A comparative study of the MATRICS and IntegNeuro cognitive assessment batteries

Steven M. Silverstein ab, Judith Jaeger c, Anne-Marie Donovan-Lepore d, Sandra M. Wilkniss e, Adam Savitz f, Igor Malinovsky a, Danielle Hawthorne a, Shane Raines f, Sarah Carson a, Stephanie Marcello e, Stephen R. Zukin c, Stephen Furlong c, Gersham Dent c

a University of Medicine and Dentistry of New Jersey-University Behavioral HealthCare, Piscataway, NJ, USA
b University of Medicine and Dentistry of New Jersey-Robert Wood Johnson Medical School, Piscataway, NJ, USA
c AstraZeneca, Wilmington, DE, USA
d Zucker-Hillside Hospital, Queens, NY, USA
e Thresholds, Inc., Chicago, IL, USA
f New York Presbyterian Hospital-Weill Medical College of Cornell University-Payne Whitney Westchester, White Plains, NY, USA

First published on: 07 May 2010
A comparative study of the MATRICS and IntegNeuro cognitive assessment batteries

Steven M. Silverstein,1,2 Judith Jaeger,3 Anne-Marie Donovan-Lepore,4 Sandra M. Wilkniss,5 Adam Savitz,6 Igor Malinovsky,1 Danielle Hawthorne,1 Shane Raines,3 Sarah Carson,1 Stephanie Marcello,1 Stephen R. Zukin,3 Stephen Furlong,3 and Gersham Dent3

1University of Medicine and Dentistry of New Jersey–University Behavioral HealthCare, Piscataway, NJ, USA
2University of Medicine and Dentistry of New Jersey–Robert Wood Johnson Medical School, Piscataway, NJ, USA
3AstraZeneca, Wilmington, DE
4Zucker-Hillside Hospital, Queens, NY, USA
5Thresholds, Inc., Chicago, IL, USA
6New York Presbyterian Hospital–Weill Medical College of Cornell University–Payne Whitney Westchester, White Plains, NY, USA

Cognitive impairment is prevalent in schizophrenia and is related to poorer functional and treatment outcomes. Cognitive assessment is therefore now a routine component of clinical trials of new treatments for schizophrenia. The current gold-standard for cognitive assessment in clinical trials for schizophrenia is the MATRICS (Measurement and Treatment Research to Improve Cognition in Schizophrenia) Consensus Cognitive Battery (MCCB), which was developed based on expert consensus and incorporates paper-and-pencil tests (and one computerized measure) with an established history in the field of neuropsychology. Recently, however, interest has increased in using computerized batteries for clinical trials. In this study, we tested 155 people with schizophrenia and 75 healthy control participants on both the MCCB and IntegNeuro, a touch-screen-based computerized battery with previously demonstrated high levels of reliability and validity, to determine comparability between test scores. In addition, we assessed test–retest reliability and practice effects over a one-month interval for both batteries and determined correlations between cognitive test scores and scores on functional outcome measures. High levels of agreement were observed between total battery composite scores ($r > .80$) and, in a canonical correlation analysis, between all critical single test scores from each battery ($r_c > .90$). The batteries demonstrated essentially equivalent sensitivity in discriminating between patients and controls and equivalent levels of test–retest reliability and practice effects. Correlations between cognitive test scores and functional outcome measures were equivalent between the two batteries and low in nearly all cases. The number of missing data points was greater with IntegNeuro, highlighting the requirements for test administrator involvement even with computerized batteries.

Keywords: Schizophrenia; Cognition; Assessment; Computers; Psychometric properties.

This work was supported by a grant to the first author from AstraZeneca. The following authors conceived and designed the study described in this paper: Steven M. Silverstein, Judith Jaeger, Stephen Zukin, Stephen Furlong, and Gersham Dent. The following authors facilitated data management and data sharing for this study: Igor Malinovsky, Danielle Hawthorne, and Sarah Carson. The following authors served as site principal investigators and coordinated staff training and data collection at their sites: Sandra M. Wilkniss, Adam Savitz, and Anne-Marie Donovan-Lepore. Stephanie Marcello assisted with research assistant training, participant recruitment, and data collection. Shane Raines analyzed the data for this study. S. M. Silverstein served as a paid consultant for the Brain Resource Company on several occasions prior to the initiation of this study. He currently receives funding from the National Institute of Mental Health (NIMH) and AstraZeneca. A. Savitz receives funding from the Stanley Medical Research Institute. S. M. Wilkniss receives funding from NIMH. J. Jaeger, S. Raines, S. R. Zukin, S. Furlong, and G. Dent are employees of AstraZeneca and report no conflicts of interest. A.-M. Donovan-Lepore, S. Marcello, I. Malinovsky, D. Hawthorne, and S. Carson report no biomedical financial interests or conflicts of interest. Address correspondence to Steven M. Silverstein, University of Medicine and Dentistry of New Jersey, 151 Centennial Avenue, Piscataway, NJ, 08854, USA (E-mail: silvers1@umdnj.edu).
Cognitive assessment is important in schizophrenia, given the number of impairments associated with the disorder (e.g., attention, memory, learning, problem solving, etc.; Green et al., 2008) and the associations between these findings and poorer functional outcome (Green, 1996). This is now a priority area for the National Institute of Mental Health (NIMH). For example, two foci of the recent Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) initiative (Green et al., 2004; Nuechterlein et al., 2008) were to identify relevant neurocognitive treatment targets and develop a test battery that assesses them. The MATRICS project built upon the earlier development of standardized batteries for the assessment of neurocognitive functioning in schizophrenia, including the Repeatable Battery for the Neuropsychological Assessment of Schizophrenia (RBANS; Gold, Quern, Iannone, & Buchanan, 1999; Hobart, Goldberg, Bartko, & Gold, 1999) and the Brief Assessment of Cognition in Schizophrenia (BACS; Keefe, 1999; Keefe et al., 2004). However, it differed from past test battery development efforts in that it was based on the input of several stakeholder groups, including academic researchers, people with schizophrenia, the NIMH, the Food and Drug Administration (FDA), and the pharmaceutical industry. The resulting MATRICS Cognitive Consensus Battery (MCCB) is a paper-and-pencil battery with the exception of the computerized Identical Pairs Continuous Performance Test (CPT-IP). Data on reliability and validity of the battery became available in early 2008 (Nuechterlein et al., 2008). A useful feature of the MCCB is that its scores reflect a conorming process on a large (312) and diverse sample (Kern et al., 2008).

A general trend in the assessment field involves the development of computerized neurocognitive test batteries (Baker et al., 1985; Devivo, Rothland, Price, & Fein, 1999; Elwood, 2001; Kane & Kay, 1992; Morris, Evenend, Sahakian, & Robbins, 1987), and interest has increased in using computerized tests for clinical trials of treatments for schizophrenia (e.g., O’Halloran et al., 2008). Recently, a touch-screen-based, computerized cognitive assessment battery, IntegNeuro (Brain Resource, Ltd), was developed as part of the standardized methodology used with the Brain Resource International Database (BRID; Gordon, Cooper, Rennie, Hermens, & Williams, 2005; http://www.brainnet.net/). It was developed with input from a consortium of scientists involved in establishing the database. The goals behind creation of IntegNeuro, and a companion psychophysiological battery called LabNeuro, were to establish standardized, computer-based cognitive and psychophysiological assessment batteries (Paul et al., 2005) and to pool data from the world-wide use of these tools into a single international database (BRID). This database currently contains complete data from over 6,000 people, from age 6–100 years, and includes various diagnostic groups in addition to healthy participants, including schizophrenia, attention-deficit/hyperactivity disorder (ADHD), Alzheimer’s disease, depression, posttraumatic stress disorder (PTSD), and mild cognitive impairment (Gordon et al., 2005). IntegNeuro and LabNeuro are currently being used in clinical trials and in clinical settings throughout the world, with all data collected being included in the BRID.

IntegNeuro consists of an automated stimulus presentation protocol and response recording involving touch-screen and voice-recording software. The tests were based on existing ones known to be sensitive to brain dysfunction. Features of the battery include standardized instructions using both auditory explanations and visual examples, practice trials prior to test trials, and largely automated scoring procedures. It is also independent from mouse or keyboard familiarity. In addition, the battery includes both language and nonlanguage paradigms. Test–retest reliability of IntegNeuro, as assessed in healthy controls, has been shown to be acceptable for all measures in the battery (Williams et al., 2005). Similarly, there is consistency across international testing sites (Paul et al., 2007). Norms have been established for IntegNeuro within the first cohort of 1,000 stringently screened nonpatient participants (Clark et al., 2006). Moreover, IntegNeuro has demonstrated a factor structure similar to the MCCB, and prediction of functional outcome, in a sample of first-episode schizophrenia patients (Williams, Whitford, Flynn, Wong, Liddell, et al. 2008).

The primary aim of this study was to measure the level of agreement between IntegNeuro and the MCCB in a large sample of schizophrenia patients and clinically healthy participants. A secondary aim was to measure, and compare, test–retest reliability of IntegNeuro and MCCB in schizophrenia patients and control participants, as well as practice effects. The proposed study was the first assessment of test–retest reliability of IntegNeuro in schizophrenia and a replication of the Williams et al. (2005) reliability study in healthy participants. In addition to these analyses, relationships between test battery scores and scores on two standardized tests of functional capacity, the Social Functioning Scale (SFS; Birchwood, Smith, Cochrane, Wetton, & Copestake, 1990) and the University of California, San Diego (UCSD) Performance-Based Skills Assessment (UPSA-2; Patterson, Goldman, McKibbin, Hughes, & Jeste, 2001) were determined.

**METHOD**

**Performance sites**

Data were collected at five sites: two community mental health centers that are part of the University of Medicine and Dentistry of New Jersey’s (UMDNJ’s) University Behavioral Health Care (UBHC) system; the New York Presbyterian Hospital–Weill Medical College (WMC) of Cornell University–Westchester Division (White Plains, NY); Zucker-Hillside Hospital (Queens, NY), and Thresholds, Inc (Chicago, IL).

Final recruitment totals for the study are reported in Table 1. Note that when participants dropped out after Visit 3 they were not replaced, and thus the sample size for the test–retest component of the study is smaller than that for the initial IntegNeuro–MCCB comparison.

**Participants**

Participants were 155 people with schizophrenia (102 or 65% male) and 75 healthy (i.e., without a diagnosable
major Axis I psychiatric condition) controls (28 or 37% male). The control group was matched to the patient group based on group means for age and maternal education. All patients were outpatients or partial hospital patients. The schizophrenia group was 37% African-American, 0.6% Asian, 3.2% Hispanic, 2.5% Pacific Islander or Other, and 48.4% White (with 8.3% undetermined). The control group was 34.7% African-American, 2.7% Asian, 7% Hispanic, 1.3% Pacific Islander or Other, and 53.3% White. The control group was matched to the patient group based on group means for age and maternal education, and this was our goal as well. Therefore, relatively asymptomatic patients on stable treatment regimens were studied.

Inclusion criteria were as follows:

1. Between the ages of 18 and 55 years.
2. Clinically stable, not hospitalized, and in the residual (nonacute) phase of their illness for at least 8 weeks.
3. Maintained on current antipsychotic and other concomitant psychotropic medications for a specified period of time sufficient to minimize potential complications of assessment of cognitive status (i.e., at least 6 weeks) and on current dose for at least 2 weeks. Any dose changes during study participation could not exceed ±25% of dose at the beginning of the period.
4. No more than a “moderate” severity rating on the hallucinations and delusions items of the Positive and Negative Syndrome Scale (PANSS; Kay, Opler, & Fiszbein, 1987). Note that the PANSS rather than the Brief Psychiatric Rating Scale (BPRS; Ventura, Nuechterlein, Subotnick, Gutkind, & Gilbert, 2000), which was used in the MCCB validation study, was used for this study.
5. No more than a “moderate” severity rating on positive formal thought disorder (e.g., PANSS Conceptual Disorganization item score).
6. No more than a “moderate” severity rating on any negative symptom item (e.g., PANSS Items N1–N7).
7. A minimal level of extrapyramidal symptoms—for example, Simpson–Angus Scale (Simpson & Angus, 1970) total score (excluding items for leg drop and head drop) less than 5. Note that the MCCB validation study cutoff was 6, but that used the full 10-item scale whereas we used 8 items.
8. A minimal level of depressive symptoms, as reflected in a Calgary Depression Scale (Addington, Addington, & Schissel, 1990) total score less than 10.

Exclusion criteria were as follows:

1. Inability to give informed consent.
2. Diagnosis of schizoaffective disorder.
3. Substance dependence in the past 6 months.
4. Positive result on a laboratory toxicology screen performed during the study.
5. Blood alcohol content > 0.05% during any study visit.
6. Diagnosis of dementia or other neurodegenerative disorder.
7. Diagnosis of other neurologic disorder (e.g., epilepsy, Tourette’s syndrome).
8. Significant head injury or other brain injury leading to cognitive impairment.
9. Mental retardation (premorbid IQ < 70).
10. First-degree relative with schizophrenia (for controls).
11. Presence of a psychotic, mood, or substance abuse disorder (for controls).
12. Electroconvulsive therapy (ECT) treatments within the past 8 weeks.

Participants took part in 5 research sessions. Demographic, diagnostic, symptom, and functional assessment data were collected during Session 1. Half the sample received IntegNeuro during Sessions 2 and 4 and the MCCB during Sessions 3 and 5, and the other half of the sample received the MCCB during Sessions 2 and 4 and IntegNeuro during Sessions 3 and 5. Sessions 2 and 4 were held one month apart, as were Sessions 3 and 5. Sessions 2 and 3 took place within one week of each other, as did Sessions 4 and 5. During all testing sessions, participants provided a urine sample for a toxicology screen prior to completing any study measures in order to assure that a homogenous sample was studied without the potential effect of drugs of abuse. Participants for whom laboratory data came back as positive for an illegal substance were excluded from further participation in the study. In such cases, a new participant was enrolled to replace the excluded participant, unless exclusion occurred after Session 3. Participants also completed a breathalyzer examination during all testing visits to rule out alcohol use. Any participant with a blood alcohol content (BAC) greater or equal to 0.05% was not permitted to continue with the study.

**Cognitive assessment batteries**

The specific tests that comprise each of the batteries are listed in Table 2. Tests in the same row measure similar constructs. Each domain measured by the MCCB has at least one comparable test in IntegNeuro, with the exception of visual learning. However, the MCCB visual learning measure is a memory test, and verbal and nonverbal (visual) memory scores are highly correlated in schizophrenia (Calev, Kopol, Kugelmas, & Lerer, 1987).
suggesting that this domain can be adequately assessed via a single memory test. Nevertheless, we examined the executive maze score (which includes a visual learning component—see Appendix), as a visual learning analogue to the MCCB visual learning test. IntegNeuro measures motor speed and language function, domains not assessed by the MCCB battery.

**IntegNeuro**

IntegNeuro tests tap the following domains of cognitive function: motor speed, attention/vigilance, working memory, verbal learning, visual learning, speed of processing, language, reasoning and problem solving, and social cognition. The measures in each of the above domains are described in the Appendix. Scoring of responses was conducted using an automated software program for most tests and by hand scoring for .wav files. Hand scoring was required for the two language tests and the verbal memory test. Trained research assistants conducted the hand scoring of the .wav files. IntegNeuro typically takes 60–70 min to complete.

**MCCB**

The MCCB taps the following domains: attention/vigilance, working memory, verbal learning, visual learning, speed of processing, reasoning and problem solving, and social cognition. The measures in each of the above domains are described in the Appendix. Hand scoring was required for all tests except the computerized verbal working memory test. The MCCB typically takes 60–70 min to complete.

**Diagnostic and symptom ratings**

Diagnostic and symptom ratings were completed during the baseline session. Participants with schizophrenia were interviewed using the Structured Clinical Interview for DSM–IV Diagnosis (SCID-IV; DSM–IV = Diagnostic and Statistical Manual of Mental Disorders–Fourth Edition), Patient Version 2.0 (First, Spitzer, Gibbon, & Williams, 1995). The Overview, Mood Disorder, Psychotic Disorder, and Substance Abuse Disorder modules were administered. Control participants were interviewed with the nonpatient version of the SCID (SCID-NP). Current symptoms were assessed in schizophrenia patients using the Positive and Negative Syndrome Scale (PANSS; Kay et al., 1987). Depression was assessed using an instrument specifically developed to assess depression in schizophrenia, the Calgary Depression Scale (Addington et al., 1990).

**Functional assessment**

All patient participants completed the UPSA-2 (Patterson et al., 2001). The UPSA measures a person’s ability to follow instructions and complete everyday tasks. Patients also completed the SFS (Birchwood et al., 1990). The SFS measures seven areas of functioning that are crucial for maintaining individuals with schizophrenia in the community: employment, social withdrawal, prosocial activities, recreation, interpersonal functioning, perceived independence competence, and perceived independence performance. A total score is calculated by adding all the subscale scores. All patients were also rated on the Multi-dimensional Scale of Independent Functioning (MSIF).
Side effects

The Simpson–Angus Scale (Simpson & Angus, 1970), Abnormal Involuntary Movement Scale (AIMS; Guy, 1976), and the Barnes Akathesia Scale (Barnes, 2003) were used to assess motor side effects. On the Simpson–Angus Scale, all items except head and leg drop were administered. The Hillside Adverse Events scale was also administered to further rule out severe side effects that could affect test performance.

Breathalyzer testing

To assess blood alcohol content, we used the AlcoHawk™ ABI Professional model. This is a U.S. Department of Transportation National Highway Traffic Safety Administration (DOT/NHTSA) approved professional-grade alcohol breath analyzer that utilizes a state-of-the-art semiconductor oxide sensor to test breath alcohol content quickly and accurately.

Urine toxicity screen

The OnTrak TesTcard 9, by Varian, Inc., was used to screen for drug use. The TesTcard 9 screens urine samples for the nine most commonly abused drugs—cocaine, tetrahydrocannabinol (THC)/cannabis, barbiturates, benzodiazepines, PCP, methamphetamine, tricyclic antidepressants, morphine, and amphetamines. A positive finding in any drug category except tricyclic antidepressants was an exclusion criterion for this study.

RESULTS

Missing data across all sessions

The frequency of missing data was higher for IntegNeuro than for the MCCB. For IntegNeuro, the total amount of calculable missing data across all tests was 3.91%, and the percentage of participants whose datasets contained missing data was 32.31%. In contrast, for the MCCB, the total amount of calculable missing data across all tests was 0.89%, and the percentage of patients whose datasets contained missing data was 11.30%. For IntegNeuro, the main reason for missing data was non-completion of practice tests, which occurred in 2.74% of all cases, leading to lack of data from non-administration of the actual associated test. This accounted for 59.09% of missing data points. Overall, the primary reasons for missing data during testing sessions were incorrect responses or incorrect technique used during responding (associated with 29.09% of all missing data), palm of the hand or other body part resting on the touch screen (24.24%), lack of response (20.91%), poor recording quality or absent files (12.42%), and premature responding (10.30%). For the MCCB, the main reason for missing data was incorrect test administration, which accounted for 65% of all missing MCCB data. In all cases except one, data loss from incorrect administration occurred during the Brief Visual Memory Test–Revised. In 15% of MCCB missing data cases, the data were lost after testing, and in 20% of cases the cause of missing data was unknown.

Data transformation

All IntegNeuro scores requiring transformation and their respective transformations are listed in Table 3. As recommended in the MCCB manual, the following MATRICS scores were transformed using the natural logarithm function: Trail Making Test Part A completion time, Hopkins Verbal Learning Test, Revised (HVLT–R) total recall score, and NAB Mazes score.

Correlational analyses between test and cognitive domain scores (Pearson r and canonical)

The correlational analyses described below are organized in terms of the seven a priori cognitive factors associated with the MCCB.

Attention/vigilance

Pearson product–moment correlation. The correlation matrix that was created included the IntegNeuro Visual Working Memory scores of false positive errors, false negative errors, reaction time, and d′, and the MCCB CPT-IP 2-digit, 3-digit, 4-digit, and total d′ scores. Correlations between MCCB CPT-IP total d′ and IntegNeuro attention/vigilance measures ranged from .25 to .55 (p < .001 for all values). The highest correlation between attention/vigilance measures (r = .55) was observed between the IntegNeuro d′ and MCCB CPT-IP total d′ indices.

Working memory

Pearson product–moment correlation. The correlation matrix included the IntegNeuro Letter–Number Span total trials correct, Span of Visual Memory total score, Forward Digit Span total, and Reverse Digit Span total scores, and the MCCB Letter–Number Span and Wechsler Memory Scale–Third Edition (WMS–III) spatial span scores. Correlations between MCCB Letter–Number Span score and IntegNeuro working memory measures ranged from .42 to .74 (p < .001 for all values). The highest correlation (r = .74, N = 210) was observed with the IntegNeuro Letter–Number Span total trials correct score. Correlations between MCCB WMS–III spatial span score and IntegNeuro working memory measures ranged from .41 to .56 (p < .001 for all values). Correlations of .56 were observed for IntegNeuro measures of Span of Visual Memory total and Reverse Digit Span total score.

Verbal learning

Pearson product–moment correlation. The correlation matrix included the IntegNeuro Verbal Learning and Recall immediate recall and delayed recall scores and the
MCCB HVLT–R scores. Correlations between MCCB HVLT–R total score and IntegNeuro verbal learning measures ranged from .47 to .64 (p < .001 for all values). The highest correlation with HVLT–R total score (r = .64, N = 225) was observed for IntegNeuro Verbal Learning and Recall Trials 1–4 total score.

**Visual learning**

**Pearson product–moment correlation.** The correlation matrix included the IntegNeuro Executive Maze scores of total completion time, total errors, total overruns, and total trials, and the MCCB Brief Visuospatial Memory Test–Revised (BVMT–R) individual and total scores. Correlations between MCCB HVLT–R total score and IntegNeuro visual learning measures ranged from .26 to .64 (p < .001 for all values). Correlations with HVLT–R total score of .64 were observed for IntegNeuro Executive Maze measures of total completion time and total errors.

**Speed of processing**

**Pearson product–moment correlation.** The correlation matrix included the IntegNeuro Average Choice Reaction Time, FAS individual and total scores, Switching of Attention completion time scores, and word generation score, and the MCCB Category Fluency, Trail Making Test (TMT) Part A, and BACS symbol coding scores. Correlations between MCCB Category Fluency score and IntegNeuro speed of processing measures ranged from .35 to .72 (p < .001 for all values). The highest correlation with Category Fluency score (r = .72, N = 226) was observed for IntegNeuro Word Generation total score. Correlations between MCCB TMT Part A score and IntegNeuro speed of processing measures ranged from .35 to .62 (p < .001 for all values). The highest correlation with TMT Part A score (r = .62, N = 222) was observed for IntegNeuro Switching of Attention Part 1 completion time. Correlations between MCCB BACS symbol coding score and IntegNeuro speed of processing measures ranged from .45 to .74 (p < .001 for all values). The highest correlation with BACS symbol coding score (r = .74, N = 217) was observed for IntegNeuro Switching of Attention Part 2 completion time.

**Reasoning and problem solving**

**Pearson product–moment correlation.** The correlation matrix included the IntegNeuro Executive Maze scores of total completion time, total errors, total overruns, and total trials, and the MCCB NAB Mazes score. Correlations between MCCB NAB Mazes score and IntegNeuro reasoning and problem-solving measures ranged from .15 to .55 (p < .05 for all values). A correlation with NAB Maze score of .55 was observed for the IntegNeuro measure of total completion time.

### TABLE 3
Transformations used for IntegNeuro™ and MCCB scores

<table>
<thead>
<tr>
<th>Battery</th>
<th>Measure</th>
<th>Transformation</th>
</tr>
</thead>
<tbody>
<tr>
<td>IntegNeuro™</td>
<td>Visual Working Memory reaction time</td>
<td>1/x</td>
</tr>
<tr>
<td></td>
<td>Forward Digit Span total score</td>
<td>/x</td>
</tr>
<tr>
<td></td>
<td>Reverse Digit Span total score</td>
<td>/x</td>
</tr>
<tr>
<td></td>
<td>LNS total trials correct</td>
<td>sin⁻¹(3/24)</td>
</tr>
<tr>
<td></td>
<td>Span of Visual Memory total score</td>
<td>sin⁻¹(3/16)</td>
</tr>
<tr>
<td></td>
<td>Verbal Learning and Recall Immediate Recall 1</td>
<td>sin⁻¹(3/12)</td>
</tr>
<tr>
<td></td>
<td>Verbal Learning and Recall Immediate Recall 2</td>
<td>sin⁻¹(3/12)</td>
</tr>
<tr>
<td></td>
<td>Verbal Learning and Recall Immediate Recall 3</td>
<td>sin⁻¹(3/12)</td>
</tr>
<tr>
<td></td>
<td>Verbal Learning and Recall Immediate Recall 4</td>
<td>sin⁻¹(3/12)</td>
</tr>
<tr>
<td></td>
<td>Verbal Learning and Recall Immediate Recall Trials 1–4 total score</td>
<td>sin⁻¹(3/48)</td>
</tr>
<tr>
<td></td>
<td>Verbal Learning and Recall Immediate Recall Trials 5 score</td>
<td>sin⁻¹(3/12)</td>
</tr>
<tr>
<td></td>
<td>Verbal Learning and Recall Delayed Recall Trial 6 score</td>
<td>sin⁻¹(3/12)</td>
</tr>
<tr>
<td></td>
<td>Verbal Learning and Recall Delayed Recall Trial 7 score</td>
<td>sin⁻¹(3/12)</td>
</tr>
<tr>
<td></td>
<td>Executive Maze total completion time</td>
<td>ln(1/x)</td>
</tr>
<tr>
<td></td>
<td>Executive Maze total errors</td>
<td>ln(1/x)</td>
</tr>
<tr>
<td></td>
<td>Executive Maze total overruns</td>
<td>ln(1/x)</td>
</tr>
<tr>
<td></td>
<td>Average Choice Reaction Time</td>
<td>1/x</td>
</tr>
<tr>
<td></td>
<td>Switching of Attention Part 1 completion time</td>
<td>1/x</td>
</tr>
<tr>
<td></td>
<td>Switching of Attention Part 2 completion time</td>
<td>1/x</td>
</tr>
<tr>
<td></td>
<td>Emotion Identification percentage correct</td>
<td>sin⁻¹(3/100)</td>
</tr>
<tr>
<td></td>
<td>Emotion Identification reaction time</td>
<td>1/x</td>
</tr>
<tr>
<td></td>
<td>Emotion Identification total percentage correct</td>
<td>sin⁻¹(3/100)</td>
</tr>
<tr>
<td>MCCB</td>
<td>HVLT–R Learning total score</td>
<td>ln(x)</td>
</tr>
<tr>
<td></td>
<td>Trail Making Test Part A score</td>
<td>ln(1/x)</td>
</tr>
<tr>
<td></td>
<td>NAB Mazes score</td>
<td>ln(x)</td>
</tr>
</tbody>
</table>

Social cognition

Pearson product–moment correlation. The correlation matrix included the IntegNeuro Emotion Identification percentage correct and reaction time scores and the MCCB Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT) managing emotions (ME) score. Correlations between MCCB MSCEIT ME score and IntegNeuro social cognition measures ranged from .06 to .40. A correlation of .40 ($p < .001$, $N = 227$) was observed between the MSCEIT ME score and IntegNeuro Emotion Identification total percentage correct.

Domain and composite scores

Before calculation of domain and composite scores, each critical IntegNeuro measure was corrected for the participant’s age and gender. This age and gender correction was conducted for each variable using formulas for the age- and gender-specific mean trend and age-corrected standard deviations as generated by Brain Resource from their large normative database. Age- and gender-corrected IntegNeuro values were then calculated by subtracting the age and gender specific trend and dividing by the age-adjusted standard deviation.

Domain scores with multiple measures included in the calculation were calculated as the mean of its components. The composite score was calculated as the mean of its domain score components. Two composite scores were calculated: one composite score with all domains including the visual learning domain score and one composite score excluding the visual learning domain score to prevent any possible bias by representing Executive Maze performance twice. MCCB domain and composite scores were calculated as age- and gender-corrected $t$ scores as described in the MCCB manual.

Domain scores

Speed of processing  186  .78
Attention/processing  192  .55
Working memory  190  .70
Verbal learning  225  .59
Visual learning  212  .62
Reasoning and problem solving  225  .54
Social cognition  227  .36
Composite score  184  .85


Relationship between IntegNeuro™ and MCCB scores

The effect sizes (Cohen’s $d$) of the difference between schizophrenia patients and healthy controls on each cognitive domain score, for each battery, are listed in Table 5. The effect sizes are similar across the two batteries except for the verbal learning and reasoning and problem-solving domains. For these two domains, the IntegNeuro measures were more sensitive to differences between schizophrenia patients and healthy controls. While the IntegNeuro composite score also appears to be more sensitive to detecting differences between participant groups, the effect sizes from the two batteries were not significantly different.

Battery-wise comparisons

Finally, the set of IntegNeuro variables was correlated with the set of MCCB variables. IntegNeuro variables included in the analysis were: Visual Working Memory $d'$, Letter–Number Span trials correct, Span of Visual Memory total score, Verbal Learning and Recall Immediate Recall Trials 1–4 total score, FAS average letter score, word generation total score, average choice reaction time, Switching of Attention (Parts 1 and 2) completion times, Executive Maze time to completion, and Emotion Identification total percentage correct. MCCB variables included in the analysis were: CPT-IP total $d'$ score, Letter–Number Span score, WMS–III spatial span score, HVLT–R total score, BVMT total score, category fluency score, TMT Form A time to completion, BACS symbol-coding score, NAB Maze total score, and MSCEIT ME score. The canonical correlation analysis indicated a single function with $r = .91$ and $r^2 = .84$. There were five significant pairs of canonical variables. The first canonical correlation coefficient was .91 with the first IntegNeuro canonical variable explaining 39% of the MCCB variance. All other IntegNeuro canonical variables explained less than 3% of the overall MCCB variance; the contributions of each were considered negligible.

Degree of deviation of IntegNeuro and MATRICS test scores from control norms

The effect sizes (Cohen’s $d$) of the difference between schizophrenia patients and healthy controls on each cognitive domain score, for each battery, are listed in Table 5. The effect sizes are similar across the two batteries except for the verbal learning and reasoning and problem-solving domains. For these two domains, the IntegNeuro measures were more sensitive to differences between schizophrenia patients and healthy controls. While the IntegNeuro composite score also appears to be more sensitive to detecting differences between participant groups, the effect sizes from the two batteries were not significantly different.

Relationship between cognitive test scores and functional assessment measures

SFS

For IntegNeuro, correlations between total battery score and SFS subscale scores ranged between .05 and .18. The correlation between the IntegNeuro total
battery composite score and SFS total score was $r = .09$ ($p = .34$, $N = 122$). For the MCCB, correlations between total battery score and SFS subscale scores ranged between .012 and .15. The correlation between the MCCB total battery composite score and SFS total score was $r = .05$ ($p = .61$, $N = 131$).

**UPSA-2**

For IntegNeuro, correlations between total battery score and UPSA-2 subscale scores ranged from .05 to .35. The correlation between the IntegNeuro total battery composite score and UPSA-2 total score was $r = .33$ ($p < .001$, $N = 115$). For the MCCB, correlations between total battery score and UPSA-2 subscale scores ranged from .13 to .37. The correlation between the MCCB total battery composite score and UPSA2 total score was $r = .42$ ($p < .001$, $N = 121$). Data on the MSIF were not analyzed for this study, and will be presented in a forthcoming paper.

**Test–retest reliability**

Test–retest reliability over a 1-month interval was calculated for the critical indices for both batteries. These values are listed in Tables 6 (for IntegNeuro) and 7 (for the MCCB). In general, test–retest reliability was high for all MCCB scores and adequate to high for IntegNeuro test scores except for CPT, Choice Reaction Time (RT), Visual Span, and Verbal Learning and Recall recall scores. The reliability values for these scores were significantly lower than those in a previously published report (Williams et al., 2005). While the lower reliability coefficients for these four scores compared to their MCCB counterparts may reflect a real difference between the test batteries, the values may reflect range restriction. Specifically, these measures had large amounts of missing data due to participants failing the practice test prior to attempting the real test (i.e., failed practice tests for these measures accounted for the majority of missing IntegNeuro data). In contrast, for the MCCB, there are no criteria that must be passed prior to

**TABLE 5**

<table>
<thead>
<tr>
<th>Cognitive domain</th>
<th>MCCB</th>
<th>IntegNeuro™</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attention/vigilance</td>
<td>1.39</td>
<td>1.06</td>
</tr>
<tr>
<td>Working memory</td>
<td>1.05</td>
<td>1.08</td>
</tr>
<tr>
<td>Verbal learning</td>
<td>0.96</td>
<td>1.58</td>
</tr>
<tr>
<td>Visual learning</td>
<td>1.34</td>
<td>1.36</td>
</tr>
<tr>
<td>Speed of processing</td>
<td>1.65</td>
<td>1.46</td>
</tr>
<tr>
<td>Reasoning and problem solving</td>
<td>0.81</td>
<td>1.55</td>
</tr>
<tr>
<td>Social cognition</td>
<td>0.95</td>
<td>0.89</td>
</tr>
<tr>
<td>Composite score</td>
<td>1.67</td>
<td>1.99</td>
</tr>
</tbody>
</table>

Note. Effect size = Cohen’s $d$ (i.e., difference of means divided by healthy control SD). MCCB = MATRICS Consensus Cognitive Battery. MATRICS = Measurement and Treatment Research to Improve Cognition in Schizophrenia.

**TABLE 6**

IntegNeuro test–retest reliability over a 4-week interval

<table>
<thead>
<tr>
<th>Cognitive domain</th>
<th>All participants</th>
<th>Healthy control participants only</th>
<th>Schizophrenia participants only</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N$</td>
<td>ICC</td>
<td>$r$</td>
</tr>
<tr>
<td>Attention vigilance</td>
<td>167</td>
<td>.53</td>
<td>.53</td>
</tr>
<tr>
<td>Visual Memory $d'$</td>
<td>179</td>
<td>.70</td>
<td>.72</td>
</tr>
<tr>
<td>Working memory</td>
<td>188</td>
<td>.53</td>
<td>.53</td>
</tr>
<tr>
<td>LNS total trials correct</td>
<td>210</td>
<td>.72</td>
<td>.73</td>
</tr>
<tr>
<td>Span of Visual Memory total score</td>
<td>212</td>
<td>.81</td>
<td>.85</td>
</tr>
<tr>
<td>Verbal learning</td>
<td>210</td>
<td>.73</td>
<td>.74</td>
</tr>
<tr>
<td>Immediate Recall Trials 1–4 total score</td>
<td>162</td>
<td>.65</td>
<td>.66</td>
</tr>
<tr>
<td>Visual learning</td>
<td>200</td>
<td>.74</td>
<td>.78</td>
</tr>
<tr>
<td>Executive Maze total errors</td>
<td>202</td>
<td>.65</td>
<td>.67</td>
</tr>
<tr>
<td>Reasoning and problem solving</td>
<td>201</td>
<td>.74</td>
<td>.78</td>
</tr>
<tr>
<td>Emotional Identification total percentage correct</td>
<td>210</td>
<td>.79</td>
<td>.80</td>
</tr>
</tbody>
</table>

Note. ICC calculated according to Shrout and Fleiss’s ICC (1,1). $N =$ number of participants. ICC = intraclass correlation coefficient. $r =$ Pearson product–moment correlation coefficient. LNS = Letter–Number Span.
completing equivalent tests, and so test data are obtained for all participants who attempt a test. It is therefore possible that the lower reliability seen with these IntegNeuro tests could be due partly to a restriction of range due to the loss of data from the most impaired patients. This hypothesis is supported by the larger amount of missing data from patients than healthy controls and the assumption that the majority of missing data due to failed practice tests would occur in the most severely ill patient cases. Raju and Brand (2003) explained how restriction of range affects Pearson product–moment correlations and how correlations can be adjusted to take this into account. Their correction can be expressed as:

\[
R = \sqrt{\frac{1}{1 + x}} \times R \text{ (original correlation value)}
\]

where \(x = \text{(variance of control group)/(variance of combined groups)}\). Range restriction is apparent—and the correction applied—when the ratio for \(x\) is at least .05 (indicating a smaller variance, i.e., restriction, than that for the combined groups). For this data set, there were no cases where this ratio was larger than .50.

Based on the Raju and Brand correction formulas, we reanalyzed the test–retest reliability data from the MCCB to include only those participants who passed the corresponding IntegNeuro test practice test and assumed a restriction of range. In addition, to make the CPT values more equivalent, we based the MCCB CPT-IP \(d’\) score on the two-digit condition only, since this was the closest to the IntegNeuro CPT condition (which was single digit), and therefore this comparison afforded the conceptually clearest test of the extent of similarity in test–retest reliability. The results of these reliability recalculations are presented in Table 8, where it can be seen that the reliability results are more equivalent than those presented in Tables 6 and 7. Using the revised statistics, across all participants, the average reliability coefficients \(r\) are .81 for the MCCB and .74 for IntegNeuro. For schizophrenia participants only, the values are .67 for the MCCB and .68 for IntegNeuro, and for controls only the values are .79 for the MCCB and .76 for IntegNeuro.

### Practice effects

In addition to determining reliability across two administrations, we determined whether scores changed significantly across the two administrations (i.e., practice effects). The significance values and effect sizes associated with these analyses are reported in Table 9. Significant performance differences were observed for 7 of the 10 MCCB tests and 4 of the 11 IntegNeuro tests. The average effect size for IntegNeuro tests was 0.15 (range = 0.03 to 0.34), and for the MCCB it was 0.13 (range = 0 to 0.26).

### TABLE 8
Test–retest reliability values after correcting for range restriction

<table>
<thead>
<tr>
<th>MCCB</th>
<th>All Controls Schiz.</th>
<th>IntegNeuro</th>
<th>All Controls Schiz.</th>
<th>All (N) Schiz (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category Fluency</td>
<td>.75 .67 .69</td>
<td>Category Fluency</td>
<td>.76 .85 .72</td>
<td>210 141</td>
</tr>
<tr>
<td>Trails A</td>
<td>.74 .69 .66</td>
<td>Switching of Attention Part 1</td>
<td>.68 .52 .74</td>
<td>203 133</td>
</tr>
<tr>
<td>BACS Symbol Coding</td>
<td>.93 .88 .87</td>
<td>Choice RT</td>
<td>.66 .85 .56</td>
<td>163 105</td>
</tr>
<tr>
<td>CPT-IP</td>
<td>.85a .87a .77a</td>
<td>Verbal Interference</td>
<td>.78 .72 .73</td>
<td></td>
</tr>
<tr>
<td>Verbal LNS</td>
<td>.83 .80 .80</td>
<td>Verbal LNS</td>
<td>.72 .68 .64</td>
<td>178 117</td>
</tr>
<tr>
<td>WMS–III Visual Spatial Span</td>
<td>.77 .79 .71</td>
<td>Span of Visual Memory</td>
<td>.56 .65 .50</td>
<td>187 120</td>
</tr>
<tr>
<td>HVLT–R</td>
<td>.78 .78 .66</td>
<td>Verbal Recall and Learning</td>
<td>.73 .79 .69</td>
<td>211 142</td>
</tr>
<tr>
<td>BMVT–R</td>
<td>.80 .79 .69</td>
<td>Executive Maze errors</td>
<td>.85 .84 .79</td>
<td>197 126</td>
</tr>
<tr>
<td>NAB–Mazes</td>
<td>.81 .78 .76</td>
<td>Executive Maze time</td>
<td>.86 .85 .77</td>
<td>211 140</td>
</tr>
<tr>
<td>MSCEIT–Managing Emotions</td>
<td>.81 .83 .77</td>
<td>Emotion Identification</td>
<td>.82 .87 .76</td>
<td>211 140</td>
</tr>
</tbody>
</table>

### TABLE 9
Practice effects

<table>
<thead>
<tr>
<th>MCCB</th>
<th>t = 2.04, p = .04, d = 0.13</th>
<th>Category Fluency</th>
<th>t = 0.98, p = .10, d = 0.14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trails A</td>
<td>t = 4.20, p &lt; .001, d = 0.26</td>
<td>Switching of Attention (Trails A)</td>
<td>t = 1.58, p = .12, d = 0.12</td>
</tr>
<tr>
<td>BACS Symbol Coding</td>
<td>t = 3.03, p = .003, d = 0.13</td>
<td>Choice RT</td>
<td>t = 0.55, p = .58, d = 0.05</td>
</tr>
<tr>
<td>CPT-IP total d'</td>
<td>t = 3.38, p &lt; .001, d = 0.19</td>
<td>Visual Working Memory (1-back) d'</td>
<td>t = 1.55, p = .13, d = 0.16</td>
</tr>
<tr>
<td>Verbal LNS</td>
<td>t = 2.08, p = .04, d = 0.10</td>
<td>Verbal LNS</td>
<td>t = 2.65, p = .009, d = 0.20</td>
</tr>
<tr>
<td>WMS–III Spatial Span</td>
<td>t = 3.00, p = .003, d = 0.19</td>
<td>Spatial Span</td>
<td>t = −0.71, p = .48, d = 0.07</td>
</tr>
<tr>
<td>HVLT–R</td>
<td>t = 0.05, p = .96, d = 0.00</td>
<td>Verbal Learning and Recall</td>
<td>t = −2.70, p = .008, d = 0.22</td>
</tr>
<tr>
<td>BMVT–R</td>
<td>t = 1.80, p = .07, d = 0.12</td>
<td>Executive Maze (errors)</td>
<td>t = 4.65, p &lt; .001, d = 0.28</td>
</tr>
<tr>
<td>NAB Mazes</td>
<td>t = 2.75, p = .007, d = 0.15</td>
<td>Executive Maze (time)</td>
<td>t = 5.68, p &lt; .001, d = 0.34</td>
</tr>
<tr>
<td>MSCEIT–Managing Emotions</td>
<td>t = 0.97, p = .34, d = 0.06</td>
<td>Emotion Identification</td>
<td>t = −0.59, p = .55, d = 0.03</td>
</tr>
</tbody>
</table>

### DISCUSSION

The specific aim of this research study was to determine the strength of relationships between scores from the IntegNeuro and MATRICS cognitive batteries. This was tested by having 155 schizophrenia and 75 control participants complete both batteries, with administration order randomized across participants and groups. In addition, a month after taking each battery, participants completed each battery a second time, to aid in comparing the test–retest reliability and practice effect characteristics of each battery in the same sample.

Results of the correlational analyses indicated that the two test batteries are comparable. The canonical correlation between the sets of critical test indices for each battery was high (.91) and accounted for 84% of the variance in the test battery scores. Moreover, the Pearson correlation between the single composite total score for each battery was also high (.81).

Analysis of individual cognitive domains indicated that while in some domains the two batteries were comparable, in other domains this was not the case. The strongest relationships between MCCB and IntegNeuro domains were observed for working memory (where the highest correlation between test indices was .73) and speed of processing (where the highest correlation was .71). The lowest correlations were in the domains of attention/vigilance, where the correlations ranged from...
.20 to .55, and social cognition, where the correlation between total score on each measure was .42. This variability in correlations can be accounted for by the nature of the tests in each battery. For example, in the social cognition domain, the MCCB test is a verbally mediated problem-solving task that involves anticipating the effectiveness of different emotion-based behaviors in an interpersonal situation. In contrast, the IntegNeuro social cognition measure involves recognizing affect while looking at pictures of faces and is equivalent to the original Penn emotion identification task. Because social cognition is a multidimensional construct (Green et al., 2008), it would not be expected that the correlation between these two tests would be high. For the attention/vigilance factor, the lowest correlations were observed with the IntegNeuro false-positive error variable. However, because most participants make few false-positive errors, this variable, although important, is not a good proxy for attentional ability. The most relevant IntegNeuro variable from this domain is d', which reflects sensitivity to the target, and is calculated using signal detection analyses. The correlation between d' and the MCCB CPT-IP total d' score was .55, which indicates a greater degree of comparability.

The correlations observed in this study are within the same range as those from past studies determining comparability between neurocognitive test batteries. For example, in Keefe et al.'s (2004) validation study of the BACS, correlations between BACS measures/factors and scores from other neuropsychological tests of similar cognitive functions were .61 for verbal memory, from .45 to .59 for executive functioning, .74 for verbal fluency, and, for composite scores, .76 for patients and .90 for controls. Velligan et al. (2004) obtained a correlation of .72 between total scores on the Brief Cognitive Assessment (BCA) and a battery composed of similar measures. A study of the CogTest battery reported correlations between individual CogTest measures and analogous neuropsychological tests to range from .30 to .70 ( Sharma & Bilder, 2004). A validation study for the RBANS (Hobart et al., 1999) reported correlations between .20 and .75 between individual RBANS tests and comparable neuropsychological tests, and correlations between .52 and .75 between RBANS total scores and Wechsler Adult Intelligence Scale (WAIS) and WMS composite scores. In the IntegNeuro validation study (Paul et al., 2005), all correlations between individual IntegNeuro measures and paper-and-pencil versions of similar tests exceeded .53. A recent study compared paper-and-pencil versions of the tests that make up the Clinical Antipsychotic Trials of Intervention Effectiveness study (CATIE; Keefe et al., 2007) and MATRICS batteries to computerized versions of these same tests—the Computerized Multiphasic Interactive Neurocognitive DualDisplay System (CMINDS; O'Halloran et al., 2008). This study found that the CMINDS tests produced equivalent scores and were characterized by equivalent levels of test–retest reliability (but see Kern, Green, Nuechterlein, & Keefe, 2009). Intraclass correlations between comparable MCCB and CMINDS tests ranged from .64 to .95 during the first testing on each measure, and test–retest reliability of CMINDS tests ranged from .56 to .94. In short, the observed correlations between IntegNeuro and the MCCB indicate that IntegNeuro scores parallel MCCB scores to an equal or greater extent than other batteries' scores correlated with the tests against which these batteries were validated in published studies. This is important because, to date, there are no agreed-upon criteria for determining whether test batteries are equivalent or not. Because we believe test battery comparability to be on a dimensional scale, as opposed to a dichotomous variable, we have chosen not to put forth such criteria in this paper, but to present the data and our interpretation of it, recognizing that it might be possible to evaluate the results differently.

IntegNeuro demonstrated adequate test–retest reliability in this study. Intraclass correlations (ICC) for critical single test scores ranged from .40 to .80 for controls and from .40 to .74 for the schizophrenia group. Pearson r values ranged from .46 to .84 for controls and from .41 to .78 for the schizophrenia group. For MCCB, the ICC values ranged from .69 to .79 for controls, and from .66 to .86 for the schizophrenia group. The latter is similar to the range of .58 to .85 reported for schizophrenia patients in the MCCB normative study (Nuechterlein et al., 2008). MCCB Pearson r values ranged from .67 to .87 for controls and from .66 to .87 for the schizophrenia group. Overall, the ICC values for MCCB tests were approximately .09 units higher for controls and .14 units higher for patients than those for IntegNeuro tests. Pearson r values were .11 units higher for controls and .12 units higher for the patient sample than those for IntegNeuro tests. When adjusted for range restriction in the IntegNeuro data due to patient participants' missing data points, however, the average differences between batteries, in Pearson r units, were .03 for controls in favor of the MCCB and .01 for patients in favor of IntegNeuro.

Neither battery demonstrated strong correlations with the functional outcome measures the SFS or UPSA-2—although correlations between battery composite scores and UPSA-2 composite scores were statistically significant—consistent with the findings from the MCCB normative study (Kern et al., 2008). This may reflect the more complex and multitedetermined phenomenon of real-world functioning compared to specific neurocognitive test performance and is consistent with the recent NIMH initiative to include functional outcome measures as coprimary outcome variables in clinical trials. It is notable, however, that the correlations we observed in this study between cognitive test scores and functional assessment measures are substantially lower than those reported in other studies (e.g., Bowie, Reichenberg, Patterson, Heaton, & Harvey, 2006; Pietrzak et al., 2009).

Practice effects were statistically significant in many cases, although effect sizes of these score improvements were small in all cases. The practice effects in this study were over 50% lower than those reported among healthy controls and first-episode schizophrenia patients in a recent study (Goldberg et al., 2007), but the latter study used a different test battery, although there was some
overlap in terms of subtests. While it would seem, intuitively, that the presence of practice effects would work against obtaining significant treatment effects, it is important to note the following: (a) Practice effects are a ubiquitous problem with cognitive assessment measures; (b) these can be reduced in clinical trials by including a training phase/administration prior to baseline testing; (c) practice effects do not preclude reliable subsequent and additive change in response to treatment when ceiling effects are not an issue; and (d) when practice effects are known, these differences can be used as the “null” hypothesis (rather than no change) against which treatment effects are measured. The data presented in this study can be useful for future studies using these batteries by providing such a baseline against which treatment-related change can be compared.

An advantage of both batteries is that, even when relatively asymptomatic patients are assessed, the difference between patient and healthy control groups is large (i.e., average Cohen’s ds for MCCB and IntegNeuro scores, including composite scores, were 1.23 and 1.37, respectively; see Table 5). This is important because the larger the difference between groups, the greater the range of change scores that is possible (i.e., large between-group differences increases sensitivity to change). In the case of two domains, verbal learning and reasoning and problem solving, the sensitivity of IntegNeuro was greater than 0.5 d units (SD) than that of the MCCB, suggesting that where these variables are the focus of interest regarding treatment effects, IntegNeuro would be the preferred assessment battery.

In summary, this study demonstrated the comparability of the IntegNeuro and MCCB batteries in terms of equivalence between cognitive domain and total battery scores, test–retest reliability, practice effects, sensitivity to cognitive deficits in schizophrenia, and correlations with functional outcome measures. These data suggest that IntegNeuro is a valid alternative to the MCCB for a cognitive assessment battery for use in clinical trials. Because there are several advantages to the use of a standardized, computerized assessment battery over paper-and-pencil measures—in terms of standardization and automatization of test instructions and scoring—use of IntegNeuro may have the additional advantage of being less resource intensive. An important caveat, however, is that test administrators for computerized batteries such as IntegNeuro must be vigilant to aspects of participant behavior that invalidate responses and lead to missing data (e.g., having sleeves drag on the touch screen; other response errors due to lack of familiarity with computers, etc). In this study, the frequent occurrence of such errors outweighed the normal rationale for using computer-administered tests, that being the negative effects of test administration errors on paper-and-pencil tests. This echoes findings from the CATIE study, in which the three cognitive assessment measures with the most missing data were all computerized tests (Keefe & Harvey, 2008). It appears that, in this study’s use of IntegNeuro, the missing data problem reflected primarily automatic skipping of tests for which the practice test had failed. These problems have been addressed via the development by Brain Resource, Ltd. of: (a) a web-based checklist that each new test administrator completes to document comprehension of testing requirements; and (b) a new software option that allows the administrator the options of repeating the practice test, and running the test on which a participant has failed practice items (Brain Resource 2009). The point remains, however, that only when computerized test administrators ensure that data integrity is not compromised by participant and test administration factors will computerized testing truly confer its potential advantages.

REFERENCES


APPENDIX

Description of the tests in each battery

IntegNeuro

Sensorimotor domain

1. Simple Motor Tapping Task: Participants were required to tap a circle on the touch screen with their index finger, as fast as possible for 60 s. The dependent variable was total number of taps with the dominant hand.

Attention/vigilance domain

1. Continuous Performance Task: Letters (B, C, D, or G) were presented to the participant on the computer screen (for 200 ms), separated by an interval of 2.5 s. If the same letter appeared twice in a row, the participant was asked to press buttons with the index finger of each hand. Speed and accuracy of response were equally stressed in the task instructions. There were 125 stimuli presented in total, 85 being nontarget letters and 20 being target letters (i.e., repetitions of the previous letter). The dependent variables were the number of errors of omission and false positives, and d′.

Working memory domain

1. Digit Span: Participants were presented with a series of digits on the touch screen (e.g., 4, 2, 7, etc.; 500-ms presentation), separated by a 1-s interval. The participant was then immediately asked to enter the digits on a numeric keypad on the touch screen. In the first part of the test, participants were required to recall the digits in forward order (Digits Forwards); in the second part, they were required to recall them in reverse order (Digits Backwards). In each part, the number of digits in each sequence was gradually increased from 3 to 9, with two sequences at each level. The dependent measure for each part was the maximum number of digits the participant recalled without error.

2. Span of Visual Memory: This test is a computerized adaptation of the Spatial Span test from the Wechsler Memory Scale (Wechsler, 1997). Participants were presented with squares arranged in a random pattern on the computer screen. The squares were highlighted in a sequential order on each trial. Participants were required to repeat the order in which the squares were highlighted by touching the squares with their forefinger. Both forward and reverse trials were conducted. The total correct was the dependent variable.

3. Letter–Number Span–Verbal (Gold, Carpenter, Randolph, Goldberg, & Weinberger, 1997). This is an orally administered test in which the respondent is instructed to mentally reorder and repeat strings of letters (alphabetical) and numbers (ascending). The trials gradually increase in difficulty. The primary score is the total number of trials repeated and reordered correctly. All trials were administered, and there was no discontinuation rule used.

Verbal learning domain

1. Verbal Recall and Learning: The participants were read a list of 12 words, which they were asked to memorize. The list contained 12 concrete words from the English language. Words were closely matched on concreteness, number of letters, and frequency. The list was presented orally four times (and received by the participant using headphones). On each of the four trials, the participant was required to recall as many words as possible by speaking directly into the attached microphone. The participant was then presented with a list of distractor words and was asked to recall them after presentation. Immediately following this, the participant was asked to recall the 12 words from the original list (short-delay recall trial). A long-delayed recall trial was completed approximately 20 min later after a number of intervening tasks. Recognition trials were then completed after the delayed trial. The dependent variables were the number of words correctly recalled across the four learning trials, the immediate recall trial, and the delayed recall trial and the total number of correctly identified words on the recognition trial.

Visual learning domain

1. Executive Maze Task: This task was a computerized adaptation of the Austin Maze (Walsh, 1985). The participant was presented with a grid (8 × 8 matrix) of circles on the computer screen. The object of the task was to identify the hidden path through the grid, from the beginning point at the bottom of the grid to the end point at the top. The participant was able to navi-
gate around the grid by pressing arrow keys (up, down, left, right). A total of 24 consecutive correct moves were required to complete the maze. The participant was presented with one tone (and a red cross at the bottom of the screen) if they made an incorrect move and a different tone (and a green tick at the bottom of the screen) if they made a correct move. The purpose of the task was therefore to assess how quickly the participant learned the route through the maze and their ability to remember that route. Only one maze was presented across trials, and the test ended when the participant completed the maze without error or after 10 min had elapsed. The dependent variable was the total maze time. It should be noted that while this measure is identified as a measure of executive function, the requirement to retain the maze in memory for two successive trials introduces an added visual memory component to the task, and therefore scores on this measure were also examined in relationship to the MCCB visual learning domain.

**Speed of processing domain**

1. **Letter Fluency:** Participants were required to generate, by speech, words that began with the letters F, A, and S. 60 s were allowed for each letter, and proper nouns were not allowed. Responses were recorded via the microphone and hand scored. Intrusive or perseverative responses were not included in the total number correct. The total number of correct words generated across the three trials was the dependent measure.

2. **Category Fluency: Animal Naming** (Spreen & Strauss, 1998). This is an oral test in which a respondent names as many animals as he or she can in 1 min. The primary score is the total number of animals named in 60 s. No points are awarded for perseverations and/or intrusions.

3. **Choice Reaction Time Task:** Participants were required to attend to the computer screen as one of four target circles was illuminated in a pseudorandom sequence over a series of trials. For each trial, the participant was required to place their index finger in preparation on a start circle displayed on the touch screen. On each trial, the participant then had to touch the illuminated circle as quickly as possible following presentation. A total of 20 trials were administered with a random delay between trials of 2–4 s. The dependent variable was the mean reaction time across trials.

4. **Switching of Attention Task:** This test is a computerized adaptation of the Trail Making Test (Army Individual Test Battery, 1944; Reitan, 1958). It consists of two parts. In the first part, the participant was presented with a pattern of 25 numbers in circles and was asked to touch them in ascending numerical sequence (i.e., 1, 2, 3 . . . ). As each number is touched in correct order, a line is drawn automatically to connect it to the preceding number in the sequence. This allowed the participant to visualize the path touched. This task tests psychomotor speed and the basic ability to hold attention on a simple task. The second part of the test is described below. The dependent variable was time to completion.

**Language domain**

1. **Verbal Interference:** This task taps the ability to inhibit automatic and irrelevant responses and has similarities to the Stroop task (Golden, 1978). The participant was presented with colored words presented serially, one at a time. Each word was drawn from the following set of lower case words: red, yellow, green, and blue. The color of each word is drawn from the following set of colors: red, yellow, green, and blue. Below each colored word is a response pad with the four possible words displayed in black and in fixed format. The test has two parts. In Part 1, the participant is required to identify the name of each word as quickly as possible after it is presented on the screen. In Part 2, the participant is required to name the color of each word as quickly as possible. Each part lasts for 1 min. Responses are made on the screen by touching the appropriate word on the response pad. The dependent variable in each part was the number of words correctly identified. This test is a computerized version of the verbal interference test from the Brief Assessment of Cognition in Schizophrenia (BACS).

2. **Spot-the-Word Task:** This task is a computerized adaptation of the Spot the Real Word Test (Baddeley, Emslie, & Nimmo-Smith, 1993). On each trial of this task, participants were presented with two words on the touch screen. One of the two words was a valid word in the English language (“true” target word), and the second was a nonword foil. Participants were required to identify, by touching the screen, which of the two words was the true target. The total correct score was the dependent measure.

**Reasoning and problem-solving domain**

1. **Executive Maze Task:** This is a computerized task that assesses constructs equivalent to those assessed by the Austin Maze (Walsh, 1985). The participant was presented with a grid (8 × 8 matrix) of circles on the computer screen. The object of the task was to identify the hidden path through the grid, from the beginning point at the bottom of the grid to the end point at the top. The participant was able to navigate around the grid by pressing arrow keys (up, down, left, right). A total of 24 consecutive correct moves were required to complete the maze. The participant was presented with one tone (and a red cross at the bottom of the screen) if they made an incorrect move and a different tone (and a green tick at the bottom of the screen) if they made a correct move. The purpose of the task was therefore to assess how quickly the participant learned the route through the maze and their ability to remember that route. Only one maze was presented across trials, and the test ended when the participant completed the maze twice without error or after 10 min had elapsed. The dependent variable was the total maze time. It
should be noted that while this measure is identified as a measure of executive function, the requirement to retain the maze in memory for two successive trials introduces an added memory component to the task, and therefore this measure taps more cognitive domains than pure executive function.

Social cognition domain

1. Emotional Identification and Delayed Recall: This is a test of emotional identification based on the one developed at the University of Pennsylvania (Gur, Sara, Hagendoorn, Marom, Hughett, et al., 2002), but modified to centralize eye placement (Williams, Mathersul, Palmer, Gur, Gur, et al., 2009). It has been adapted to a Microsoft Windows environment by BRC. Participants are presented with a series of faces with different emotional expressions (fear, disgust, happy, neutral). They are required to press on the prompts on the bars on the screen that correctly identify the emotional valence presented by the face. The dependent variable is the total correct.

MCCB

Attention/vigilance domain

1. Continuous Performance Test–Identical Pairs (Cornblatt, Risch, Faris, Friedman, & Erlenmeyer-Kimling, 1988). In this computer-administered measure of sustained attention the respondent is instructed to click a mouse button in response to any two consecutive matching numbers that briefly flash on the screen. The there are three sets of stimuli: 2-digit numbers, 3-digit numbers, and 4-digit numbers. The primary score is mean $d'$ value across 2-, 3-, and 4-digit conditions, where $d'$ is an index of signal/noise discrimination that is computed by the CPT-IP computer program.

Working memory domain

1. Wechsler Memory Scale–Third Edition: Spatial Span–nonverbal (Wechsler, 1997). The respondent is instructed to use a board with 10 irregularly spaced blocks to “tap” cubes in the same or reversed sequence as the administrator. The primary score is the sum of raw scores on Forward and Backward conditions.

2. Letter Number Span–verbal (Gold et al., 1997). This is an orally administered test in which the respondent is instructed to mentally reorder and repeat stings of letters (alphabetical) and numbers (ascending). The trials gradually increase in difficulty. The primary score is the total number of trials repeated and reordered correctly (unlike its IntegNeuro counterpart, there is a discontinuation rule).

Verbal learning domain

1. Hopkins Verbal Learning Test–Revised (Brandt & Benedict, 2001). This is an orally administered test in which a list of 12 words from three taxonomic categories is presented. The respondent is instructed to recall as many words as possible after each of three learning trials. The primary score is the total number of words recalled correctly over three learning trials.

Visual learning domain

1. Brief Visuospatial Memory Test–Revised (Benedict, 1997). In this paper-and-pencil test the respondent is instructed to reproduce (draw) six geometric figures that are presented to him or her for 10 s. The reproduction is scored on the basis of accuracy in placement of each figure, as well as the accuracy of drawing. There are three learning trials, and the primary score is the total recall sum for all learning trials.

Speed of processing domain

1. Brief Assessment of Cognition in Schizophrenia: Symbol Coding (Keefe, 1999). This is a timed paper and pencil test in which respondent uses a key to match digits to nonsense symbols. The primary score is the total number correct.

2. Category Fluency: Animal Naming (Spreen & Strauss, 1998). This is an oral test in which a respondent names as many animals as he or she can in 1 min. The primary score is the total number of animals named in 60 s. No points are awarded for perseverations and/or intrusions.

3. Trail Making Test: Part A (Army Individual Test Battery, 1944; Reitan, 1958). This is a timed paper-and-pencil test in which the respondent draws a line to connect consecutively numbered circles places irregularly in a sheet of paper. The participant is instructed to connect the numbers without lifting the pencil. The primary score is “time to completion.”

Reasoning and problem-solving domain

1. Neuropsychological Assessment Battery: Mazes (White & Stern, 2003). This timed paper-and-pencil test includes seven mazes that the respondent is instructed to navigate from beginning to end without lifting a pencil or crossing the lines. The complexity of mazes is increased with each trial, and points are awarded on the basis of “time to completion” (i.e., the shorter the completion time, the higher is the amount of points earned).

Social cognition domain

1. Mayer–Salovey–Caruso Emotional Intelligence Test: Managing Emotions (Mayer, Salovey, & Caruso, 2002). This task measures the respondent’s ability to make decisions, informed by his or her emotional state. In the task, the person rates the predicted effectiveness of various alternative behaviors in interpersonal situations where emotions are activated and must be regulated.