back conditions were correlated with SCoRS. In sum, the AX-CPT and working memory trials of the N-back were associated with overall functioning on the UPSA but neither set of tasks was related to functioning on the SCoRS.

4. Discussion

The cognitive neuroscience measures investigated in this study are more time-consuming and burdensome than standard neuropsychological tests. To justify their inclusion in clinical trials, cognitive neuroscience tests must display better psychometric properties than standard neuropsychological tests, or more specificity with the effects of cognitive treatment or with functional outcomes. We have not seen

Table 4
Baseline correlations of AX-CPT and N-back outcomes with MCCB.

MCCB Domain Score or Test	AX-CPT				N-back		
	d-Prime	AX Hits	AY reject	BX reject	0-Back (N = 52)	1-Back (N = 52)	2-Back (N = 51)
Attention/vigiliance (CPT-IP)	**0.42****	0.26	0.20	**0.37****	0.22	**0.51****	**0.38****
Processing speed	**0.57****	**0.54****	0.06	**0.33***	0.25	**0.52****	**0.60****
BACS symbol coding	**0.55****	**0.50****	−0.11	**0.32***	0.25	**0.58****	**0.57****
Category fluency	**0.42****	**0.40****	0.19	0.25	0.25	0.24	**0.37****
Trails A	**0.39****	**0.40****	0.10	0.23	0.12	**0.40****	**0.48****
Reasoning/problem solving (NAB mazes)	0.27	0.28	−0.05	0.18	0.07	0.27	**0.29***
Social cognition (MSCEIT managing emotions)	**0.32***	0.27	−0.15	0.15	0.10	0.11	0.13
Verbal memory (Hopkins verbal learning test-R)	**0.35***	0.26	0.10	0.15	**0.33***	**0.38****	**0.61****
Visual memory (Brief visuospatial memory test)	**0.50****	**0.32***	0.03	**0.33***	**0.38****	**0.46****	**0.39****
Working memory	**0.44****	**0.40****	0.03	**0.29***	0.13	**0.45****	**0.67****
Letter number sequencing	**0.44****	**0.47****	0.04	0.24	0.07	**0.34***	**0.57****
WMS III spatial span	**0.31***	0.21	0.02	0.27	0.17	**0.43****	**0.56****
Overall composite	**0.61****	**0.49****	0.09	**0.39****	**0.31***	**0.55****	**0.60****

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

evidence for either of these criteria with the AX-CPT and N-back tests employed in the current study.

Although the current study does not support the widespread adoption of the AX-CPT or N-back in clinical trials, it does allow some conclusions regarding which of the many outcome variables produced by these tests are most promising. Identification of the most promising variables for use in future clinical studies may help reduce the need to correct for multiple comparisons and thereby reduce the likelihood of Type II errors. Of the three N-back conditions, the 2-back d-prime scores correlated most strongly with cognition and functional capacity. The 2-back d-prime correlated most strongly with the cognitive domain of working memory, consistent with its classification as a working memory task, while the 1-back correlated with attention/vigilance more strongly than did the 2-back. While other studies have used 2-back reaction times as an outcome measure, our results suggest it is less promising than the 2-back d-prime measure as it did not correlate significantly with functional capacity or any neuropsychological domains. Thus, of all the N-back variables investigated in this study, the 2-back d-prime appears to have the best combination of construct validity, favorable psychometric properties and relation to functional capacity.

The psychometric properties of the AX-CPT used in this study largely conform to the criteria of optimal cognitive neuroscience tasks as described for CNTRICS (Barch et al., 2008), though this is not the version developed and optimized as part of the CNTRAC consortium (Henderson et al., 2012). The d-prime-context ICC in the current study just meets the minimum criterion and is consistent with previous reports of approximately 0.70. In contrast to Strauss and colleagues (Strauss et al., 2014), but consistent with Barch and colleagues' (Barch et al., 2003) findings, there was no evidence of an overall practice effect for the AX-CPT in the current study. Single condition variables for the test were likewise free from practice effects but exhibited much lower reliability than d-prime-context, consistent with previous reports.

Perhaps the most concerning psychometric property exhibited by the AX-CPT in this study was the appreciable ceiling effect for d-prime-context scores at both baseline and 4 weeks. The Dot Pattern Expectancy (DPX) variant of the AX-CPT task, developed as part of the CNTRACs consortium (Henderson et al., 2012), has fewer ceiling effects, many fewer subjects failing to complete the task due to clearer instructions, more user friendly administration, and test-retest reliability within the recommended range. Thus, this optimized DPX is likely a better choice for use in clinical trials.

Although the current study addresses the feasibility and psychometric properties of the AX-CPT and N-Back in a multi-site trial, we did not observe an effect of the investigational drug (Buchanan et al., 2011a). Thus we were unable to assess the sensitivity of these measures to cognitive changes associated with effective drug treatment. Additionally, the current study included assessments of functional capacity, but not direct measures of occupational or social functioning. It will be important that future studies investigate the sensitivity of these and other cognitive neuroscience based measures to cognitive change and their relationship to functional outcomes.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scog.2019.100161.

Table 5
Baseline correlations of AX-CPT and N-back outcomes with SCoRS and UPSA.

	AX-CPT d-prime N = 46	0-Back d-prime N = 52	1-Back d-prime N = 52	2-Back d-prime N = 51	2-Back target RT	2-Back novel distractor RT	2-Back repeat distractor RT
SCRS Int Global	0.04	0.09	−0.05	−0.17	0.00	0.00	−0.11
UPSA summary	**0.57****	0.08	**0.29***	**0.56****	0.21	0.05	**0.42****
Comp/planning	**0.34***	0.03	0.05	0.27	−0.04	0.12	0.20
Financial	**0.40****	0.11	0.35	**0.45****	0.11	−0.08	**0.43****
Communication	**0.54****	0.19	0.31	**0.52****	0.06	0.13	**0.34***
Transportation	**0.41****	−0.02	0.22	**0.32***	0.11	−0.12	0.17
Household	**0.35***	0.07	0.11	**0.35***	**0.32***	0.01	**0.39****
Medication	**0.37****	−0.06	0.26	**0.48****	**0.29***	0.14	0.24

RT = response time, SCoRS Global = SCoRS interviewer global score, UPSA Summary = UPSA total summary score, Orgplan = UPSA organization and planning subscale, Financial = UPSA financial skills subscale, Communication = UPSA communication subscale, Tranportation = UPSA transportation subscale, Household = UPSA household management subscale, Medication = UPSA medication management subscale.
* p < .05.
** p < .01.

Role of funding source

This study was funded by National Institute of Mental Health Contract HHSN278200441003C to the University of California, Los Angeles (SRM, Principal Investigator). Double-blind medications were provided by Merck and Company.

Declaration of competing interest

Dr. Buchanan in the past 3 years has served on advisory boards for Avanir, Astellas Pharma, ITI, Inc., and Roche and has served as a consultant for Takeda, and Upsher-Smith Laboratories, Inc.

Dr. Csernansky in the last 3 years has served as a consultant for Indivior Pharmaceuticals and Aptinyx, Inc.

Dr. Goff has received research support from Avanir Pharmaceuticals.

Dr. Green has been a consultant to AiCure, Lundbeck, and Takeda, and he is on the scientific board ofCadent. He has received research funds from FORUM.

In the past 3 years, Dr. Jarskog has received research grant funding from Auspex/Teva, Boehringer-Ingelheim and Otsuka.

In the past 3 years, Dr. Kimhy has served as a consultant for Neurocog Trials.

Dr. Lieberman neither accepts nor receives any personal financial remuneration for consulting, speaking or research activities from any pharmaceutical, biotechnology or medical device companies. He receives support administered through Columbia University and the Research Foundation for Mental Hygiene in the form of funding and medication supplies for investigator initiated research from Denovo, Taisho, Pfizer, Sunovion, and Genentech, and for company sponsored phase II, III and IV studies from Alkermes, Allergan and Boehringer Ingelheim, and is a consultant to or member of the advisory board of Intracellular Therapies, Lilly, Pierre Fabre, Pear Therapeutics and Psychogenics for which he receives no remuneration. He is a paid consultant for Clintara, a clinical research services organization, and holds a patent from Repligen that yields no royalties.

Dr. McEvoy has received research grant funding from Alkermes, BoehringerIngelheim, Neurocrine, Takeda, and TEVA and has given disease state talks for Neurocrine.

Dr. Marder has served as a consultant for Abbvie, Roche, Otsuka, Teva, Neurocrine, Newron, and Lundbeck. Dr. Marder has received research support from Neurocrine, Takeda, and Boeringer-Ingelheim.

Dr. Richard Keefe currently or in the past 3 years has received honoraria, served as a consultant, speaker, or advisory board member for Abbvie, Acadia, Aeglea, Akebia, Akili, Alkermes, Allergan, ArmaGen, Astellas, Avanir, AviNeuro/ChemRar, Axovant, Biogen, Boehringer-Ingelheim, Cerecor, CoMentis, Critical Path Institute, FORUM, Gammon Howard & Zeszotarski, Global Medical Education (GME), GW Pharmaceuticals, Intracellular Therapeutics, Janssen, Kempharm, Lundbeck, Lysogene, MedScape, Mentis Cura, Merck, Merrakris Therapetics, Minerva Neurosciences Inc., Mitsubishi, Montana State University, Monteris, Moscow Research Institute of Psychiatry, Neuralstem, Neuronix, Novartis, NY State Office of Mental Health, Orygen, Otsuka, Paradigm Testing, Percept Solutions, Pfizer, Pharm-Olam, Regenix Bio, Reviva, Roche, Sangamo, Sanofi, SOBI, Six Degrees Medical, Sunovion, Takeda, Targacept, Teague Rotenstreich Stanaland Fox & Holt, Thrombosis Research Institute, University of Moscow, University of Southern California, University of Texas Southwest Medical Center, WebMD, and Wilson Therapeutics. Dr. Keefe has currently or in the past 3 years received research funding from the National Institute of Mental Health and Boehringer-Ingelheim. Dr. Keefe receives royalties from versions of the BAC testing battery, the MATRICS Battery (BACS Symbol Coding), and the Virtual Reality Functional Capacity Assessment Tool (VRFCAT). He is also a shareholder in NeuroCog Trials, Inc. and Sengenix.

Mr. Kraus, Dr. Gold, Dr. Barch, Ms. Walker, Dr. Chun, Dr. Javitt, Dr. Mesholam-Gately, Dr. Seidman, Ms. Ball, Dr. Kern, Dr. McMahon and Dr. Robinson report no competing interests.

References

Addington, D., Addington, J., Schissel, B., 1990. A depression rating scale for schizophrenics. Schizophr. Res. 3 (4), 247–251. http://www.sciencedirect.com/science/article/pii/092099649090005R (/8//).

Barch, D.M., Carter, C.S., MacDonald III, A.W., Braver, T.S., Cohen, J.D., 2003. Context-processing deficits in schizophrenia: diagnostic specificity, 4-week course, and relationships to clinical symptoms. J. Abnorm. Psychol. 112 (1), 132–143.

Barch, D.M., Carter, C.S., the CNTRICS Executive Committee, 2008. Measurement issues in the use of cognitive neuroscience tasks in drug development for impaired cognition in schizophrenia: a report of the second consensus building conference of the CNTRICS initiative. Schizophr. Bull. 34 (4), 613–618. http://schizophreniabulletin.oxfordjournals.org/cgi/content/abstract/34/4/613. http://schizophreniabulletin.oxfordjournals.org/cgi/reprint/34/4/613.pdf.

Best, M.W., Bowie, C.R., 2017. A review of cognitive remediation approaches for schizophrenia: from top-down to bottom-up, brain training to psychotherapy. Expert. Rev. Neurother. 17 (7), 713–723.

Buchanan, R.W., Davis, M., Goff, D., Green, M.F., Keefe, R.S., Leon, A.C., Nuechterlein, K.H., Laughren, T., Levin, R., Stover, E., Fenton, W., Marder, S.R., 2005. A summary of the FDA-NIMH-MATRICS workshop on clinical trial design for neurocognitive drugs for schizophrenia. Schizophr. Bull. 31 (1), 5–19.

Buchanan, R.W., Keefe, R.S.E., Lieberman, J.A., Barch, D.M., Csernansky, J.G., Goff, D.C., Gold, J.M., Green, M.F., Jarskog, L.F., Javitt, D.C., Kimhy, D., Kraus, M.S., McEvoy, J.P., Mesholam-Gately, R.I., Seidman, L.J., Ball, M.P., McMahon, R.P., Kern, R.S., Robinson, J., Marder, S.R., 2011a. A randomized clinical trial of MK-0777 for the treatment of cognitive impairments in people with schizophrenia. Biol. Psychiatry 69 (5), 442–449. http://www.sciencedirect.com/science/article/pii/S000632231001053X.

Buchanan, R.W., Keefe, R.S.E., Umbricht, D., Green, M.F., Laughren, T., Marder, S.R., 2011b. The FDA-NIMH-MATRICS guidelines for clinical trial design of cognitive-enhancing drugs: what do we know 5 years later? Schizophr. Bull. 37 (6), 1209–1217. http://schizophreniabulletin.oxfordjournals.org/content/37/6/1209.abstract http://schizophreniabulletin.oxfordjournals.org/content/37/6/1209.full.pdf.

Chun, C.A., Ciceron, L., Kwapil, T.R., 2019. A meta-analysis of context integration deficits across the schizotypy spectrum using AX-CPT and DPX tasks. J. Abnorm. Psychol (in press).

Cohen, J.D., Perlstein, W.M., Braver, T.S., Nystrom, L.E., Noll, D.C., Jonides, J., Smith, E.E., 1997. Temporal dynamics of brain activation during a working memory task. Nature 386 (6625), 604–608.

Cohen, J.D., Barch, D.M., Carter, C., Servan-Schreiber, D., 1999. Context-processing deficits in schizophrenia: converging evidence from three theoretically motivated cognitive tasks. J. Abnorm. Psychol. 108 (1), 120–133. http://search.ebscohost.com/login.aspx?direct=true&db=pdh&AN=abn-108-1-120&site=ehost-live&scope=site.

Davidson, M., Keefe, R.S.E., 1995. Cognitive impairment as a target for pharmacological treatment in schizophrenia. Schizophr. Res. 17 (1), 123–129. http://www.sciencedirect.com/science/article/B6TC2-3Y6PFFN-17/2/13e8ff4b80ec00148603c83a30ae16df.

First, M.B., Spitzer, R.L., Gibbon, M., Williams, J., 1997. Structural Clinical Interview for DSM-IV Axis Disorders (SCID-IV). Biometrics Research Department, New York State Psychiatric Institute, New York.

Firth, J., Stubbs, B., Rosenbaum, S., Vancampfort, D., Malchow, B., Schuch, F., Elliott, R., Nuechterlein, K.H., Yung, A.R., 2017. Aerobic exercise improves cognitive functioning in people with schizophrenia: a systematic review and meta-analysis. Schizophr. Bull. 43 (3), 546–556. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5464163/.

Glahn, D.C., Ragland, J.D., Abramoff, A., Barrett, J., Laird, A.R., Bearden, C.E., Velligan, D.I., 2005. Beyond hypofrontality: a quantitative meta-analysis of functional neuroimaging studies of working memory in schizophrenia. Hum. Brain Mapp. 25 (1), 60–69.

Gold, J.M., Barch, D.M., Carter, C.S., Dakin, S., Luck, S.J., MacDonald, A.W., Ragland, J.D., Ranganath, C., Kovacs, I., Silverstein, S.M., Strauss, M., 2012. Clinical, functional, and intertask correlations of measures developed by the cognitive neuroscience test reliability and clinical applications for schizophrenia consortium. Schizophr. Bull. 38 (1), 144–152. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3245578/.

Green, M.F., 1996. What are the functional consequences of neurocognitive deficits in schizophrenia? Am. J. Psychiatry 153 (3), 321–330. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=8610818 http://psychiatryonline.org/data/Journals/AJP/3659/321.pdf.

Harvey, P.D., Howanitz, E., Parrella, M., White, L., Davidson, M., Mohs, R.C., Hoblyn, J., Davis, K.L., 1998. Symptoms, cognitive functioning, and adaptive skills in geriatric patients with lifelong schizophrenia: a comparison across treatment sites. Am. J. Psychiatry 155 (8), 1080–1086. http://ajp.psychiatryonline.org/cgi/content/abstract/155/8/1080.

Hawley, C., Fineberg, N., Roberts, A., Baldwin, D., Sahadevan, A., Sharman, V., 2003. The use of the Simpson Angus scale for the assessment of movement disorder: a training guide. Int. J. Psychiatry Clin. Pract. 7 (4), 349–2257. https://doi.org/10.1080/13651500310002986.

Henderson, D., Poppe, A.B., Barch, D.M., Carter, C.S., Gold, J.M., Ragland, J.D., Silverstein, S.M., Strauss, M.E., MacDonald, A.W., 2012. Optimization of a goal

maintenance task for use in clinical applications. Schizophr. Bull. 38 (1), 104–113. http://schizophreniabulletin.oxfordjournals.org/content/38/1/104.abstract http://schizophreniabulletin.oxfordjournals.org/content/38/1/104.full.pdf.

Hyman, S.E., Fenton, W.S., 2003. Medicine: what are the right targets for psychopharmacology? Science 299 (5605), 350–351. http://www.sciencemag.org.

Keefe, R.S., 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. Psychiatry 163 (3), 426–432. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16513863.

Kern, R.S., Nuechterlein, K.H., Green, M.F., Baade, L.E., Fenton, W.S., Gold, J.M., Keefe, R.S.E., Mesholam-Gately, R., Mintz, J., Seidman, L.J., Stover, E., Marder, S.R., 2008. The MATRICS consensus cognitive battery, part 2: co-norming and standardization. Am. J. Psychiatry 165 (2), 214–220. http://ajp.psychiatryonline.org/cgi/content/abstract/165/2/214.

Kimhy, D., Vakhrusheva, J., Bartels, M.N., Armstrong, H.F., Ballon, J.S., Khan, S., Chang, R.W., Hansen, M.C., Ayanruoh, L., Lister, A., Castrén, E., Smith, E.E., Sloan, R.P., 2015. The impact of aerobic exercise on brain-derived neurotrophic factor and neurocognition in individuals with schizophrenia: a single-blind, randomized clinical trial. Schizophr. Bull. 41 (4), 859–868. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4466187/.

Kurtz, M.M., Seltzer, J.C., Shagan, D.S., Thime, W.R., Wexler, B.E., 2007. Computer-assisted cognitive remediation in schizophrenia: what is the active ingredient? Schizophr. Res. 89 (1–3), 251–260. http://www.sciencedirect.com/science/article/B6TC2-4M6SBJX-2/2/a18db5c92cb27d4ce65f080d49852ce5.

Lees, J., Applegate, E., Emsley, R., Lewis, S., Michalopoulou, P., Collier, T., Lopez-Lopez, C., Kapur, S., Pandina, G.J., Drake, R.J., 2015. Calibration and cross-validation of MCCB and CogState in schizophrenia. Psychopharmacology 232 (21), 3873–3882. https://doi.org/10.1007/s00213-015-3960-8.

Lewis, D.A., Cho, R.Y., Carter, C.S., Eklund, K., Forster, S., Kelly, M.A., Montrose, D., 2008. Subunit-selective modulation of GABA type a receptor neurotransmission and cognition in schizophrenia. Am. J. Psychiatry 165 (12), 1585–1593. http://ajp.psychiatryonline.org/cgi/content/abstract/165/12/1585.

McGurk, S.R., Twamley, E.W., Sitzer, D.I., McHugo, G.J., Mueser, K.T., 2007. A meta-analysis of cognitive remediation in schizophrenia. Am. J. Psychiatry 164 (12), 1791–1802. http://ajp.psychiatryonline.org/cgi/content/abstract/164/12/1791.

Medalia, A., Revheim, N., Casey, M., 2000. Remediation of memory disorders in schizophrenia. Psychol. Med. 30 (6), 1451–1459.

Nuechterlein, K.H., Green, M.F., Kern, R.S., Baade, L.E., Barch, D.M., Cohen, J.D., Essock, S., Fenton, W.S., Frese III, 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. Psychiatry 165 (2), 203–213. http://ajp.psychiatryonline.org/cgi/content/abstract/165/2/203.

Overall, J.E., Gorham, D.R., 1962. The brief psychiatric rating scale. Psychol. Rep. 10(3), 799–812. http://www.amsciepub.com/doi/abs/10.2466/pr0.1962.10.3.799 2011/12/28.

Owoso, A., Carter, C.S., Gold, J.M., MacDonald, A.W., Ragland, J.D., Silverstein, S.M., Strauss, M.E., Barch, D.M., 2013. Cognition in schizophrenia and schizo-affective disorder: impairments that are more similar than different. Psychol. Med. 43 (12), 2535–2545. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4149253/.

Patterson, T.L., Goldman, S., McKibbin, C.L., Hughs, T., Jeste, D.V., 2001. UCSD performance-based skills assessment: development of a new measure of everyday functioning for severely mentally ill adults. Schizophr. Bull. 27 (2), 235–245.

ServanSchreiber, D., Cohen, D.J., Steingard, S., 1996. Schizophrenic deficits in the processing of context - a test of a theoretical model. Arch. Gen. Psychiatry 53 (12), 1105–1112. http://sfx.lib.duke.edu:9003/sfx_local?url_ver=Z39.88-2004&rft.genre=article&rft.issn=0003-990X&rft.date=1996&rft.jtitle=ARCHIVES%20OF%20GENERAL%20PSYCHIATRY&rft.volume=53&rft.issue=12&rft.spage=1105&rft.epage=1112&rft.atitle=Schizophrenic%20deficits%20in%20the%20processing%20of%20context%20-%20A%20test%20of%20a%20theoretical%20model%20&rft.aulast=ServanSchreiber&rft.aufirst=D.

Sheffield, J.M., Gold, J.M., Strauss, M.E., Carter, C.S., MacDonald, A.W., Ragland, J.D., Silverstein, S.M., Barch, D.M., 2014. Common and specific cognitive deficits in schizophrenia: relationships to function. Cogn. Affect. Behav. Neurosci. 14 (1), 161–174. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4083227/.

Stanislaw, H., Todorov, N., 1999. Calculation of signal detection theory measures. Behav. Res. Methods 31 (1), 137–149.

Strauss, M.E., McLouth, C.J., Barch, D.M., Carter, C.S., Gold, J.M., Luck, S.J., MacDonald, A.W., Ragland, J.D., Ranganath, C., Keane, B.P., Silverstein, S.M., 2014. Temporal stability and moderating effects of age and sex on CNTRaCS task performance. Schizophr. Bull. 40 (4), 835–844. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4059430/.

Subramaniam, K., Luks, T.L., Garrett, C., Chung, C., Fisher, M., Nagarajan, S., Vinogradov, S., 2014. Intensive cognitive training in schizophrenia enhances working memory and associated prefrontal cortical efficiency in a manner that drives long-term functional gains. Neuroimage 99, 281–292. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4498800/.

Vinogradov, S., 2019. Has the time come for cognitive remediation in schizophrenia… again? Am. J. Psychiatry 176 (4), 262–264.

Wechsler, D., 2001. Wechsler Test of Adult Reading. Psychological Corporation, San Antonio, Texas.