Neuropsychology
2015, Vol. 29, No. 2, 235–246

0894-4105/15/$12.00 http://dx.doi.org/10.1037/neu0000093

Psychometric Properties of the Penn Computerized Neurocognitive Battery

Tyler M. Moore
Perelman School of Medicine, University of Pennsylvania

Steven P. Reise
University of California, Los Angeles

Raquel E. Gur
Perelman School of Medicine, University of Pennsylvania

Hakon Hakonarson
Children's Hospital of Philadelphia, Philadelphia, Pennsylvania

Ruben C. Gur
Perelman School of Medicine, University of Pennsylvania and Philadelphia Veterans Administration Medical Center, Philadelphia, Pennsylvania

Objective: The Penn Computerized Neurocognitive Battery (CNB) was designed to measure performance accuracy and speed on specific neurobehavioral domains using tests that were previously validated with functional neuroimaging. The goal of the present study was to evaluate the neuropsychological theory used to construct the CNB by confirming the factor structure of the tests composing it. ***Method:*** In a large community sample ($N = 9{,}138$; age range 8–21), we performed a correlated-traits confirmatory factor analysis (CFA) and multiple exploratory factor analyses (EFAs) on the 12 CNB measures of Efficiency (which combine Accuracy and Speed). We then performed EFAs of the Accuracy and Speed measures separately. Finally, we performed a confirmatory bifactor analysis of the Efficiency scores. All analyses were performed with Mplus using maximum likelihood estimation. ***Results:*** Results strongly support the a priori theory used to construct the CNB, showing that tests designed to measure executive, episodic memory, complex cognition, and social cognition aggregate their loadings within these domains. When Accuracy and Speed were analyzed separately, Accuracy produced 3 reliable factors: executive and complex cognition, episodic memory, and social cognition, while speed produced 2 factors: tests that require fast responses and those where each item requires deliberation. The statistical "Fit" of almost all models described above was acceptable (usually excellent). ***Conclusions:*** Based on the analysis from these large-scale data, the CNB offers an effective means for measuring the integrity of intended neurocognitive domains in about 1 hour of testing and is thus suitable for large-scale clinical and genomic studies.

Keywords: computerized neurocognitive battery, psychometrics, factor analysis, Philadelphia Neurodevelopmental Cohort

The recent incorporation of genomics into clinical neuroscience has generated an unprecedented need for behavioral assessment of domains that can be linked to brain systems and serve as "biomarkers" of psychopathology (Insel & Cuthbert, 2009). The scale of such studies requires efficient computerized testing of a broad range of abilities using tests that have been validated with functional neuroimaging. Such "neurobehavioral probes" (Gur, Erwin & Gur, 1992) have been assembled into a computerized neurocognitive battery (CNB) in which these tasks have been adapted to assure adequate psychometric properties, such as reliability and validity (Gur et al., 2001, 2010) and its linkage to brain systems (Roalf, Ruparel et al., 2013). The CNB has been applied in large-scale genomic studies (Aliyu et al., 2006; Greenwood et al., 2007; Gur et al., 2007; Almasy et al., 2008), treatment research (Gur et al., 2001; Grant, Huh, Perivoliotis, Stolar, & Beck, 2012), and the military (Thomas et al., 2013). Although its individual tests' scores have been largely reliable (and their inferences valid), the overall latent structure of the battery has not been sufficiently evaluated. Using confirmatory and exploratory factor analysis, we aimed to examine the factor structure of the CNB, which is of paramount importance in justifying how the instrument is scored and interpreted.

The CNB currently comprises 14 tests grouped into five domains of neurobehavioral function. These broad domains were

This article was published Online First September 1, 2014.

Tyler M. Moore, Brain Behavior Laboratory, Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania; Steven P. Reise, Department of Psychology, University of California, Los Angeles; Raquel E. Gur, Brain Behavior Laboratory, Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania; Hakon Hakonarson, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania; Ruben C. Gur, Brain Behavior Laboratory, Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania and Philadelphia Veterans Administration Medical Center, Philadelphia, Pennsylvania.

This work was supported by NIMH Grants MH089983, MH089924, and MH019112.

Correspondence concerning this article should be addressed to Tyler Moore or Ruben Gur, Brain Behavior Laboratory, Perelman School of Medicine, University of Pennsylvania, 3400 Spruce Street–10th Floor Gates, Philadelphia, PA 19104. E-mail: tymoore@upenn.edu or gur@upenn.edu

selected because they represent well-established brain systems, and each test in the battery was used in functional neuroimaging studies to probe aspects of these domains (Roalf, Ruparel et al., 2013). There are three tests in each of four domains measuring Executive Control, Episodic Memory, Complex Cognition, and Social Cognition as summarized in Table 1. A fifth domain (Sensorimotor Speed) was measured by two tests, but the current analysis examined accuracy and these tests only provide speed measures and were therefore not included. Here we examined the factorial structure of the CNB in a sample of over 9,000 individuals age 8 to 21 who were administered the CNB as part of their participation in the Philadelphia Neurodevelopmental Cohort (PNC; Gur et al., 2012, 2013).

Method

Sample

The PNC sample included children (age 8–21) recruited through an NIMH-funded Grand Opportunity (GO) study characterizing clinical and neurobehavioral phenotypes in a genotyped prospectively accrued community cohort. All study participants were previously consented for genomic studies when they presented for pediatric services within the Children's Hospital of Philadelphia (CHOP) health care network. At that time they provided a blood sample for genetic studies, authorized access to Electronic Medical Records (EMRs) and gave written informed consent/assent to be recontacted for future studies. Of the 50,540 genotyped subjects, 18,344 met criteria and were randomly selected, with stratification for age, sex, and ethnicity.

The sample included ambulatory children in stable health, proficient in English, physically and cognitively capable of participating in an interview and performing the computerized neurocognitive testing. Youths with disorders that impaired motility or cognition (e.g., significant paresis or palsy, intellectual disability) were excluded. Notably, participants were not recruited from psychiatric clinics and the sample is not enriched for individuals who seek psychiatric help. A total of 9,138 enrolled in the study between 11/2009–10/2011 and were included in this analysis. Participants provided informed consent/assent after receiving a complete description of the study and the Institutional Review Boards at Penn and CHOP approved the protocol.

CNB Tests and Administration

The tests included in the CNB have been described in detail in Gur et al. (2010, 2012). The tests are briefly described below by domain:

Executive. The Penn Conditional Exclusion Test (PCET; Kurtz, Wexler, & Bell, 2004) measures the executive functions of abstraction and mental flexibility (ABF), critical for effective problem solving. It assesses the ability to derive principles and concepts from feedback, as well as the ability to detect and adjust to changing rules. The PCET uses the "odd man out" paradigm, in which participants must determine which object in a group does not belong. The exclusion rule can be based on the shape or configuration of the objects (e.g., a square would not fit in with three stars), the size of the objects, or the thickness of the lines outlining the objects. The participant is given feedback ("correct" or "incorrect") after each response, and the test-administration program automatically changes the exclusion rule after 10 consecutively correct responses (without informing the participant). The participant must then use the feedback to determine what the new exclusion rule is, and after 10 consecutively correct responses, the rule is changed again. The test is scored based on demonstrated learning (proportion of correct responses multiplied by the number of learned rules; 1 is added to accommodate participants who were unable to discover any rule).

The Penn Continuous Performance Test (PCPT) measures vigilance and visual attention (ATT) independent of working memory or perceptual factors. Vertical and horizontal lines in 7-segment displays appear on the screen (at a rate of one second each), and the participant must press the spacebar when the lines are configured as complete numbers (first half of task) or complete letters (second half of task). Each half lasts 1.5 minutes, and during each 1-s response window, the stimulus is presented for only 300 milliseconds (leaving 700 milliseconds of blank screen).

The Penn Letter N-Back Test measures working memory (WM), the ability to keep and refresh goal-related information. Participants attend to a continual series of letters that flash on the screen

Table 1
Domain-Specific Scales of the Computerized Neurocognitive Battery, by Neurobehavioral Function

Neurobehavioral function	Domain	Test
	Mental Flexibility (ABF)	Penn Conditional Exclusion Test (PCET)
Executive Control	Attention (ATT)	Penn Continuous Performance Test (PCPT)
	Working Memory (WM)	Letter N-Back (LNB) task
	Verbal Memory (VME)	Penn Word Memory task (PWMT)
Episodic Memory	Face Memory (FME)	Penn Face Memory task (PFMT)
	Spatial Memory (SME)	Visual Object Learning Test (VOLT)
	Language Reasoning (LAN)	Penn Verbal Reasoning Test (PVRT)
Complex Cognition	Nonverbal Reasoning (NVR)	Penn Matrix Reasoning Test (PMRT)
	Spatial Ability (SPA)	Penn Line Orientation Test (PLOT)
	Emotion Identification (EMI)	Penn Emotion Identification Test (PEIT)
Social Cognition	Emotion Differentiation (EMD)	Penn Emotion Differentiation Test (PEDT)
	Age Differentiation (AGD)	Penn Age Differentiation Test (PADT)

(one at a time) and press the spacebar according to three different rules (called the 0-back, 1-back, and 2-back). During the 0-back condition, the participant must simply respond to a currently present target ("X"). During the 1-back condition, he or she must press the spacebar when the letter on the screen is the same as the previous letter. During the 2-back condition, he or she must press the spacebar when the letter on the screen is the same as the letter before the previous letter (i.e., 2 letters back). In all trials, the interstimulus interval (ISI) is 2.5 seconds, and the stimuli (letters) themselves are presented for 0.5 seconds each. The participant practices all three principles before testing.

Episodic memory. The Penn Word Memory Test (PWMT) measures episodic memory for verbal material (VMEM). In the first part of this test, participants are shown 20 words (for one second each) that they will be asked to identify later. During the recognition phase, participants are shown a series (one at a time) of 40 words, 20 of which are the stimuli they were asked to memorize, and the other 20 are distractors (matched for length, imageability, and concreteness). For each word, the participant must decide whether he or she has seen the word in the memorization phase on a four-choice scale (*definitely not*, *probably not*, *probably yes*, or *definitely yes*).

The Penn Facial Memory Test (PFMT) measures episodic memory for faces (FMEM). The task is identical to the Penn Word Memory Test (above), except that the participant is asked to memorize faces instead of words. PFMT distractor faces are matched for age, ethnicity, and gender.

The Visual Object Learning Test (VOLT) measures episodic memory for shapes (SMEM). The task is nearly identical to the PWMT and PFMT (above), except that the participant is asked to memorize 10 Euclidean shapes instead of 20 words or faces.

Complex cognition. The Penn Verbal Reasoning Test (PVRT) measures language-mediated complex cognition ability (LAN). The task involves a series of analogy problems patterned after Educational Testing Service factor-referenced test kit.

The Penn Matrix Reasoning Task (PMRT) measures nonverbal reasoning ability (NVR) using matrix reasoning problems as used in the Raven's Progressive Matrices Test (Raven, 1989, 2000; Raven, Raven, & Court, 2000) and the Matrix Reasoning subscale of the WAIS-III (Raven, Raven, & Court, 2003).

The Penn Line Orientation Test (PLOT) measures the complex reasoning domain of spatial ability (SPA). The participant is shown two lines on the computer screen that differ in length and orientation, and must press a button to rotate one of the lines until its orientation (angle relative to a horizontal line) is the same as the other (nonrotating) line.

Social cognition. The Penn Emotion Identification Test (EMI) measures the social cognition domain of emotion identification—specifically, the ability to decode and correctly identify facial expressions of emotion. Participants are shown 40 faces (one at a time), and must determine whether the emotion expressed by the actor's face is happiness, sadness, anger, fear, or none at all. There are four female four male faces for each emotion ($4 \times 2 \times 5 = 40$).

The Penn Emotion Differentiation Test (EMD) measures the social cognition domain of emotion intensity differentiation—the ability to decode the intensity of facial expressions of emotion. Participants are shown two faces at a time, both expressing the same emotion, and must determine which of the two faces expresses the emotion more intensely. Differential intensity was obtained by morphing a neutral face to one of four emotions (happy, sad, anger, fear).

The Penn Age Differentiation Test (AGD) measures the social cognition domain of the ability to decode the age of a face. Participants are shown two faces at a time, both neutral, and must determine which of the two faces is older. The stimuli were constructed from young faces morphed into old faces, providing graded levels of difficulty.

Data Analysis

Accuracy and speed values were recorded and entered into a relational database (Oracle), Raw accuracy and speed values were transformed to their standard equivalents (*z*-scores) based on means and *SD*s for the entire sample (*z*-scores for median response time were multiplied by -1 to produce Speed values where, as for accuracy, higher scores reflect better performance). Unstandardized descriptive statistics are shown in Table 2. All confirmatory and exploratory factor analyses were performed with Mplus (Version 7; Muthén & Muthén, 2012) using maximum likelihood (ML) estimation. Because all variables in the present study were continuous *z*-scores that departed from normality only minimally, the robust ML estimator (MLR) was not used. The specific scores being analyzed reflect efficiency, which is the sum of an individual's standardized accuracy and speed scores. For example, if an individual had an accuracy score of 2.50 (very accurate) and a speed score of -2.50 (very slow), his or her efficiency score would be 0.

Results

Confirmatory Factor Analysis

Methodologically, the most rigorous step in evaluating the latent structure of a battery is to estimate a confirmatory model based on theory. Researchers often base confirmatory analyses

Table 2
Descriptive Statistics for Accuracy and Speed on the 12 Penn CNB Tests

	Accuracy (total correct)			Median RT (ms)		
Test	Median	Mean	*SD*	Median	Mean	*SD*
AGD	43	41.68	7.16	2300	2423	700
FME	31	30.50	4.31	1904	1995	480
EMI	34	33.25	3.40	1882	1950	393
VME	38	36.82	3.34	1444	1510	323
LAN	11	10.72	3.04	5648	6197	2409
EMD	45	43.26	7.04	2911	3012	778
NVR	11	11.96	4.65	5776	6907	3823
SME	16	15.54	2.44	1736	1807	426
WM	18	17.54	2.61	517	548	140
ABF	2.04	1.91	0.73	2227	2409	929
ATT	53	50.99	8.24	483	494	68
SPA	9	9.35	4.37	8808	9381	3369

Note. Median accuracy score for ABF was calculated as [(# of categories learned) * (% correct responses)]; *SD* = standard deviation; RT = response time.

purely on their own exploratory analyses; however, given that the CNB was designed explicitly to measure four neurobehavioral domains, we began by testing a theory-based, 4-factor confirmatory model. Note that in contemporary psychometric studies, it is unusual to present a confirmatory analysis before presenting any exploratory analyses. In this case, however, the hypothesized structure of the battery is based on a very specific neuropsychological theory. Thus, the goal of the CFA presented in this section is truly to confirm previously articulated claims about brain networks related to specific domains of cognitive performance. That is, the order in which an investigator presents confirmatory and exploratory analyses is a matter of preference, and depends on what he or she is trying to "confirm." Presenting confirmatory analyses after exploratory analyses implies that one is trying to confirm the exploratory analyses themselves; whereas, presenting confirmatory analyses first implies one is trying to confirm an a priori theory. Here, we take the latter approach, because the hypothesized CNB structure is based in neuropsychological theory.

Figure 1 shows the standardized results of the 4-factor confirmatory model where factors are free to correlate—a "correlated factors model." The fit of the model is acceptable by conventional standards (Hu & Bentler, 1999). The comparative fit index (CFI) is 0.95, the root mean square error of approximation (RMSEA) is 0.055 ± 0.0002, and the standardized root mean residual (SRMR) is 0.030. Fit indices for this model and all subsequent models presented can be found in Table 3. Additionally, all loadings are moderate or strong and in the hypothesized direction. The interfactor correlations are also strong, suggesting an underlying general factor that influences all responses. Indeed, the high correlation (0.94) between executive functioning and complex cognition suggests that perhaps only three factors are necessary to explain the correlations among the CNB scores. This is because 88% (0.94^2) of executive functioning is explained by complex cognition (and vice versa), suggesting they are mostly the same construct. Therefore, we sought to explore the structure further.

Preparation for Exploratory Analyses: How Many Factors?

Before exploring the latent structure of a battery, one must estimate the number of factors necessary to explain a satisfactory amount of the covariance among the items (or tests, in this

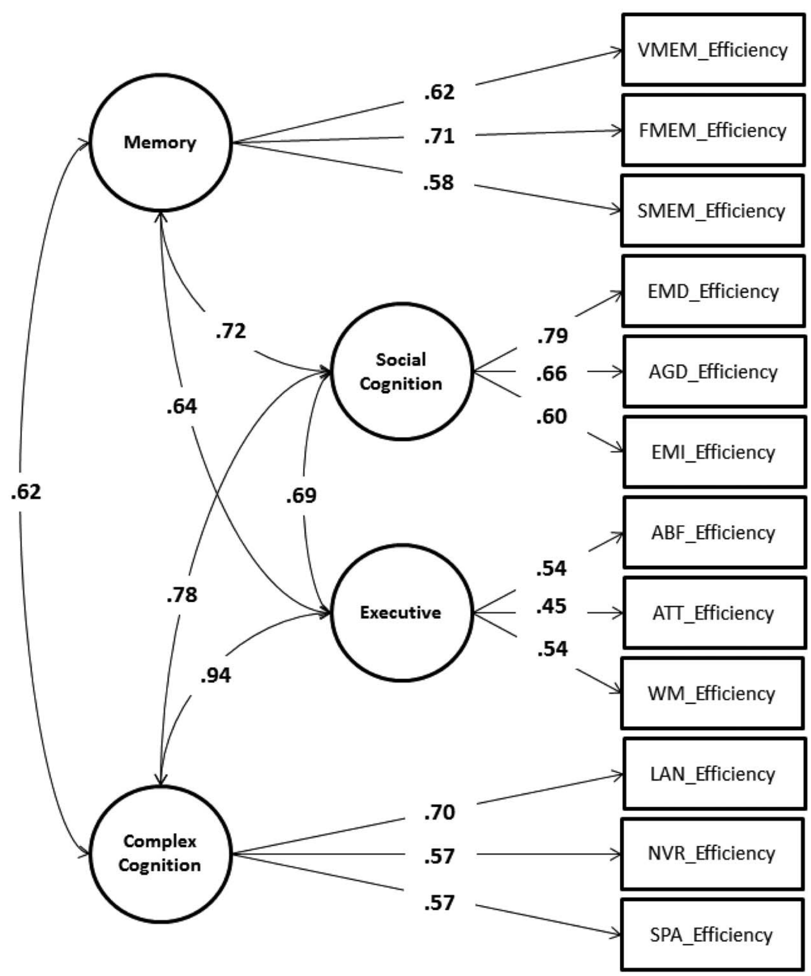

Figure 1. Confirmatory correlated-traits model of the CNB Efficiency scores.

Table 3
Fit Indices of Confirmatory and Exploratory Factor Models of Efficiency, Accuracy, and Speed

Analysis type	Measure	Factors	Fit indices		
			CFI	RMSEA	SRMR
CFA	Efficiency	4	.95	.055	.030
EFA	Efficiency	1	.86	.087	.050
EFA	Efficiency	2	.93	.070	.033
EFA	Efficiency	3	.97	.049	.021
EFA	Efficiency	4	.99	.036	.012
EFA	Accuracy	1	.87	.075	.045
EFA	Accuracy	2	.94	.055	.031
EFA	Accuracy	3	.98	.034	.016
EFA	Accuracy	4	.99	.028	.011
EFA	Speed	1	.82	.098	.061
EFA	Speed	2	.89	.085	.039
EFA	Speed	3	.95	.059	.024
EFA	Speed	4	.99	.040	.015
Bifactor CFA	Efficiency	4	.96	.054	.029

Note. All models are correlated traits models, except the bifactor CFA model of Efficiency; EFA = exploratory factor analysis; CFA = confirmatory factor analysis; CFI = comparative fit index; RMSEA = root mean-square error of approximation; SRMR = standardized root mean-square residual.

case). In the CFA described above, it was not necessary to use any empirical methods to determine the number of factors, because that number (four) was determined by theory.

If the main goal is to explore the data, several numeric methods exist for estimating the optimal number of factors (see Hoyle & Duvall, 2004). The method often regarded as most rigorous is parallel analysis (Horn, 1965; Peres-Neto, Jackson, & Somers, 2005), which compares the progressive eigenvalues of the data being analyzed to the eigenvalues of randomly generated data of the same dimensions. The rationale is that one should not extract factors beyond the point of explaining variance that could be explained by random chance. Figure 2 shows the parallel analysis results (in a scree plot) for the CNB efficiency data using both principal components and maximum likelihood factor extraction. As can be appreciated from Figure 2, the first factor explains a relatively large amount of the covariance, indicating that it might be reasonable to treat all 12 scales as belonging to a single dimension. However, the parallel analysis suggests that four factors are necessary to account for a sufficient amount of covariance, which is indicated by the fact that four factor analysis eigenvalues (triangles) lie above the lower dotted line.

Another common method used to determine the appropriate number of factors is to estimate a full range of solutions (e.g., one–four factors) and examine them for feasibility and interpretability. In practice, this is probably the most common method for selecting an exploratory factor solution, because estimating one additional model (changing only the number of factors) is easy and fast with current software. Also, it can often be useful to see multiple solutions; indeed, when evaluating the structure of a battery, it would probably be unwise to inspect only one exploratory solution.

Exploratory Factor Analysis of Efficiency Scores

Despite the acceptable fit of the confirmatory model above, it is possible that exploratory analysis will reveal further subtleties about the structure of the CNB. Table 4 shows the unidimensional, 2-, 3-, and 4-factor exploratory solutions of the CNB Efficiency scores using an oblique rotation (direct oblimin with default $\delta = 0$). A common way to evaluate factor solutions is to examine their fit indices, and, with the exception of the unidimensional solution, all solutions in Table 4 are at least bordering on acceptable fit. Specifically, the CFI's of the 1- through 4-factor solutions were 0.86, 0.93, 0.97, and 0.99, respectively; their RMSEAs were 0.087, 0.070, 0.049, and 0.036, respectively; and their SRMRs were 0.050, 0.033, 0.021, and 0.012, respectively.

The poor fit of the unidimensional solution suggests that the parameter estimates should be interpreted with caution. Inspection of the modification indices (Sorbom, 1989) reveals several correlated residuals, which is what one expects if there are additional factors that need to be modeled. Such correlated residuals indicate that the unidimensional loadings are inflated. Note that modification indices are normally used in confirmatory (not exploratory) factor analysis; however, because the unidimensional CFA is identical to the unidimensional EFA, one can obtain modification indices in this special case of EFA. With the above cautions in mind, however, one should at least note the two largest and two smallest loadings—specifically, Emotion Differentiation and Language had the largest (0.71 and 0.64, respectively), and Attention and Working Memory had the smallest (0.40 and 0.48, respectively).

When two correlated factors are extracted, the Executive measures (ABF, ATT, and WM) and Complex Cognition (LAN, NVR, and SPA) tests remain together as Factor 1, and the Episodic Memory tests (VMEM, FMEM, and SMEM) "break away" to form their own factor. Notably, without exception, all three Social Cognition tests (EMI, EMD, and AGD) cross-load on both factors. Note also that EMD favors the Executive/ComplexCog factor, whereas EMI prefers the Memory factor. The 3-factor solution is very similar to the 2-factor solution, except that the three Social Cognition tests break away to form their own factor. Note, however, that EMI still prefers the Memory factor, which is consistent with the finding that memory impairment in clinical populations is associated with emotion-recognition deficits (Gur et al., 2006; Hargrave, Maddock, & Stone, 2002; Kohler et al., 2005).

Finally, when four factors are extracted, the solution almost perfectly matches the theory used to construct the Penn CNB. The only clear exception is the moderate loading of Abstraction and Mental Flexibility (ABF) on the Complex Cognition factor rather than its intended factor (Executive Functioning). It appears that the Penn Conditional Exclusion Test, which measures ABF, requires complex cognition in addition to executive control. It also allows the test taker more time to contemplate the answer than the CPT (ATT) or Letter-N-Back (WM). A second notable cross-loading is the 0.28 loading of LAN on Factor 1 (Executive). This finding is consistent with evidence implicating working memory and attention in reading (Casco, Tressoldi, & Dellantonio, 1998; Daneman & Carpenter, 1980; Vidyasagar, 2004). Finally, the cross-loading of SMEM on Factor 3 (Com-

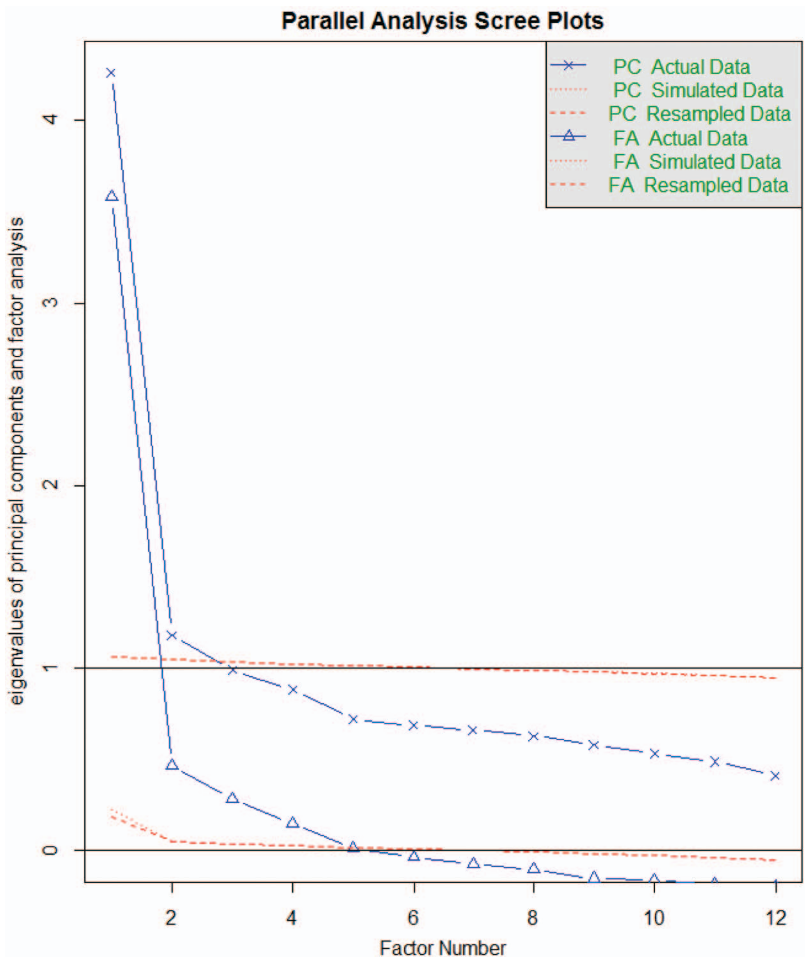

Figure 2. Parallel analysis scree plots of 12-variable CNB. See the online article for a color version of this figure.

plex Cognition) is consistent with the established association between spatial memory and IQ (Passolunghi & Lanfranchi, 2012; Shang & Gau, 2011). The general conformity of the 4-factor solution to the theory used to design the CNB adds support to the confirmatory results reported above. Especially important is the finding that, despite the very large correlation between the Complex Cognition and Executive Functioning factors (see Figure 1), they do split into two factors (although imperfectly) when explored using oblimin rotation.

Exploratory Factor Analysis of Accuracy and Speed Scores

The analyses reported above combined accuracy and speed, to form measures of "efficiency." Such measures are most comparable to previous studies using traditional batteries, where accuracy and speed are confounded. It could be informative, however, to examine the factor structure of the CNB accuracy and speed measures separately to answer two important questions. First, do the accuracy and speed scores produce factor patterns that are interpretable enough to warrant separate accuracy and speed scales capable of predicting outcomes above and beyond the efficiency scores? This would suggest a need for further investigation. Second, do the patterns of accuracy and speed scores match those of the efficiency scores, that is, the structures shown in Table 4 follow inevitably from combining accuracy and speed, or do they differ, that is, the structures in Table 4 are the result of speed and accuracy interacting differently for each test? This issue is related to the question of how accuracy and speed interact with each other. For example, accuracy and speed might be negatively correlated, implying a speed–accuracy trade-off, or their relationship might depend on the type of test, in which case it would be unlikely that the structures of speed or accuracy alone will mimic the structure of efficiency. We do not directly address the question of how accuracy and speed relate, especially because such an investigation would require analysis within individuals. However, the results provide some hints.

Table 5 shows the unidimensional, 2-, 3-, and 4-factor exploratory solutions of the CNB Accuracy scores using an oblique rotation (direct oblimin with default $\delta = 0$). As in the case of the Efficiency scores (see Table 4), one should first examine the fit indices of the four solutions. With the exception of the unidimensional model, all solutions in Table 5 fit at least moderately well. Specifically, the CFI's of the 1- through 4-factor solutions were 0.87, 0.94, 0.98, and 0.99, respectively; their RMSEAs were 0.075,

Table 4
Unidimensional, 2-, 3-, and 4-Factor Exploratory Solutions of the CNB Efficiency Scores (With Oblique Rotation)

		2-Factor		3-Factor			4-Factor			
Test	Uni	F1	F2	F1	F2	F3	F1	F2	F3	F4
ABF	.48	.53		.48					.49	
ATT	.40	.36		.40			.46			
WM	.46	.52		.58			.58			
VMEM	.52		.51		.56			.53		
FMEM	.57		.80		.75			.70		
SMEM	.48		.48		.49			.53	.25	
LAN	.64	.72		.63			.28		.37	
NVR	.52	.59		.47					.51	
SPA	.53	.56		.47					.50	
EMI	.59	.24	.43		.32	.29		.28		.37
EMD	.71	.52	.25			.71				.79
AGD	.59	.30	.34			.69				.66
		Factor correlations (phi matrices)								
		F1	F2	F1	F2	F3	F1	F2	F3	F4
	F1	—		—			—			
	F2	.61	—	.51	—		.36	—		
	F3			.61	.58	—	.55	.42	—	
	F4						.44	.57	.60	—

Note. Uni = unidimensional; F = factor; rotation = oblimin; loadings < .20 removed.

0.055, 0.034, and 0.028, respectively; and their SRMR's were 0.045, 0.031, 0.016, and 0.011, respectively. Because of the poor fit of the unidimensional model, its loadings should be interpreted with caution as likely to be inflated. Nonetheless, it is worth noting that the two largest loadings are for Language (0.68) and Nonverbal Reasoning (0.65), and the two smallest are for Emotion Identification (0.31) and Attention (0.34).

When two correlated factors are extracted, the social cognition variables (EMI, EMD, and AGD) clearly form their own factor with no cross-loadings over 0.20. The 0.60 correlation between the factors, however, indicates that a strong general factor ("accuracy") underlies all scale scores. Note also that EMI loads substantially weaker on F2 than the other two social cognition variables that define F2.

When three correlated factors are specified, the memory measures (VMEM, FMEM, and SMEM) form their own factor (F2), the social cognition scales remain together (F3), and F1 is defined by the remaining six tests. These remaining six tests defining F1 were designed to measure the neurobehavioral functions of executive control (ABF, ATT, and WM scales) and complex cognition (LAN, NVR, and SPA tests). This suggests that, of the four neurobehavioral functions, the two that are most similar to each other are executive control and complex cognition. Only two cross-loadings were >0.20: EMI loads 0.25 on the memory factor (F2), and SMEM loads 0.28 on the executive/cognition factor (F1). Finally, note that the correlations among the three factors remain moderate-to-high, suggesting a general factor underlying all test scores.

The 4-factor solution shown in Table 4 is less interpretable than the other three solutions, because F4 does not have any strong loadings and does not correlate highly with any of the other factors. It also remains nearly identical to the 3-factor solution, which suggests that the 3-factor solution is probably optimal. We evaluated that possibility using post hoc confirmatory analyses.

We estimated the same four models as above using the CNB Speed scores. Table 6 shows the unidimensional, 2-, 3-, and 4-factor explor-

Table 5
Unidimensional, 2-, 3-, and 4-Factor Exploratory Solutions of The CNB Accuracy Scores (With Oblique Rotation)

		2-Factor		3-Factor			4-Factor			
Test	Uni	F1	F2	F1	F2	F3	F1	F2	F3	F4
ABF	.45	.48		.47			.50			
ATT	.34	.37		.34			.39			
WM	.51	.56		.50			.53			
VMEM	.38	.36			.45			.48		
FMEM	.50	.31	.23		.65			.58		
SMEM	.45	.47		.28	.37			.53		−.21
LAN	.68	.68		.66			.70			
NVR	.65	.67		.71			.64			
SPA	.60	.61		.60			.51			
EMI	.31		.25		.24	.21				.27
EMD	.59		.64			.66			.64	
AGD	.52		.78			.74			.77	
		Factor correlations (phi matrices)								
		F1	F2	F1	F2	F3	F1	F2	F3	F4
	F1	—		—			—			
	F2	.63	—	.58	—		.61	—		
	F3			.59	.51	—	.58	.49	—	
	F4						−.12	.02	.08	—

Note. Uni = unidimensional; F = factor; rotation = oblimin; loadings < .20 removed.

Table 6
Unidimensional, 2-, 3-, and 4-Factor Exploratory Solutions of The CNB Speed Scores (With Oblique Rotation)

		2-Factor		3-Factor			4-Factor			
Test	Uni	F1	F2	F1	F2	F3	F1	F2	F3	F4
ABF	.47	.47				.41			.41	
ATT	.30		.61	.59			.82			
WM	.26		.54	.57			.49			
VMEM	.67	.56	.22		.74			.77		
FMEM	.71	.68			.77			.68		
SMEM	.65	.66			.66			.69		−.27
LAN	.50	.44		.20		.46			.48	.25
NVR	.33	.43				.47			.46	
SPA	.53	.55				.49			.49	
EMI	.64	.54	.21	.22	.28	.32		.26	.37	.33
EMD	.68	.71				.79			.78	
AGD	.68	.75				.65			.66	
MOT	.25		.39	.35			.26			
		Factor correlations (phi matrices)								
		F1	F2	F1	F2	F3	F1	F2	F3	F4
	F1	—		—			—			
	F2	.38	—	.32	—		.33	—		
	F3			.27	.69	—	.28	.66	—	
	F4						.18	.07	.05	—

Note. Uni = unidimensional; F = factor; rotation = oblimin; loadings < .20 removed.

atory solutions of the CNB Speed scores using direct oblimin rotation with default $\delta = 0$. As in the case of the Efficiency and Accuracy scores, the unidimensional model (CFI = 0.82; RMSEA = 0.098; SRMR = 0.061) is insufficient to explain the relationships among the Speed scores, and its loadings should therefore be interpreted with caution. It is noteworthy, however, that the highest and lowest loadings in the unidimensional Speed model do not come close to matching those in the Accuracy model. Whereas LAN and NVR were the best indicators of how accurate participants would be overall, FMEM, EMD, and AGD were the best indicators of how fast they would be (loadings = 0.71, 0.68, and 0.68, respectively). Further, whereas ATT and EMI were the worst indicators of how accurate participants would be overall, WM and NVR were the worst indicators of how fast they would be (loadings = 0.26 and 0.33, respectively).

When two Speed factors are extracted, the fit (CFI = 0.89; RMSEA = 0.085; SRMR = 0.039) begins to approach acceptable levels (using liberal criteria). It also makes intuitive sense: the tasks requiring constant vigilance (ATT, WM, and MOT) form their own factor, while all other tasks remain together. Note also that the correlation between the two speed factors (0.38) is much lower than when two accuracy factors are extracted (0.63; see Table 5), indicating that the two speed-related processes are relatively separate.

When three Speed factors are extracted, the fit (CFI = 0.95; RMSEA = 0.059; SRMR = 0.024) becomes acceptable by typical standards. The structure remains mostly the same as in the 2-factor solution, except that the three Memory tasks break away from the non-high-vigilance tasks to form their own factor. This could indicate that the speed with which participants are willing to commit to a memory-related response involves a process that is somewhat unique from the process involved in committing to other low-vigilance tasks (such as Emotion Differentiation). Note, however, that Factors 2 and 3 are highly correlated (0.69), indicating that, the difference is not substantial.

Finally, the 4-factor solution in Table 6 has the same combination of excellent fit yet poor interpretability as the 4-factor model of Accuracy (see Table 5). That is, despite the excellent fit of the model (CFI = 0.99; RMSEA = 0.040; SRMR = 0.015), the fourth factor contributes very little: it is weak (max loading = 0.33) and is not correlated to the other three factors (max intercorrelation = 0.18). Thus, interpretation of the 4-factor solution is probably not worthwhile beyond pointing out that the three factors from the 3-factor solution remain strong in the 4-factor solution.

Post Hoc Confirmatory Bifactor Analysis

We began our investigation of the CNB structure with a theory-based confirmatory model of Efficiency scores (see Figure 1) in which the four neurobehavioral domain factors were allowed to correlate. Such a model is appropriate for testing the theoretical basis of the battery, but if one wishes to use the measurement model within a larger structural equation model—for example, a model in which the latent CNB factors predict an external outcome—the correlations among the factors can become problematic. Thus, in an effort to provide future investigators with a confirmatory model that does not come with the drawback of interfactor correlations, we sought a model that controls for these correlations, allowing the subfactors to be orthogonal. This can be accomplished by modeling a general (overall) factor along with the neurobehavioral function subfactors.

A convenient way to model a general factor and the individual neurobehavioral function factors is to use a bifactor model (Holzinger & Swineford, 1937; also see Reise, 2012; Reise, Moore, & Haviland, 2010). In a typical bifactor model, each variable (in this case, test score) loads on two factors, one general and one specific. All factors are orthogonal, such that the specific factors (in this case, neurobehavioral domains) explain only the covariance not explained by the general factor. Likewise, the general factor (overall performance) explains the covariance among all 12 test scores independent from the covariance explained by the specific neurobehavioral domains. That the specific factors are orthogonal after controlling for the general factor is a key strength of the bifactor model, because it allows one to test the relationships between the general factors and important external variables (e.g., gray matter density of a specific brain region of interest) without the "contamination" caused by general performance. Obtaining such an independent measure of a specific type of neurocognition is crucial in testing the validity of a battery.

The exploratory analyses of Efficiency scores indicated that one test (ABF) does not load on the factor for which it was intended (Executive Functioning), but rather loads with the three Complex Cognition tests. Given this discovery, we can build a better-fitting confirmatory model (post hoc) that we can later use to test the structural validity of the CNB. Such an approach might seem puzzling, given the good fit of the model in Figure 1, in which ABF loads on Executive Functioning, as theoretically intended. One might ask, why not keep the assignment of tests to neurobehavioral functions the same in the bifactor model as they were in the original correlated-traits model? This is a legitimate question, and if the exploratory analysis (see Table 4) left any doubt as to where ABF should load, the best approach would probably be to assign tests to neurobehavioral functions as indicated in Figure 1. However, the exploratory analyses unambiguously indicate that ABF should load on Complex Cognition. Indeed, when a bifactor model is specified in which ABF loads on Executive Functioning, Mplus encounters computational problems that can be fixed only by placing constraints on the model.

Figure 3 shows the 4-factor bifactor model of the CNB Efficiency scores. The fit is acceptable (CFI = 0.96; RMSEA = 0.054 ± 0.002; SRMR = 0.029), and comparisons of the information criteria to those of the correlated traits model (see Figure 1) favor the bifactor model. Specifically, the Akaike Information Criterion for the correlated traits and bifactor models are 362928 and 362745, respectively, and the Bayesian Information Criteria are 363228 and 363080, respectively.

The overall conclusions to be drawn from these results are very similar to those suggested by the 4-factor correlated traits model. Specifically, a strong general factor (Efficiency or Performance) explains most of the covariance among all 12 scales, and the covariance not explained by that factor is explained by four neurobehavioral function factors. Here, however, the Abstraction and Mental Flexibility test (ABF) is now loading on the Complex Cognition factor rather than the Executive Control factor. Also, note that because an orthogonal factor cannot be identified with only two indicators, the loadings of Attention and Working Memory on the Executive Control factor had to be constrained to equality.

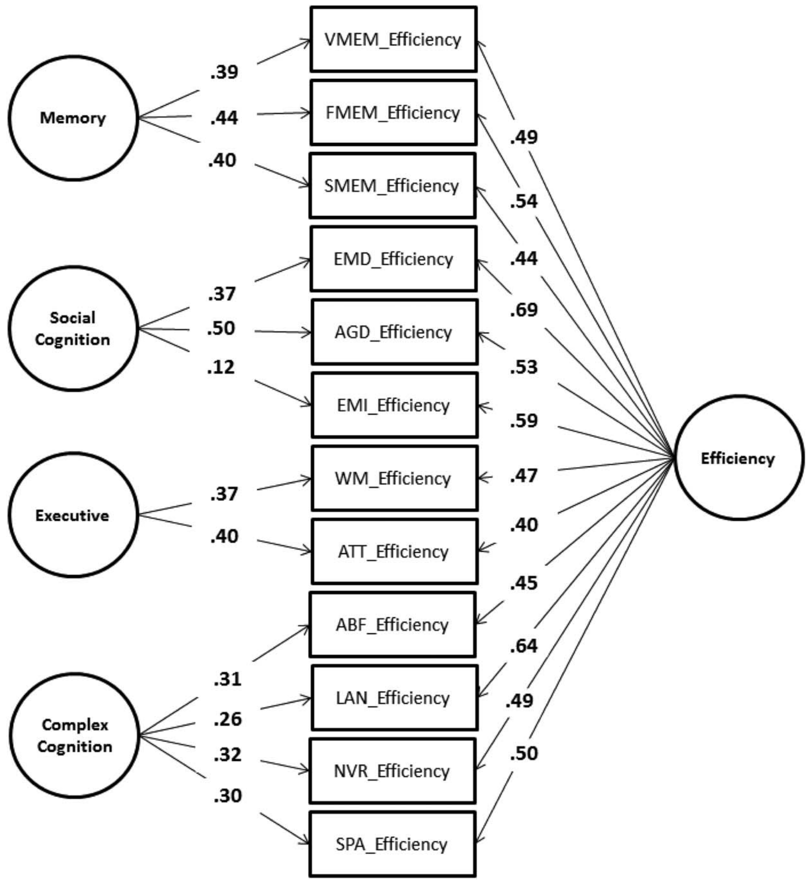

Figure 3. Confirmatory bifactor model of the CNB Efficiency scores.

The bifactor model provides some unique insights into the CNB structure. First, Social Cognition factor in Figure 1 is dominated by EMD (loading = 0.79), whereas in the bifactor model (see Figure 3) it is dominated by AGD (loading = 0.50). This is because much of the variance explained by EMD in the correlated-traits model shifted to the general factor in the bifactor model, suggesting it is a strong indicator of the overall trait measured by the CNB (cognitive performance). A second insight provided by the bifactor model pertains to the general factor itself—indeed, the general factor loadings are excellent indicators of which tests are the "best overall." Using this standard, Figure 3 suggests that the language reasoning and emotion differentiation scores are the best measures of general performance, followed closely by Emotion Identification. The worst indicator of general performance is Attention.

Finally, an important strength of factor modeling (especially bifactor modeling) is that it allows for informed calculation of subscale scores. Rather than relying exclusively on theory to create subscales, one can use the information gathered from factor analysis to, a) determine which items (or tests) should compose a score, and b) weight the items such that their relative contributions to the score variance appropriately reflect their correlations with the theoretical factor itself. The bifactor model is especially useful for calculating subscale scores, because its factors are orthogonal, that is, an investigator could use them simultaneously in a predictive model (such as regression). The problem, however, is that factor scores must be well-determined in order to acquire the properties (e.g., orthogonality) of the factors themselves (see Grice, 2001), which is why indices of factor score determinacy are important.

The most common way to measure the determinacy of a factor score is to calculate the correlation (ρ) between that score and the theoretical factor used to calculate it. (See Grice, 2001, for an explanation of why this correlation is not 1.00.) What is considered an "acceptable" value of ρ is subjective, but Gorsuch (1983, p. 260) recommends 0.80 as the minimum cutoff. By this standard, the general Efficiency bifactor score (calculated from the model in Figure 3) is usable (ρ = 0.89), but the Complex Cognition, Executive, Social, and Memory subscales are not (ρ = 0.54, 0.55, 0.62, 0.63, respectively). We therefore recommend that if an investigator wishes to use one overall score for the NB, he or she should use the general factor from the bifactor model, but if he or she wishes to calculate subscales, the correlated-traits model is preferable.

Discussion

Traditional neuropsychological tests are not feasible in the context of large-scale genomic, epidemiologic, and treatment studies in basic and clinical neuroscience where massive data are collected electronically and entered into common databases for integration.

Designers of such studies, even if they care deeply about measuring neuropsychological constructs, would not be able to contemplate the inclusion of any nonelectronic component, let alone testing that takes place over hours and requires experts for scoring and interpretation. Thus, while obviously behavior is the product of brain function, and arguably the ultimate target of clinical neuroscience, there is a risk that neuropsychology will be left out of the genomic revolution that is now sweeping medical research. The web-based CNB, on the other hand, fits well in such studies because it is brief, electronic, can be administered across platforms on any web-enabled device, requires minimal training for administration, scores results "on the fly" and can feed data directly into an electronic data repository. Furthermore, unlike traditional neuropsychological tests where only the summary scores can be included in a database, computerized tests can be evaluated for quality assurance after acquisition and probed for item-wise effects on either accuracy or speed. These features have resulted in rapid and wide adoption of the CNB in large-scale population-based and clinical studies as well as military research (Aliyu et al., 2006; Almasy et al., 2008; Grant et al., 2012; Greenwood et al., 2007, 2011; Gur et al., 2007, 2012; Thomas et al., 2013).

Despite the widespread use of the CNB, its latent structure had not been thoroughly examined. The results of the present large-scale study support the theoretical approach to the CNB design, and offer some insights into uses and interpretations of CNB scores. The rationale used to develop the CNB is strongly supported, with two possible exceptions. Specifically, one of the tests (the Penn Conditional Exclusion Test) appears to be a better measure of complex cognition than of what it was intended to measure, executive functioning. Furthermore, two of the four neurobehavioral domains thought to compose overall cognitive performance are highly correlated and might not reflect separate abilities. This latter point might lead one to question whether a 4-factor model is necessary over a 3-factor model—that is, one can reject the unidimensional model based on poor fit, but the acceptable fit of the 3-factor model necessitates some further explanation as to why it is not preferred over the 4-factor solution. Overall, the exceptions were neither surprising nor indicative of poor measurement. They may not necessitate post hoc theory adjustment, but instead lead to practical considerations, especially with regard to scoring subscales. In our case, abstraction and mental flexibility were hypothesized to tap the functioning of frontal lobe systems, and hence we expected performance to correlate with the other tests of executive functioning. However, the complexity of the test seems to require greater amount of complex reasoning than executive abilities. Perhaps easier abstraction and mental flexibility tests would be better at isolating the executive aspects of this domain.

The high correlation between the Executive Control and Complex Cognition factors (see Figure 1) indicates that a large portion (about 85%) of the variance in one can be predicted from the other. From a measurement perspective, this means there is no good reason to treat them as two separate factors. From a neuropsychological perspective, however, there is reason to expect Executive Control and Complex Cognition to be correlated because it is difficult to solve a complex problem without applying executive functioning and some executive functions require complex reasoning. The high factor intercorrelation may also relate to an overlap in frontoparietal brain systems needed to perform these tasks (Roalf, Ruparel et al., 2013).

The main practical conclusion is that an investigator wishing to calculate subscale scores for each domain of the CNB (rather than one global score) should combine the Executive Control and Complex Cognition domains. That is, even though the exploratory analysis (of efficiency scores) suggests there truly are two separate executive functioning and complex cognition factors, they are so highly correlated in the confirmatory model (see Figure 1) that their two subscale scores would likely correlate to the point of redundancy. Regardless of whether the Executive Control and Complex Cognition domains are truly separate in terms of neural systems (as suggested by current neuropsychological theory), they should not be treated separately when calculating scores unless there is a hypothesis that one and not the other is correlated with a specific measure of brain function.

There are two important exceptions to the above point. First, even though general psychometric practice suggests that two highly correlated latent variables should not be treated separately, exceptions to the phenomenon—for example, a person who scores very high on Executive Control but very low on Complex Cognition—can be highly informative. A useful example is height and weight, which are highly correlated and yet not redundant because a discrepancy between them is an important index of an abnormality. Similarly, cerebral blood flow is highly correlated with metabolism (Paulson, Strandgaard, & Edvinsson, 1990), but disturbances in this autoregulation are important indicators of brain dysfunction (Duffy, Howse, & Plum, 1975; Chapman, Meldrum, & Siesjö, 1977). Therefore, despite the high correlation between them, one can gain valuable information by measuring both.

The second argument in favor of creating subscales involves the bifactor model. One could justifiably separate the battery into four orthogonal factors if and only if the model is an appropriate bifactor model (see Figure 3). Even in that case, however, one could not use the bifactor model to calculate scores (due to factor score indeterminacy; see Grice, 2001). Instead, one would have to put the bifactor measurement model within the context of a structural equation model, assuring that the specific factors within the bifactor remain truly orthogonal. Further, given the correlations among the factors and the ratio of first to second eigenvalues, the CNB can be thought of as measuring one global trait (perhaps general intelligence, also known as "g"). Again, however, if it is absolutely imperative that a researcher obtain four orthogonal latent variables from the CNB, that can be accomplished using a bifactor model with SEM.

Finally, if an investigator does want to calculate subscale scores, he or she should use the correlated-traits model (see Figure 1) rather than the bifactor model. Recall that scores calculated using the bifactor model are too poorly determined to be used as valid measures of the factors they represent. The correlated-traits model, however, produces well-determined factors: ρ for Complex Reasoning, Executive, Social, and Memory are 0.89, 0.87, 0.90, and 0.87, respectively, which are beyond the recommended cutoff of 0.80 (Gorsuch, 1983). The drawback of using the correlated-traits scores is that they are correlated, and will therefore introduce nuisance collinearity when used simultaneously in a predictive model (such as regression). Thus, investigators wishing to use CNB performance to predict outcomes (or to be predicted by other external variables) face the following choices:

1. Use the bifactor model (which comprises orthogonal subfactors and a general factor) within the context of a larger SEM model, rather than calculating scores to be used in that model.

2. Use the correlated-traits model to calculate scores, and use those scores in a separate predictive model, with the understanding that the scores will not be orthogonal.

Both options come with specific advantages and disadvantages discussed above. A worthwhile note of caution, also, is that calculation of factor scores is a very complex topic with multiple caveats. Because we present confirmatory models here, we assume scores will be calculated using the factor loadings as regression coefficients; however, this is only one of many methods, and readers are directed to Grice (2001) for review.

Future Directions

The most important step in evaluating the psychometric properties of the CNB is to establish its validity. The excellent structural validity demonstrated here is important—indeed, it is necessary, because test scores cannot be valid unless they are reliable—but it is perhaps even more important to demonstrate that the CNB measures what it's supposed to measure. By "validity," in this case we mean *true* validity (see Borsboom, 2005)—that is, an instrument is valid if and only if it measures what it is supposed to measure.

Often, the validity of an instrument is assessed by correlating its scores with several criteria (external variables), with the hope that it predicts the behaviors/outcomes it was designed to predict. Neurobehavioral measures like the CNB, however, are designed to probe neurobiological processes. Thus, a core strength of the CNB is that it was built with the aim of measuring the integrity of specific brain systems. Therefore, its true validity can be directly tested with functional neuroimaging (see Roalf, Ruparel et al., 2013).

Another convenient development in methodology that will help with the exploration of validity is the recent adoption by many researchers of the bifactor model itself (See Reise, 2012). A strength of the bifactor model that was not exploited here is that it can be used within a structural equation model to test the relationships among tests and neurobehavioral functions independent of the general factor (Efficiency in this case; see Chen, West, & Sousa, 2006, for an example using quality-of-life variables).

References

Aliyu, M. H., Calkins, M. E., Swanson Jr., C. L., Lyons, P. D., Savage, R. M., May, R., . . . Go, R. C. (2006). Project among African-Americans to explore risks for schizophrenia (PAARTNERS): recruitment and assessment methods. *Schizophrenia Research, 87,* 32–44. doi:10.1016/j.schres.2006.06.027

Almasy, L., Gur, R., Haack, K., Cole, S., Calkins, M., Peralta, J., . . . Gur, R. (2008). A genome screen for quantitative trait loci influencing schizophrenia and neurocognitive phenotypes. *The American Journal of Psychiatry, 165,* 1185–1192. doi:10.1176/appi.ajp.2008.07121869

Borsboom, D. (2005). *Measuring the mind: Conceptual issues in contemporary psychometrics*. Cambridge, UK: Cambridge University Press.

Casco, C., Tressoldi, P. E., & Dellantonio, A. (1998). Visual selective attention and reading efficiency are related in children. *Cortex, 34,* 531–546. doi:10.1016/S0010-9452(08)70512-4

Chapman, A. G., Meldrum, B. S., & Siesjö, B. K. (1977). Cerebral metabolic changes during prolonged epileptic seizures in rats. *Journal of Neurochemistry, 28,* 1025–1035. doi:10.1111/j.1471-4159.1977.tb10665.x

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research, 41,* 189–225. doi:10.1207/s15327906mbr4102_5

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450–466. doi:10.1016/S0022-5371(80)90312-6

Duffy, T. E., Howse, D. C., & Plum, F. (1975). Cerebral energy metabolism during experimental status epilepticus. *Journal of Neurochemistry, 24,* 925–934. doi:10.1111/j.1471-4159.1975.tb03657.x

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Grant, P. M., Huh, G. A., Perivoliotis, D., Stolar, N. M., & Beck, A. T. (2012). Randomized trial to evaluate the efficacy of cognitive therapy for low-functioning patients with schizophrenia. *Archives of General Psychiatry, 69,* 121–127. doi:10.1001/archgenpsychiatry.2011.129

Greenwood, T. A., Braff, D. L., Light, G. A., Cadenhead, K. S., Calkins, M. E., Dobie, D. J., . . . Schork, N. J. (2007). Initial heritability analyses of endophenotypic measures for schizophrenia: The consortium on the genetics of schizophrenia. *Archives of General Psychiatry, 64,* 1242–1250. doi:10.1001/archpsyc.64.11.1242

Greenwood, T. A., Lazzeroni, L. C., Murray, S. S., Cadenhead, K. S., Calkins, M. E., Dobie, D. J., . . . Braff, D. L. (2011). Analysis of 94 candidate genes and 12 endophenotypes for schizophrenia from the Consortium on the Genetics of Schizophrenia. *The American Journal of Psychiatry, 168,* 930–946. doi:10.1176/appi.ajp.2011.10050723

Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods, 6,* 430–450. doi:10.1037/1082-989X.6.4.430

Gur, R. C., Erwin, R. J., & Gur, R. E. (1992). Neurobehavioral probes for physiologic neuroimaging studies. *Archives of General Psychiatry, 49,* 409–414. doi:10.1001/archpsyc.1992.01820050073013

Gur, R. E., Kaltman, D., Melhem, E. R., Ruparel, K., Prabhakaran, K., Riley, M., . . . Gur, R. C. (2013). Incidental findings in youths volunteering for brain MRI research. *American Journal of Neuroradiology, 34,* 2021–2025.

Gur, R. E., Kohler, C. G., Ragland, J. D., Siegel, S. J., Lesko, K., Bilker, W. B., & Gur, R. C. (2006). Flat affect in schizophrenia: Relation to emotion processing and neurocognitive measures. *Schizophrenia Bulletin, 32,* 279–287. doi:10.1093/schbul/sbj041

Gur, R. E., Nimgaonkar, V. L., Almasy, L., Calkins, M. E., Ragland, J. D., Pogue-Geile, M. F., . . . Gur, R. C. (2007). Neurocognitive endophenotypes in a multiplex multigenerational family study of schizophrenia. *American Journal of Psychiatry, 164,* 813–819.

Gur, R. C., Ragland, J. D., Moberg, P. J., Turner, T. H., Bilker, W. B., Kohler, C., & Gur, R. E. (2001). Computerized neurocognitive scanning: I. Methodology and validation in healthy people. *Neuropsychopharmacology, 25,* 766–776. doi:10.1016/S0893-133X(01)00278-0

Gur, R. C., Richard, J., Calkins, M. E., Chiavacci, R., Hansen, J. A., Bilker, W. B., . . . Gur, R. E. (2012). Age group and sex differences in performance on a computerized neurocognitive battery in children age 8–21. *Neuropsychology, 26,* 251–265. doi:10.1037/a0026712

Gur, R. C., Richard, J., Hughett, P., Calkins, M. E., Macy, L., Bilker, W. B., & Gur, R. E. (2010). A cognitive neuroscience-based computerized battery for efficient measurement of individual differences: Standardization and initial construct validation. *Journal of Neuroscience Methods, 187,* 254–262. doi:10.1016/j.jneumeth.2009.11.017

Hargrave, R., Maddock, R. J., & Stone, V. (2002). Impaired recognition of facial expressions of emotion in Alzheimer's. *The Journal of Neuropsychiatry and Clinical Neurosciences, 14,* 64–71. doi:10.1176/appi.neuropsych.14.1.64

Holzinger, K. J., & Swineford, F. (1937). The bi-factor method. *Psychometrika, 2,* 41–54.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30,* 179–185. doi:10.1007/BF02289447

Hoyle, R. H., & Duvall, J. L. (2004). Determining the number of factors in exploratory and confirmatory factor analysis. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences* (pp. 301–315). Thousand Oaks, CA: Sage. doi:10.4135/9781412986311.n16

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55. doi:10.1080/10705519909540118

Insel, T. R., & Cuthbert, B. N. (2009). Endophenotypes: Bridging genomic complexity and disorder heterogeneity. *Biological Psychiatry, 66,* 988–989. doi:10.1016/j.biopsych.2009.10.008

Kohler, C. G., Anselmo-Gallagher, G., Bilker, W., Karlwish, J., Gur, R. E., & Clark, C. M. (2005). Emotion-discrimination deficits in mild Alzheimer disease. *The American Journal of Geriatric Psychiatry, 13,* 926–933. doi:10.1097/00019442-200511000-00002

Kurtz, M. M., Wexler, B. E., & Bell, M. D. (2004). The Penn Conditional Exclusion Test (PCET): Relationship to the Wisconsin Card Sorting Test and work function in patients with schizophrenia. *Schizophrenia Research, 68,* 95–102.

Muthén, L. K., & Muthén, B. O. (2012). *Mplus User's Guide*. *Seventh Edition*. Los Angeles, CA: Authors.

Passolunghi, M. C., & Lanfranchi, S. (2012). Domain-specific and domain-general precursors of mathematical achievement: A longitudinal study from kindergarten to first grade. *British Journal of Educational Psychology, 82,* 42–63. doi:10.1111/j.2044-8279.2011.02039.x

Paulson, O. B., Strandgaard, S., & Edvinsson, L. (1990). Cerebral autoregulation. *Cerebrovascular and Brain Metabolism Reviews, 2,* 161–192.

Peres-Neto, P. R., Jackson, D. A., & Somers, K. M. (2005). How many principal components? Stopping rules for determining the number of non-trivial axes revisited. *Computational Statistics & Data Analysis, 49,* 974–997. doi:10.1016/j.csda.2004.06.015

Raven, J. (1989). The Raven Progressive Matrices: A review of national norming studies and ethnic and socioeconomic variation within the United States. *Journal of Educational Measurement, 26,* 1–16. doi:10.1111/j.1745-3984.1989.tb00314.x

Raven, J. (2000). The Raven's Progressive Matrices: Change and stability over culture and time. *Cognitive Psychology, 41,* 1–48. doi:10.1006/cogp.1999.0735

Raven, J., Raven, J. C., & Court, J. H. (2000). *Standard progressive matrices*. East Sussex, UK: Psychology Press.

Raven, J., Raven, J. C., & Court, J. H. (2003). *Manual for Raven's progressive matrices and vocabulary scales. Section 1: General overview* (1998 ed., updated 2003). San Antonio, TX: Pearson Assessment.

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47,* 667–696. doi:10.1080/00273171.2012.715555

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment, 92,* 544–559. doi:10.1080/00223891.2010.496477

Roalf, D. R., Ruparel, K., Verma, R., Elliott, M. A., Gur, R. E., & Gur, R. C. (2013). White matter organization and neurocognitive performance variability in schizophrenia. *Schizophrenia Research, 143,* 172–178. doi:10.1016/j.schres.2012.10.014

Shang, C. Y., & Gau, S. S. (2011). Visual memory as a potential cognitive endophenotype of attention deficit hyperactivity disorder. *Psychological Medicine, 41,* 2603–2614. doi:10.1017/S0033291711000857

Sörbom, D. (1989). Model modification. *Psychometrika, 54,* 371–384.

Thomas, M. L., Brown, G. G., Gur, R. C., Hansen, J. A., Nock, M. K., Heeringa, S., . . . Stein, M. B. (2013). Parallel psychometric and cognitive modeling analyses of the Penn Face Memory Test in the Army Study to Assess Risk and Resilience in Servicemembers. *Journal of Clinical and Experimental Neuropsychology, 35,* 225–245. doi:10.1080/13803395.2012.762974

Vidyasagar, T. R. (2004). Neural underpinnings of dyslexia as a disorder of visuo-spatial attention. *Clinical and Experimental Optometry, 87,* 4–10. doi:10.1111/j.1444-0938.2004.tb03138.x

Received January 8, 2014
Revision received March 4, 2014
Accepted April 3, 2014 ■