Development and testing of a web-based battery to remotely assess cognitive health in individuals with schizophrenia

Bruno Biagianti [a],*, Melissa Fisher [b], Benjamin Brandrett [c], Danielle Schlosser [c,d], Rachel Loewy [c], Mor Nahum [e,1], Sophia Vinogradov [b,1]

[a] Posit Science, Inc., United States of America
[b] Department of Psychiatry, University of Minnesota, United States of America
[c] Department of Psychiatry, University of California, San Francisco, United States of America
[d] Verily Life Sciences, United States of America
[e] School of Occupational Therapy, Hebrew University of Jerusalem, Israel

ARTICLE INFO

ABSTRACT

Cognitive impairment in schizophrenia is often severe, enduring, and contributes significantly to chronic disability. A standardized platform for identifying cognitive impairments and measuring treatment effects in cognition is a critical aspect of comprehensive evaluation and treatment for individuals with schizophrenia. In this project, we developed and tested a suite of ten web-based, neuroscience-informed cognitive assessments that are designed to enable the interpretation of specific deficits that could signal that an individual is experiencing cognitive difficulties. The assessment suite assays speed of processing, sustained attention, executive functioning, learning and socio-affective processing in the auditory and visual modalities. We have obtained data from 283 healthy individuals who were recruited online and 104 individuals with schizophrenia who also completed formal neuropsychological testing. Our data show that the assessments 1) are acceptable and tolerable to users, with successful completion in an average of under 40 min; 2) reliably measure the distinct theoretical cognitive constructs they were designed to assess; 3) can discriminate schizophrenia patients from healthy controls with a fair degree of accuracy (AUROC > 0.70); and 4) have promising construct, convergent, and external validity. Further optimization and validation work is in progress to finalize the evaluation process prior to promoting the dissemination of these assessments in real-world settings.

1. Introduction

1.1. Importance of cognition in schizophrenia

Individuals with schizophrenia (SZ) as a group show a wide range of cognitive impairments, especially within the domains of speed of processing, attention, executive functioning, learning and socio-affective processing (Green et al., 2004), with deficits up to 2.5 standard deviations below healthy control subjects (Bratti and Bilder, 2006). There is nonetheless a great deal of inter-individual variability, with some studies showing unimpaired cognitive abilities in some domains (Kremen et al., 2000). When cognitive impairment is defined as a performance deficit of at least one standard deviation below the mean calculated from a healthy control population on one or more areas of cognitive function, the estimated percentage of SZ patients who show cognitive impairment is 55–80% (Kremen et al., 2000; Palmer et al., 1997; Bryson et al., 1993). Importantly, cognitive impairments endure over time, are identifiable early in the course of illness, are present among individuals with prodromal risk syndrome, and contribute to the prediction of conversion to psychosis (Bratti and Bilder, 2006; Cannon et al., 2016).

Studies have shown that these impairments are neither the consequences of positive or negative symptoms, nor related to motivation or global intellectual deficit, nor to anti-psychotic medication. While positive and negative symptoms contribute to morbidity, multiple studies have demonstrated that cognitive impairments in SZ contribute most to chronic disability and unemployment, accounting for 20–60% of the variance in functional outcome of individuals with SZ (Green, 1996).

While the development of new pharmacological, rehabilitative, and psychotherapeutic treatments to enhance cognition has emerged as one of the most pressing challenges in the therapeutics of SZ (Vinogradov, 2017), cognitive profiling in SZ is not routinely implemented in clinical settings, with the result that cognitive deficits largely go unaddressed.

Therefore, there is a growing need for reliable and valid evaluative tools to assess cognition that can be administered and interpreted easily.

1.2. The development of online neurocognitive assessments (ONAs)

Attempts to measure cognition in SZ initially relied on well standardized clinical neuropsychological measures chosen for their history of use in drug development trials for antipsychotics. The MCCB, for example, has been used in the context of many clinical trials, and it has been shown to have good psychometric properties (Nuechterlein et al., 2008; Keefe et al., 2011; Green et al., 2008), but its administration is costly and time-consuming. Since its design, a plethora of brief and reliable instruments for assessing cognitive functioning by non-trained clinicians has been implemented in clinical practice, both as paper and pencil tests (including Cogstate, RBANS, BACS, BCA, and B-CATS (Maruff et al., 2009; Gold et al., 1999; Keefe et al., 2013; Velligan et al., 2004; Hurford et al., 2011)) or as computerized and/or web-based assessments (Silverstein et al., 2007; Elwood, 2001; Levaux et al., 2007; Gordon et al., 2005; Carter et al., 2009). However, the recent explosion of technical advances in the field of human neuroimaging, neurophysiology, and systems neuroscience has led to an increase in knowledge regarding the neural systems that support cognitive processes in human subjects, and the potential pathophysiological processes in SZ (Krystal et al., 2017; Cuthbert and Insel, 2013). It is has now become clear that cognitive processes and the behaviors reflecting those processes are mediated by activity within specific neural circuits, and that drugs and non-pharmacological treatments act at the biological level to change neural circuit functioning (Young and Geyer, 2015). Therefore, in order to measure treatment effects on cognition in SZ, we need to translate measures from basic cognitive neuroscience into behavioral tools that assess the function of specific neural systems that are critically impaired in SZ (Carter and Barch, 2007).

Initiatives like CNTRICS were able to identify measures of cognitive constructs from the field of basic cognitive neuroscience, which led to development, optimization and validation of four behavioral tasks (Gold et al., 2012). While these tasks identify specific SZ-related cognitive impairments, link to neural systems in functional neuroimaging studies, and correlate with outcomes of interest, they still require specific hardware and software. Fundamentally, the challenge lies in coupling neuroscience-informed psychometrically-sound tools with accessible and ubiquitous technologies. Our project, aligned with the CNTRICS theoretical framework, aimed to develop short, online behavioral measures of discrete cognitive processes that individuals with SZ can easily and remotely complete with minimal assistance, in both clinical and non-clinical settings. Importantly, these assessments were designed to enable the interpretation of specific deficits that could be linked to neural systems in SZ, with unique characteristics that hold potential as treatment targets (Young and Geyer, 2015).

The first step in the development of the ONAs was to decide on the critical cognitive domains and processes that are known to be impaired in SZ and in need of targeted treatment. In line with the principles of team science, we integrated theoretical perspectives, technical expertise and empirical knowledge of cognitive neuroscientists working in human and or animal model systems, clinical researchers, and preclinical translational behavioral neuroscientists (Carter et al., 2009). Through a consensus building process, we yielded a scheme of 5 critical domains impaired in SZ: perception, attention, executive functioning, learning, and socio-affective processing (Carter and Barch, 2007). Within each domain, we designated the specific cognitive construct to be considered most relevant to the cognitive impairments of SZ (e.g. in the domain of perception, we chose the construct of speed of processing). We defined "cognitive construct" as a distinct process that could be measured at the behavioral level and for which there existed clearly hypothesized and measureable neural-circuit mechanisms (Moore et al., 2013).

Each of these constructs was then targeted for measurement. For each construct, we identified a cognitive neuroscience paradigm that could selectively and parametrically measure the construct at the behavioral level. For each paradigm, we evaluated promising measurement approaches from the basic cognitive neuroscience literature and developed psychometric tasks in the visual and auditory modality (e.g. within the construct of speed of processing, the paradigm we chose was speeded stimulus discrimination, and an example of the task was frequency discrimination called sounds sweeps). Assessments also needed to meet the following criteria: 1) Be able to be self-administered, 2) Be short, with each assessment < 4 min, 3) Use standardized instructions that are simple and understandable; 4) Include practice trials prior to test trials, 5) Be easily interpretable. Based on these criteria, 10 assessment instruments were developed. A detailed list of these tasks is presented in Supplementary Material.

1.3. Testing and validating online neurocognitive assessments (ONAs)

In this study, the evaluation of ONAs for use in SZ trials was completed thanks to a multistep data-driven process. First, we remotely administered the battery to a convenience sample of healthy controls (HC) and examined if the assessments measured discrete cognitive functions and mapped onto the constructs they were designed to assess. Second, we extended the remote and unsupervised administration of ONAs to a sample of SZ patients, to examine if ONAs captured impairments in SZ and provided information about the structure of cognition in this population. Finally, we conducted convergent and external validity analyses to support using the ONAs in clinical practice and trials of treatments for SZ.

2. Experimental/materials and methods

2.1. Study participants

SZ participants were recruited and studied as part of three independent RCTs of cognitive training: ClinicalTrials.gov NCT01973270 (recent onset schizophrenia, $N = 68$, age 21.9 ± 3.2, study ongoing), NCT01817387 (recent onset schizophrenia, $N = 21$, age 21.3 ± 2.9, study completed and outcome data reported in Schlosser et al. (2018)), and NCT02105779 (chronic schizophrenia, $N = 15$, age 36.7 ± 12.0, study completed and cognitive outcome data reported in Fisher et al. (2017)). Participants with recent onset schizophrenia (RO) were drawn from two research programs at the University of California, San Francisco. Participants with chronic schizophrenia (CSZ) were drawn from research programs at the San Francisco VA Medical Center. Recruitment and informed consent procedures for each site were reviewed and approved by each site's Institutional Review Boards. Healthy volunteers were recruited online: information about the study aims and procedures was posted on the university listservs, Study Finder and Reddit.

2.1.1. Inclusion and exclusion criteria
Criteria for SZ patients were: (1) age 16–65 years, (2) Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, (DSM-IV) diagnosis of schizophrenia, schizophreniform, schizoaffective disorder, or psychosis NOS, (3) good general physical health, (4) fluent in English, (5) no neurological disorder or clinically significant head injury, (6) no current, clinically significant substance abuse that would impede assessment or training, (7) ability to give valid informed consent, (8) no medication changes in the prior month, and (9) stable outpatient status (no hospitalization within the past 3 months). The criteria for healthy control participants included: (1) age 16–60 years, (2) no endorsement of DSM-IV Axis I psychiatric disorders, (3) good general physical health, (4) fluent in English, (5) no neurological disorder or clinically significant head injury, (6) no endorsement of DSM-IV diagnosis of substance dependence, (7) ability to give valid informed consent.

2.2. Procedures

Eligible, consented HC's were sent a link to access the Online Neurocognitive Assessments (ONAs) and completed the tests remotely without supervision. After an intake evaluation that determined study eligibility, SZ participants underwent an in-person structured diagnostic clinical interview and a battery of clinical and neuropsychological tests, as well as the ONAs. SZ participants completed the ONAs prior to starting the cognitive training intervention. Staff aided all participants with accessing the ONAs but did not provide any coaching. All study participants gave written informed consent for the study and were compensated for their participation in all assessments. Payment was contingent on participation and not performance.

2.3. Measures

All neuropsychological assessments and clinical interviews for SZ participants were performed by highly trained raters, directly supervised by the same senior researchers (M.F., D.S, & R.L). The training of clinical and neuropsychological raters has been described in detail in previous publications (Fisher et al., 2015).

2.3.1. Online neurocognitive assessments (ONAs)

The ONAs were developed by Posit Science. Each domain of interest (perception, attention, executive functioning, learning, and socio-affective processing) is assessed in both auditory and visual modalities through distinct tasks. Table 1 summarizes the list of ONAs used in the study, organized according to domain, cognitive construct, and paradigm. A full description of these tasks, along with web-links for researchers interested in using the assessments in their research, are included in the Supplementary Material.

Each assessment takes approximately 3–4 min to complete. Practice trials are presented at the beginning of each task, to help users familiarize with the stimuli and logic of the assessment. In the first practice trial, users are shown the correct answer. They are then asked to complete a few practice trials. If users provide wrong answers during the practice trials (except for the sustained attention tasks, see Supplementary Material), additional trials are presented.

2.3.2. Diagnostic assessment

At study entry, each SZ participant received a standardized diagnostic evaluation performed by research personnel trained in research diagnostic techniques. Evaluations included the Structured Clinical Interview for DSM-IV Axis I Disorders (First et al., 2002), as well as review of clinical records and interview with patient informants (e.g., psychiatrists, therapists, social workers).

2.3.3. Neuropsychological, clinical, and functional outcome assessments

The MATRICS Consensus Cognitive Battery (MCCB) (Nuechterlein et al., 2008) was administered to all SZ participants. In addition, the Hopkins Verbal Learning Test-Revised (HVLT-R) and Brief Visuospatial Memory Test-Revised (BVMT-R) delayed recall trials were administered. All tests were scored and rescored by a second staff member blind to the first scoring. The MCCB computerized scoring program was used to compute age and gender adjusted T-scores and the composite scores for individuals aged 20 or older, and published norms for individuals under 20 (Loewy et al., 2016). T-scores for the HVLT-R and BVMT-R delayed recall trials were computed using normative data from the published manuals. All SZ participants were assessed with the Positive and Negative Symptoms Scale (Kay et al., 1987). Functional outcome was measured in SZ using the modified Global Assessment of Functioning (mGAF) scale (Hall, 1995).

2.4. Data processing and statistical analyses

Prior to analysis, all variables were screened for univariate and multivariate normality. HC and SZ participants' performance on ONAs did not follow a normal distribution (with 8 out of 10 assessments showing a skewed distribution and high kurtosis). There were no significant outliers, and no floor or ceiling effects. SZ participants' performance on MCCB and mGAF followed a normal distribution.

All ONA raw scores from HC and SZ participants were rescaled to T-scores, with a mean of 50 and a standard deviation of 10, to have all variables on the same scale. ONA variables were transformed so that higher = better, and lower = worse, to simplify comparisons across tests.

The following analyses were conducted:

1. Testing our a-priori hypothesis that ONAs measured distinct cognitive processes in HC, we first examined Spearman's rank correlation coefficients between assessments. Second, we used Confirmatory Factor Analysis (CFA) with GEOMIN rotation to compare three nested models of varying complexity: a) a single factor model, where all ONA scores loaded onto a single general cognitive latent factor; b) a data-driven model of five factors allowed to correlate with each other, which included perception, attention, executive functioning, learning, and socio-affective processing; c) a data-driven model of four correlated factors where perception and executive functioning loaded onto the same component.

2. To determine which ONAs are sensitive to the effects of SZ, we included data from HC and SZ participants. First, we investigated whether age, years of education and gender influence performance on the ONAs using non-parametric bivariate correlations and logistic regressions. Next, we estimated Receiver Operating Characteristic (ROC) curves, adjusting for age and gender. ROCs are created by plotting the true positive rate against the false positive rate at various threshold settings – i.e. a measure sensitivity as a function of fall-out. To measure the accuracy of each ONA to correctly classify the group being tested into HC and SZ, we examined the Area Under the Receiver Operating Characteristic Curve (AUROC). The closer AUROC for a model comes to 1, the better the test is at discriminating the two classes. In line with the traditional academic point system (Metz, 1978), we considered fairly accurate tests with AUROCs of 0.70–0.80.

Table 1
The identification of domains, constructs, and paradigms that led to the development and production of ready-to-use online neurocognitive assessments.

Domain	Construct	Paradigm	Task
Perception	Speed of processing	Speeded stimulus discrimination	Sound Sweeps
			Visual Sweeps
Attention	Sustained attention	Test of Variables of Attention (T.O.V.A.)	Sustained Auditory Attention
			Sustain Visual Attention
Executive functioning	Set shifting	Task switching	Auditory Task Switcher
			Visual Task Switcher
Learning	Response learning and bond formation	Paired-associate learning	Auditory Associates
			Visual Associates
Socio-affective processing	Prosody	Prosody detection	Voice Choice
	Emotion processing	Facial emotion recognition	Emotion Motion

3. Next, we tested whether the structure of ONA performance data in SZ matched that of HC. The non-parametric correlation matrix in SZ revealed distributed inter-correlations with high Spearman's rank correlation coefficients. Based on these findings, we first decided to evaluate for statistical fit the best-fitting factor solution from the HC analysis (5-factor solution) in the SZ data, using CFA with GEOMIN rotation. Next, we compared the fit indexes of this model with those of 1) a single factor solution, and 2) a non-nested hierarchical model with five first-order factors that loaded onto a single second-order general factor.

4. As the CFAs identified factors that matched theoretically some of the neuropsychological constructs assessed by the MCCB, in order to assess convergent validity, we measured the strength of the relationship between the ONA scores and MCCB scores with Spearman rank-order correlations.

5. Finally, we conducted regression analyses to study whether the ONAs that are sensitive to the effects of SZ explain a fraction of functional outcome variance independent of that explained by MCCB. Using the Enter method, the ONAs individual scores were entered in the first block, and the MCCB Global Cognition T-score was entered in the second block as a single step.

Exploratory, regression and correlation analyses were conducted in IBM SPSS version 21 (IBM, 2011). All confirmatory analyses were conducted in MPlus version 7 (Muthén and Muthén, 1998-2011). Factor models were compared using the fit indexes that are described in full in the Supplementary Material. Additionally, we directly compared the nested models using the chi-square differences test.

3. Results

Three hundred and ninety-one qualifying participants (104 patients and 283 controls) were included in the study. Table 2 depicts the demographic means and SDs for the entire sample, as well as the means and SDs for the MATRICS and mGAF scores in the SZ sample. SZ participants were significantly younger and had less years of education compared to HCs (all $ps < 0.001$). There was a significantly higher proportion of males in the SZ sample compared to HCs ($\chi^2 = 26.89$, $ps < 0.001$).

The completion of all 10 ONAs required 37 ± 15 min for HC's and 39 ± 18 min for SZ's. The difference between-groups was non-significant

($p = 0.19$). There was variability for completing the battery, with a standard deviation of 17 min and a range of 33 min for the entire sample. SZ participants completed the ONAs without interruption and qualitatively reported good understanding of the instructions. A series of One-Way ANOVAs showed that study site did not significantly affect performance on ONAs, with non-significant between-groups differences (all $ps = 0.11$).

3.1. Do ONAs measure distinct cognitive processes in HC?

There were significant correlations between performances on the auditory and visual assessments of each cognitive construct (e.g., correlation between auditory and visual speed of processing tasks). Spearman's rho coefficients ranged from 0.21 to 0.32, suggesting that, while each pair of tasks tapped into the same cognitive construct, performances in the two sensory modalities only partially overlapped. There were also some significant intercorrelations among performances on tasks that assessed different domains (e.g., Visual Sweeps with Auditory Task Switcher). However, the highest rho coefficient found in the correlation matrix was 0.28, suggesting that ONAs measure distinct cognitive processes.

Results of the CFA on HC data, including fit indices for the competing models, are presented in Table 3. The results of the $\Delta\chi^2$ tests for nested models are presented in Table 4. The single-factor solution, in which all ONAs load onto a single cognitive factor, provided a poor fit for the data (see Table 3, "Single-factor"). The single-factor model was then compared to the four correlated factors solution (attention, learning, socio-affective processing, and executive functioning combined with perception). The four-factor solution yielded a significant increase in model fit over the single-factor solution [$\Delta\chi^2$ (6) = 28.056, $p < 0.001$], but the fit was still relatively poor (see Table 3, "4-factor"). Next, we compared the four-factor model to the five correlated factors solution. The five-factor solution further improved the model fit beyond the four-factor solution [$\Delta\chi^2$ (4) = 13.664, $p < 0.001$] and was a reasonable fit for the data, as the RMSEA is <0.50 and the CFI is >0.90 (see Table 3, "5-factor", and Supplementary Material). Taken together, these analyses on HCs indicate that ONAs show promising psychometric properties and that they reliably measure the five theoretical cognitive constructs they were designed to assess.

3.2. Do ONAs capture SZ-related impairments?

The second goal of the analysis was to test whether the ONAs captured SZ-related impairments. As our two samples were not matched by age, gender, and years of education, we first investigated whether these three demographic variables influenced performance on the ONAs in the combined sample. Logistic regressions indicated that gender influenced performance on some ONAs at trend-level significance, with females performing better at some tasks, and males performing better at others. Non-parametric bivariate correlations showed that age and/or years of education affected performance on the ONAs significantly or at trend-level significance, with better performance associated with younger age or more years of education. However, as education does not have the same implications in healthy subjects and schizophrenia patients, due to the developmental nature of the disorder (Swanson et al., 1998), and because age and years of education where highly correlated (rho = 0.46), we estimated ROC curves for all ONAs only adjusting for age and gender.

Results from the ROC analysis are presented in Table 5 and Fig. 1. After adjusting for age, gender and years of education, six ONAs (*Sound Sweeps, Visual Sweeps, Sustained Visual Attention, Auditory Task Switcher, Auditory Associates and Visual Associates*) showed the ability to distinguish HCs from SZs with a fair degree of accuracy (AUROCs > 0.70). SZ participants performed significantly worse on these tasks compared to HCs (see Fig. 1). *Sustained Auditory Attention, Visual Task Switcher, Emotion Motion and Voice Choice* failed or performed poorly

Table 2
Demographics.

Variable	Healthy control (N = 283)		Schizophrenia patients (N = 104)	
	Mean	SD	Mean	SD
Age (y)[a]	27.7	8.04	24.1	7.7
Gender (% males)[a]	54.60%		84.40%	
Education (y)[a]	15.5	2.6	13	1.8
MCCB Global Cognition T-score	NA		30.0	14.6
Attention T-subscore			36.7	12.3
Executive Fx T-subscore			38.7	10.7
Social cognition T-subscore			42.2	15.2
Speed of processing T-subscore			33.7	14.1
Verbal learning T-subscore			37.31	8.8
Visual learning T-subscore			35.0	13.2
Working memory T-subscore			38.2	12.2
PANSS			61.12	16.36
mGAF	NA		48.8	11.5
Diagnosis			82 schizophrenia	
			16 schizoaffective	
			2 psychosis NOS	
			4 schizophreniform	

MCCB: MATRICS Consensus Cognitive Battery;
mGAF: modified Global Assessment of Functioning scale;
PANSS: Positive and Negative Symptom Scale.
  [a] Significant between-group differences ascertained by Fisher's Exact Test results for age and education, and Chi-Square Test results for gender.

Table 3
Confirmatory factor analysis results in healthy controls.

Model	df (# est. parameters)	χ^2	p-Value	χ^2/df	AIC	BIC	$_{SA}$BIC	CFI	RMSEA [90%CI]
Single-factor	35 (30)	75.118	0.0001	2.15	15,691.830	15,801.194	15,706.063	0.683	0.064 [0.044 0.083]
4-Factor	29 (36)	47.062	0.0183	1.62	15,675.775	15,807.011	15,692.854	0.857	0.047 [0.020 0.071]
5-Factor	**25 (40)**	**33.398**	**0.1214**	**1.34**	**15,670.110**	**15,815.928**	**15,689.087**	**0.934**	**0.034 [0.000 0.062]**

df: degrees of freedom; AIC: Akaike Information; BIC: Bayesian Information Criterion; SABIC: sample size adjusted BIC; CFI: comparative fit index; RMSEA: root mean square error of approximation; CI: confidence interval. Best fitting model is formatted in bold.

at discriminating HCs from SZs (AUROCs 0.5–0.6). The AUROC for the composite score of the 6 ONAs that were sensitive to the effects of SZ was 0.788.

3.3. Do ONAs measure distinct processes in SZ?

The third goal of the analysis was to determine whether the factor model that best fitted the HC data could apply to the SZ data. Accordingly, we first inspected the correlation matrix in SZ, which revealed significant and distributed inter-task correlations with rho coefficients ranging from 0.28 to 0.73. In particular, performances on those 6 ONAs that were sensitive to the effects of SZ were significantly inter-correlated. This seemed to suggest that the overall pattern of group differences found in the ROCs could reflect shared variance across cognitive domains in SZ, with little evidence of domain-specific impairment. Based on these findings, we first decided to evaluate for statistical fit a single factor solution, in which all ONAs load onto a single cognitive factor, using CFA with GEOMIN rotation (Table 6). Next, we compared the fit indexes of this model with those of 1) the 5-factor solution from the HC analysis in the SZ dataset, and 2) a non-nested hierarchical model with five first-order factors that loaded onto a single second-order general cognitive factor.

Results of the CFA on SZ data, including fit indices for the competing models, are presented in Table 6. The single-factor solution provided a poor fit for the data, as the RMSEA was >0. 50 and the CFI was <0.90 (see Table 6, "Single-factor"). The five-factor solution was a very good fit to the data, and yielded a significant increase in model fit over the single-factor solution [$\Delta\chi^2$ (10) = 30.345, p < 0.001]. The hierarchical model was also a very god fit to the data. The smaller values for χ^2/df and the information criterion-based fit indices (i.e. AIC, BIC, SABIC - see Supplementary Material) for the hierarchical model, along with the better fit indices (see Table 6, "Hierarchical"), demonstrate that addition of a general cognitive factor yielded a slightly improved model fit over the five correlated factors solution.

3.4. Convergent and external validity of ONAs

As the five factors identified in SZ and HC by the CFAs represent cognitive domains that are assessed by the MCCB, we measured the strength of the relationship between the ONAs subscores and the MCCB subscores in the SZ sample. We found significant Spearman's rank-order correlations between the two batteries in the medium range for speed of processing (rho = 0.54), attention (rho = 0.57), executive functioning (rho = 0.40), learning (rho = 0.42), but not for emotion recognition (rho = 0.13). For across-domains and within-domain correlations, please see the Supplementary Material. Finally, there was a highly significant correlation between the composite score

of the 6 ONAs that were sensitive to the effects of SZ, and the MCCB Global Cognition T-score (rho = 0.73).

Finally, to investigate external validity, we assessed the relationship of the ONAs that are sensitive to the effects of SZ with functional outcome measures. Spearman's rank-order correlations between scores from these 6 sensitive tests and mGAF showed that the auditory speed of processing, auditory executive functioning, and auditory learning tasks all significantly correlated with mGAF (rhos = 0.26–0.42), whereas the correlations with the three visual ONAs sensitive to the effects of SZ were not significant. All MCCB subscores were significantly correlated with mGAF (rhos 0.24–0.40). In order to determine whether the fraction of mGAF variance explained by the three auditory ONAs was already explained by MCCB scores, we ran a multiple regression, entering auditory ONAs individual subscores and MCCB Global Cognition score in two blocks as separate predictors. This regression accounted for ~38% of the variance (adjusted $R^2 = 0.376$). The three auditory ONAs sensitive to SZ explained 23,6% of the mGAF variance, and adding the MCCB Global Cognition score increased the R^2 at a trend-level (F change = 3.23, p = 0.08). This suggests that the 3 auditory ONAs sensitive to the effects of SZ are capturing a fraction of the mGAF variance that is not explained by MCCB scores.

4. Discussion

As cognitive interventions become more available for people with SZ, there exists a strong need to reliably assess cognitive abilities in clinical and non-clinical settings via easy-to-use, brief tools that do not require face to face administration or special hardware and software (Berry et al., 2016; Lumsden et al., 2016; Gay et al., 2016). While several neurocognitive computerized batteries are currently available (Silverstein et al., 2007; Elwood, 2001; Levaux et al., 2007; Gordon et al., 2005; Carter et al., 2009), numerous drawbacks speak to their general utility within a real-world clinical setting. Our battery sought to address these shortcomings by successfully providing ONAs that 1) are

Table 4
$\Delta\chi^2$ difference tests showing that the 5-factor model best fits the data in healthy controls.

Model	χ^2	df	Contrast	$\Delta\chi^2$	Δdf	p-Value
Single-factor	75.118	35	–	–	–	–
4-Factor solution	47.062	29	Single- vs 4-factor	28.056	6	<0.001
5-Factor solution	33.398	25	5-Factor vs 4-factor	13.664	4	<0.001

df: degrees of freedom.

Table 5
ROC analyses.

Area under the ROC curve (AUROC)					
Test result variable(s)	Area	Std. error	Asymptotic sig.	Asymptotic 95% Confidence interval	
				Lower bound	Upper bound
Sound Sweeps	**0.702**	0.037	0.000	0.639	0.79
Visual Sweeps	**0.774**	0.031	0.000	0.714	0.834
Sustained Auditory Attention	0.675	0.040	0.000	0.596	0.753
Sustained Visual Attention	**0.750**	0.035	0.000	0.681	0.819
Auditory Task Switcher	**0.748**	0.036	0.000	0.677	0.819
Visual Task Switcher	0.679	0.045	0.000	0.593	0.767
Auditory Associates	**0.700**	0.036	0.000	0.631	0.788
Visual Associates	**0.701**	0.036	0.000	0.630	0.773
Emotion Motion	0.594	0.043	0.038	0.509	0.679
Voice Choice	0.543	0.043	0.028	0.460	0.626
COMPOSITE (6 tasks)	**0.777**	0.031	0.000	0.717	0.837

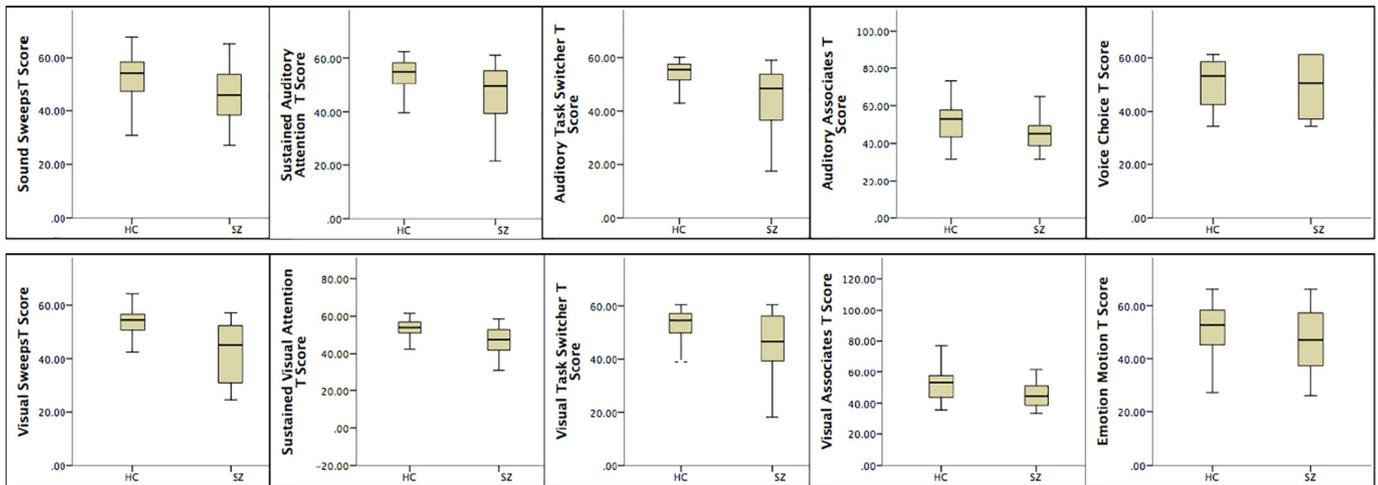Tasks with AUROCs > 0.70 are formatted in bold.

Fig. 1. Box plots of T-scores of ONAs performance for SZ patients (n = 283) and HCs (n = 104).

informed by cognitive affective neuroscience; 2) have good psychometric properties; 3) capture discrete cognitive impairments in SZ; 4) are brief, web-based, and inexpensive; and 5) do not require unique infrastructure, or administration and interpretation by trained staff. In this study, we tested the ONAs in a sample that included 283 healthy controls and 104 patients with SZ successfully, who completed the battery on a web-based platform, without assistance, in an average of under 40 min.

The psychometric analysis of ONAs for use in SZ studies was completed through a multistep data-driven process. First, data from the HCs demonstrated that the magnitude of the intercorrelations among performance in the ONAs is lower than that observed with more standard clinical neuropsychological measures (Dickinson and Gold, 2007), indicating that the measures are capturing unique constructs. Additionally, results from CFA indicated that ONAs reliably measure the five theoretical cognitive constructs they were designed to assess. Further, to the degree that the ONAs demand or share some more general features/processes (such as the ability to sustain performance over time, comprehend instructions, engage with challenging tasks, etc.), our findings clearly indicate that such shared processes are not a major contributor to variance in the observed individual differences in performance in HC, which ultimately suggests that ONAs measure distinct cognitive processes.

Subsequent sensitivity analyses revealed that six of the ten ONAs (namely *Sound Sweeps, Visual Sweeps, Sustained Visual Attention, Auditory Task Switcher, Auditory Associates and Visual Associates*) distinguished HCs from SZs with a fair degree of accuracy, with SZ participants performing significantly worse at these tasks compared to HCs. While the tasks found to have low discriminative power may simply not have tapped into the expected cognitive construct, it is also that there results are due to a sampling bias, given that our sample included SZ participants who were less symptomatic, higher functioning, and less likely to have a dual-diagnosis than some other study populations reported in the literature. Finally, although we controlled for the between-group differences between HC and SZ in age and gender during analyses, more conclusive results could be obtained if participants were matched.

Another step of the psychometric evaluation was to determine whether the factor model that best represented the patter of cognition in HCs could apply to the SZ sample. The number of separable cognitive dimensions in SZ has been long debated, and competing theoretical models currently fall into three separate approaches: (i) a single-factor model where the factor reflects a generalized deficit in performance across neuropsychological tests that may be attributable to diffuse dysfunction of the central nervous system in schizophrenia (Keefe et al., 2006); (ii) a hierarchical model which consists of multiple cognitive factors arising from a second-order general cognitive factor (Dickinson et al., 2011; Dickinson et al., 2006); and (iii) multi-factorial models of intermediate complexity where the deficit is across multiple highly correlated domains (Arnau and Thompson, 2000; McCleery et al., 2015; Burton et al., 2013). In our analyses, the hierarchical model (5 cognitive factors arising from a single, second-order cognitive factor) fitted the SZ data better than the single factor model, and better than the HC informed 5-factor model, although the advantage of the hierarchical model over the 5-factor model was slight in terms of fit indices (similarly to (Dickinson et al., 2011; Dickinson et al., 2006)). These results are indicative of a hierarchical model of cognition in SZ, which is consistent with models of cognitive structure in healthy samples (e.g., combinations of a psychometric "g" as well as separable intellectual factors) (Arnau and Thompson, 2000). Thus, SZ patients may differ from the healthy population in having substantially poorer level of cognitive performance, while retaining a very similar cognitive structure. Because these second-level factors are separable to some degree, it is important to assess discrete cognitive dimensions to clarify which domains can be targeted by specific training-based interventions, while keeping into consideration that a central mechanism, such as cognitive control, may serve a rate-limiting function for deficits and potential improvements across these various cognitive domains (Dickinson et al., 2006; Barch and Ceaser, 2012). Of course, the conclusiveness of our SZ findings is limited by the risk of overfitting: although the SZ sample size in this study was reasonably large, there were many estimated parameters, and a larger sample could have yielded a more stable solution with improved fit indices that could have hypothetically resembled that of HCs.

Table 6
Confirmatory factor analysis results in schizophrenia patients.

Model	df (# est. parameters)	χ^2	p-Value	χ^2/df	AIC	BIC	$_{SA}$BIC	CFI	RMSEA [90%CI]
Single-factor	35 (30)	55.554	0.0150	1.59	4947.457	5023.755	4929.045	0.860	0.079 [0.035 0.117]
5-Factor solution	25 (40)	25.209	0.4507	1.01	4937.111	5038.84	4912.563	0.999	0.009 [0.000 0.083]
Hierarchical 5-Factors + underlying factor	**30 (35)**	**26.648**	**0.6417**	**0.089**	**4928.550**	**5017.565**	**4907.070**	**1.000**	**0.000 [0.000 0.066]**

df: degrees of freedom; AIC: Akaike Information; BIC: Bayesian Information Criterion; SABIC: sample size adjusted BIC; CFI: comparative fit index; RMSEA: root mean square error of approximation; CI: confidence interval. Best fitting model is formatted in bold.

Convergent and external validity constituted two additional steps of our validation process. We first examined whether ONAs engaged the same neurocognitive domains as "gold-standard" assessment batteries, such as the MCCB. We found significant correlations for speed of processing, attention, executive functioning, learning, but not for emotion recognition. While such findings suggest that performance on some of these new tests aligns with previously identified constructs, more development work is needed to improve the assessment of social cognition. Finally, we measured the relationship between ONAs performance in SZ and functional outcome, as measured by mGAF. We found that three auditory assessments had significant – albeit not particularly high – associations with the mGAF and captured a fraction of its variance that was not explained by MCCB scores. Tentatively, the weak association of ONAs performance with functional outcome measures could have arisen for multiple reasons. Firstly, again, this may refer back to our sampling bias. Secondly, this could be the result of ONA measures having lower reliability than standard neuropsychological tests. Therefore, future studies will be necessary in order to more fully compare the reliability of ONAs to measures from neuropsychological batteries. Ultimately, it could be that the relatively modest correlations with functional outcome measures is a real signal, suggesting that the assessment of specific constructs may help determine with enhanced precision what a treatment does at the cognitive level of analysis, but comes at the cost of reduced relationships to clinical outcomes of interest.

In conclusion, we have identified a set of cognitive systems that show impairment in schizophrenia, and have developed a set of tasks adapted from basic cognitive neuroscience that can be used to assess these discrete cognitive processes in this patient population. As many of these cognitive systems also show impairment in other disorders, including ADHD, autism, mild cognitive impairment, and Alzheimer's disease (Solomon et al., 2008; Vaidya and Stollstorff, 2008; Arnáiz and Almkvist, 2003), studies are underway to evaluate the sensitivity of these tests to other clinical populations. Researchers interested in using these assessments and in learning about the most recent developments are encouraged to contact the corresponding author.

Nonetheless, before we recommend them as cognitive neuroscience based measurement tools, more research and development work is necessary to: 1) maximize sensitivity and selectivity in assessing the specific cognitive mechanisms of interest; 2) evaluate test-rest reliability for within-subject measurement of treatment effects; 3) ensure that optimizations designed to enhance the psychometric properties of the tasks do not alter their construct validity; and 4) simplify task administration and minimize task length. Large-scale normative studies that include data-driven revised versions of these ONAs, multiple measures of functional outcome, and more traditional neuropsychological measures, are underway to replicate these initial findings. If successful, this assessment suite has the potential to bridge the translational gap and provide clinicians and patients with a self-guided, cost-effective assessment tool that could inform strategies to radically enhance cognitive health.

Conflict of interest

BB is Senior Scientist at Posit Science, a company that produces cognitive training and assessment software. MN is a former Posit Science scientist and is currently co-PI on two NIMH-funded SBIR grants awarded to Posit Science. The other authors report no conflict of interest.

Contributors

BBi and MN guided the development of the online neurocognitive assessments. BBi, MN, and SV designed the study and wrote the protocol. All authors contributed to the implementation of the protocol and in the operationalization of the study. Authors MF, DS, RL provided patient participants. Authors BB and MN managed the literature searches and statistical analyses. Authors BBi, MF, and BBr undertook the statistical analysis, and author BB wrote the first draft of the manuscript. Authors DS, RL, MN and MF and contributed to revise the manuscript. All authors agreed with the final content of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.schres.2019.01.047.

References

Arnáiz, E., Almkvist, O., 2003. Neuropsychological features of mild cognitive impairment and preclinical Alzheimer's disease. Acta Neurol. Scand. Suppl. 179, 34–41.

Arnau, R.C., Thompson, B., 2000. Second order confirmatory factor analysis of the WAIS-III. Assessment 7, 237–246.

Barch, D.M., Ceaser, A., 2012. Cognition in schizophrenia: core psychological and neural mechanisms. Trends Cogn. Sci. 16, 27–34.

Berry, N., Lobban, F., Emsley, R., Bucci, S., 2016. Acceptability of interventions delivered online and through mobile phones for people who experience severe mental health problems: a systematic review. J. Med. Internet Res. 18, e121.

Bratti, I.M., Bilder, R.M., 2006. Neurocognitive deficits and first-episode schizophrenia; characterization and course. In: Sharma, T., Harvey, P.D. (Eds.), The Early Course of Schizophrenia. Oxford University Press, Oxford, UK.

Bryson, G.J., Silverstein, M.L., Nathan, A., Stephen, L., 1993. Differential rate of neuropsychological dysfunction in psychiatric disorders: comparison between the Halstead-Reitan and Luria-Nebraska batteries. Percept. Mot. Skills 76, 305–306.

Burton, C.Z., et al., 2013. Factor structure of the MATRICS consensus cognitive battery (MCCB) in schizophrenia. Schizophr. Res. 146, 244–248.

Cannon, T.D., et al., 2016. An individualized risk calculator for research in prodromal psychosis. Am. J. Psychiatry 173, 980–988.

Carter, C.S., Barch, D.M., 2007. Cognitive neuroscience-based approaches to measuring and improving treatment effects on cognition in schizophrenia: the CNTRICS initiative. Schizophr. Bull. 33, 1131–1137.

Carter, C.S., et al., 2009. CNTRICS final task selection: social cognitive and affective neuroscience-based measures. Schizophr. Bull. 35, 153–162.

Cuthbert, B.N., Insel, T.R., 2013. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. BMC Med. 11.

Dickinson, D., Gold, J.M., 2007. Less unique variance than meets the eye: overlap among traditional neuropsychological dimensions in schizophrenia. Schizophr. Bull. 34, 423–434.

Dickinson, D., Ragland, J., Calkins, M., Gold, J., Gur, R., 2006. A comparison of cognitive structure in schizophrenia patients and healthy controls using confirmatory factor analysis. Schizophr. Res. 85, 20–29.

Dickinson, D., Goldberg, T.E., Gold, J.M., Elvevåg, B., Weinberger, D.R., 2011. Cognitive factor structure and invariance in people with schizophrenia, their unaffected siblings, and controls. Schizophr. Bull. 37, 1157–1167.

Elwood, R.W., 2001. MicroCog: assessment of cognitive functioning. Neuropsychol. Rev. 11, 89–100.

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

Fisher, M., et al., 2015. Neuroplasticity-based auditory training via laptop computer improves cognition in young individuals with recent onset schizophrenia. Schizophr. Bull. 41, 250–258.

Fisher, M., et al., 2017. Supplementing intensive targeted computerized cognitive training with social cognitive exercises for people with schizophrenia: an interim report. Psychiatr. Rehabil. J. 40, 21–32.

Gay, K., Torous, J., Joseph, A., Pandya, A., Duckworth, K., 2016. Digital technology use among individuals with schizophrenia: results of an online survey. JMIR Ment. Health 3, e15.

Gold, J.M., Queern, C., Iannone, V.N., Buchanan, R.W., 1999. Repeatable battery for the assessment of neuropsychological status as a screening test in schizophrenia I: sensitivity, reliability, and validity. Am. J. Psychiatry 156, 1944–1950.

Gold, J.M., et al., 2012. Clinical, functional, and intertask correlations of measures developed by the cognitive neuroscience test reliability and clinical applications for schizophrenia consortium. Schizophr. Bull. 38, 144–152.

Gordon, E., Cooper, N., Rennie, C., Hermens, D., Williams, L.M., 2005. Integrative neuroscience: the role of a standardized database. Clin. EEG Neurosci. 36, 64–75.

Green, M.F., 1996. What are the functional consequences of neurocognitive deficits in schizophrenia? Am. J. Psychiatry 153, 321–330.

Green, M.F., Kern, R.S., Heaton, R.K., 2004. Longitudinal studies of cognition and functional outcome in schizophrenia: implications for MATRICS. Schizophr. Res. 72, 41–51.

Green, M.F., et al., 2008. Functional co-primary measures for clinical trials in schizophrenia: results from the MATRICS psychometric and standardization study. Am. J. Psychiatry 165, 221–228.

Hall, R.C., 1995. Global assessment of functioning. A modified scale. Psychosomatics 36, 267–275.

Hurford, I.M., Marder, S.R., Keefe, R.S.E., Reise, S.P., Bilder, R.M., 2011. A brief cognitive assessment tool for schizophrenia: construction of a tool for clinicians. Schizophr. Bull. 37, 538–545.

IBM Corp., Released 2011. IBM SPSS Statistics for Windows. IBM Corp., Armonk, NY , Version 20.0.

Kay, S.R., Fiszbein, A., Opler, L.A., 1987. The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophr. Bull. 13, 261–276.

Keefe, R.S.E., et al., 2006. Baseline neurocognitive deficits in the CATIE schizophrenia trial. Neuropsychopharmacology 31, 2033–2046.

Keefe, R.S.E., et al., 2011. Characteristics of the MATRICS consensus cognitive battery in a 29-site antipsychotic schizophrenia clinical trial. Schizophr. Res. 125, 161–168.

Keefe, R.S.E., et al., 2013. Clinical trials of potential cognitive-enhancing drugs in schizophrenia: what have we learned so far? Schizophr. Bull. 39, 417–435.

Kremen, W.S., Seidman, L.J., Faraone, S.V., Toomey, R., Tsuang, M.T., 2000. The paradox of normal neuropsychological function in schizophrenia. J. Abnorm. Psychol. 109, 743–752.

Krystal, J.H., et al., 2017. Computational psychiatry and the challenge of schizophrenia. Schizophr. Bull. 43, 473–475.

Levaux, M.-N., et al., 2007. Computerized assessment of cognition in schizophrenia: promises and pitfalls of CANTAB. Eur. Psychiatry 22, 104–115.

Loewy, R., et al., 2016. Intensive auditory cognitive training improves verbal memory in adolescents and young adults at clinical high risk for psychosis. Schizophr. Bull. 42 (Suppl. 1), S118–S126.

Lumsden, J., Edwards, E.A., Lawrence, N.S., Coyle, D., Munafò, M.R., 2016. Gamification of cognitive assessment and cognitive training: a systematic review of applications and efficacy. JMIR Serious Games 4, e11.

Maruff, P., et al., 2009. Validity of the CogState brief battery: relationship to standardized tests and sensitivity to cognitive impairment in mild traumatic brain injury, schizophrenia, and AIDS dementia complex. Arch. Clin. Neuropsychol. 24, 165–178.

McCleery, A., et al., 2015. Latent structure of cognition in schizophrenia: a confirmatory factor analysis of the MATRICS consensus cognitive battery (MCCB). Psychol. Med. 45, 2657–2666.

Metz, C.E., 1978. Basic principles of ROC analysis. Semin. Nucl. Med. 8, 283–298.

Moore, H., Geyer, M.A., Carter, C.S., Barch, D.M., 2013. Harnessing cognitive neuroscience to develop new treatments for improving cognition in schizophrenia: CNTRICS selected cognitive paradigms for animal models. Neurosci. Biobehav. Rev. 37, 2087–2091.

Muthén, L.K., Muthén, B.O., 1998-2011. Mplus User's Guide. Sixth Edition. Muthén & Muthén, Los Angeles, CA.

Nuechterlein, K.H., et al., 2008. The MATRICS consensus cognitive battery, part 1: test selection, reliability, and validity. Am. J. Psychiatry 165, 203–213.

Palmer, B.W., et al., 1997. Is it possible to be schizophrenic yet neuropsychologically normal? Neuropsychology 11, 437–446.

Schlosser, D.A., et al., 2018. Efficacy of PRIME, a mobile app intervention designed to improve motivation in young people with schizophrenia. Schizophr. Bull. 44, 1010–1020.

Silverstein, S.M., et al., 2007. Development and validation of a world-wide-web-based neurocognitive assessment battery: WebNeuro. Behav. Res. Methods 39, 940–949.

Solomon, M., Ozonoff, S.J., Cummings, N., Carter, C.S., 2008. Cognitive control in autism spectrum disorders. Int. J. Dev. Neurosci. 26, 239–247.

Swanson, C.L., Gur, R.C., Bilker, W., Petty, R.G., Gur, R.E., 1998. Premorbid educational attainment in schizophrenia: association with symptoms, functioning, and neurobehavioral measures. Biol. Psychiatry 44, 739–747.

Vaidya, C.J., Stollstorff, M., 2008. Cognitive neuroscience of attention deficit hyperactivity disorder: current status and working hypotheses. Dev. Disabil. Res. Rev. 14, 261–267.

Velligan, D.I., et al., 2004. A brief cognitive assessment for use with schizophrenia patients in community clinics. Schizophr. Res. 71, 273–283.

Vinogradov, S., 2017. The golden age of computational psychiatry is within sight. Nat. Hum. Behav. 1, 0047.

Young, J., Geyer, M., 2015. Developing treatments for cognitive deficits in schizophrenia: the challenge of translation. J. Psychopharmacol. (Oxf.) 29, 178–196.