Improving clinical cognitive testing
Report of the AAN Behavioral Neurology Section Workgroup

Kirk R. Daffner, MD*
Seth A. Gale, MD*
A.M. Barrett, MD
Bradley F. Boeve, MD
Anjan Chatterjee, MD
H. Branch Coslett, MD
Mark D'Esposito, MD
Glen R. Finney, MD
Darren R. Gitelman, MD
John J. Hart, Jr., MD
Alan J. Lerner, MD
Kimford J. Meador, MD
Alison C. Pietras, BA
Kytja S. Voeller, MD
Daniel I. Kaufer, MD

Correspondence to
Dr. Daffner:
kdaffner@partners.org

ABSTRACT

Objective: To evaluate the evidence basis of single-domain cognitive tests frequently used by behavioral neurologists in an effort to improve the quality of clinical cognitive assessment.

Methods: Behavioral Neurology Section members of the American Academy of Neurology were surveyed about how they conduct clinical cognitive testing, with a particular focus on the Neurobehavioral Status Exam (NBSE). In contrast to general screening cognitive tests, an NBSE consists of tests of individual cognitive domains (e.g., memory or language) that provide a more comprehensive diagnostic assessment. Workgroups for each of 5 cognitive domains (attention, executive function, memory, language, and spatial cognition) conducted evidence-based reviews of frequently used tests. Reviews focused on suitability for office-based clinical practice, including test administration time, accessibility of normative data, disease populations studied, and availability in the public domain.

Results: Demographic and clinical practice data were obtained from 200 respondents who reported using a wide range of cognitive tests. Based on survey data and ancillary information, between 5 and 15 tests in each cognitive domain were reviewed. Within each domain, several tests are highlighted as being well-suited for an NBSE.

Conclusions: We identified frequently used single-domain cognitive tests that are suitable for an NBSE to help make informed choices about clinical cognitive assessment. Some frequently used tests have limited normative data or have not been well-studied in common neurologic disorders. Utilizing standardized cognitive tests, particularly those with normative data based on the individual's age and educational level, can enhance the rigor and utility of clinical cognitive assessment. *Neurology*® 2015;85:910–918

GLOSSARY

AAN = American Academy of Neurology; **AD** = Alzheimer disease; **BIT** = Behavioral Inattention Test; **BNS** = Behavioral Neurology Section; **BNT** = Boston Naming Test; **CERAD** = Consortium to Establish a Registry in Alzheimer's Disease; **FAB** = Frontal Assessment Battery; **FTLD** = frontotemporal lobar degeneration; **HVLT-R** = Hopkins Verbal Learning Test–Revised; **JLO** = Judgment of Line Orientation; **MMSE** = Mini-Mental State Examination; **MoCA** = Montreal Cognitive Assessment; **NBSE** = Neurobehavioral Status Exam; **ROCFT** = Rey-Osterrieth Complex Figures Test; **TBI** = traumatic brain injury; **TMT** = Trail Making Test; **UCNS** = United Council for Neurologic Subspecialties; **WAB-R** = Western Aphasia Battery–Revised.

Cognitive testing is essential for evaluating memory and other cognitive complaints associated with neurologic disorders. General cognitive tests, such as the Montreal Cognitive Assessment (MoCA),[1] Addenbrooke's Cognitive Examination–Revised,[2] or Mini-Mental State Examination (MMSE),[3] are useful for screening but have limited diagnostic specificity. By contrast, a Neurobehavioral Status Exam (NBSE)[4] is a collection of cognitive tests, chosen by a neurologist, intended to provide a more in-depth assessment of multiple cognitive domains, usually in

Supplemental data
at Neurology.org

*These authors contributed equally to this work.

From the Center for Brain/Mind Medicine (K.R.D., S.A.G., A.C.P.), Division of Cognitive and Behavioral Neurology, Department of Neurology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; Kessler Foundation Research Center (A.M.B.), West Orange, NJ; Department of Neurology (B.F.B.), Mayo Clinic, Rochester, MN; Department of Neurology and Center for Cognitive Neuroscience (A.C., H.B.C.), University of Pennsylvania, Philadelphia; Helen Wills Neuroscience Institute (M.D.), University of California, Berkeley; Department of Neurology (G.R.F.), University of Florida College of Medicine, Gainesville; Department of Neurology (D.R.G.), Northwestern University, Feinberg School of Medicine, Chicago, IL; Center for Brain Health (J.J.H.), School of Behavioral & Brain Sciences, University of Texas at Dallas; Department of Neurology (A.J.L.), University Hospitals Case Medical Center, Case Western Reserve University School of Medicine, Cleveland, OH; Department of Neurology and Neurological Sciences (K.J.M.), Stanford Comprehensive Epilepsy Center, Stanford University School of Medicine, CA; Western Institute for Neurodevelopmental Studies and Interventions (K.S.V.), Boulder, CO; and Memory Disorders Program (D.I.K.), UNC Department of Neurology, University of North Carolina at Chapel Hill.

Go to Neurology.org for full disclosures. Funding information and disclosures deemed relevant by the authors, if any, are provided at the end of the article.

conjunction with a neurologic evaluation, and sometimes as a separate assessment procedure (*Current Procedural Terminology* [CPT] 96116). If neuropsychology services are unavailable, the NBSE is an essential component of diagnostic assessment. If neuropsychology testing services are available, the NBSE may be still be useful as an adjunct or to help in determining whether more in-depth testing is needed, particularly for evaluating personality and psychological traits. In general, the NBSE is used to identify a patient's pattern of relative cognitive strengths and weaknesses, inform differential diagnosis, and guide therapeutic management. However, there are no guidelines for selecting tests for NBSE testing and virtually no information about how the NBSE is deployed in clinical practice. An NBSE differs from a formal neuropsychological assessment in its brevity, integration with other findings of a broader neurologic evaluation, and absence of formal intelligence testing or detailed assessment of personality or psychological attributes.

Many neurologists do not consider normative values for tests, and neurologists often use cognitive tests without ample normative data or that have not been studied in the conditions they are using them to assess. The NBSE Workgroup of the American Academy of Neurology (AAN) was established to review office-based cognitive testing in clinical practice as a reference point for quality improvement. Toward this end, behavioral neurologists were surveyed about their use of cognitive testing in clinical practice. A large sample of frequently used tests was reviewed in a standardized way, emphasizing desirable features for office-based testing: brief administration time (total NBSE duration of 30–60 minutes or less), diseases studied (diagnostic utility), existence of normative values (normal, abnormal, borderline), and availability in the public domain (accessibility and low cost). While not an exhaustive, systematic review, characterizing how the NBSE is used in clinical practice helped identify a compendium of tests most suitable for domain-specific, normative, office-based testing to improve the quality and utility of cognitive testing.

METHODS **NBSE practice survey.** The executive council of the Behavioral Neurology Section (BNS) of the AAN formed a workgroup to evaluate the NBSE. The workgroup identified 5

main cognitive domains by consensus and assigned leaders for each: attention (D.G.), executive function (K.D.), language (J.H.), memory (D.K.), and spatial cognition (A.M.B.). Between 5 and 15 tests within each cognitive domain were chosen by each subgroup to include in the survey (table 1). Tests that assess multiple domains (e.g., Clock-Drawing Test) were assigned to a single domain for review purposes.

A web-based survey using Google Documents was e-mailed to all 713 members of the AAN-BNS in March 2012, inquiring about the frequency of use of these selected tests and any additional tests that members administer. Respondents rated each test as "often use," "occasionally use," or "never use." The survey also asked about time devoted to testing, patient populations evaluated, and whether global cognitive screens (e.g., MMSE, Addenbrooke's Cognitive Examination–Revised), neuropsychiatric scales (e.g., Neuropsychiatric Inventory),[5] and functional rating scales (e.g., Functional Activities Questionnaire)[6] are also administered. A more detailed account of the survey data will be conveyed in a separate report.

Evidenced-based review. Tests that survey respondents indicated were either "often" or "occasionally" used were selected for review. Each cognitive domain subgroup used PubMed, Ovid, and textbooks to review the literature on the included tests. Each test review followed a similar format: (1) historical background and test description; (2) specific cognitive functions assessed; (3) copyright status (restricted or public domain); (4) administration time; (5) normative data across the adult lifespan; (6) patient populations studied; (7) advantages/limitations of the test; and (8) key references.

RESULTS **Survey.** Two hundred of the 713 AAN-BNS members (28%) responded to the survey (table 2). Most respondents (87%) were board-certified in Neurology and 22% had United Council for Neurologic Subspecialties (UCNS) certification in the Behavior Neurology & Neuropsychiatry subspecialty. The majority of respondents (79%) were adult neurologists and most (72%) were US-based.

Review. Data compiled for each test were used to create 37 summary tables, which are included in appendix e-1 on the *Neurology*® Web site at Neurology.org. Table 3 provides an example of a summary table, in this case for the Trail Making Test (TMT). These tables cover administration time, variations of the test used, a focused review of validity, normative data, some limited psychometric properties, and a consensus summary of the test's benefits and shortcomings. The data are intended to provide a concise reference guide to the clinical neurologist, in contrast to a comprehensive review of test characteristics and psychometric properties contained in neuropsychology textbooks.[7,8] Table 4 highlights some of the reviewed tests and summarizes whether each has properties that the workgroup believes may be important for test selection in clinical practice (e.g., available normative values; a brief administration time; availability in the public domain).

Attention domain tests. Tests of attention can be broadly divided into 2 groups: "basic" or "complex" attention. Basic attention is a prerequisite for all other

Table 1	Tests reviewed by cognitive domain
Attention	
Continuous Performance Test (CPT)	
Corsi Blocks	
Digit Span Backward	
Digit Span Forward	
N-Back	
Oral Trail Making	
Paced Auditory Serial Addition Test (PASAT)	
Sequential Operation Series	
Stroop	
Trail Making Test, Parts A and B	
Executive	
Antisaccade	
Frontal Assessment Battery (FAB)	
Frontal Behavioral Inventory (FBI)	
Luria Hand-Sequencing	
Proverb Test (D-KEFS)	
Similarities subtest (WAIS-IV)	
Language	
Boston Naming Test (BNT)	
Boston Naming Test (BNT)–15-Item Short Form	
Controlled Oral Word Association Test (COWAT)	
Semantic Category Fluency	
Western Aphasia Battery–Revised (WAB-R)	
Memory	
California Verbal Learning Test (CVLT)	
CERAD Word-List Memory	
Hopkins Verbal Learning Test–Revised (HVLT-R)	
Logical Memory Test (I and II)	
NYU Paragraph Recall	
Rey Auditory Verbal Learning Test (RAVLT)	
Verbal Paired Associates	
Spatial cognition	
Behavioral Inattention Test (BIT)	
Cancellation Tasks	
Catherine Bergego Scale	
Clock-Drawing Test (CDT)	
Cube Copying Test	
Judgment of Line Orientation (JLO)	
Line bisection tests	
Navon Figures	
Rey-Osterrieth Complex Figures Test (ROCFT)	

Abbreviations: CERAD = Consortium to Establish a Registry in Alzheimer's Disease; D-KEFS = Delis-Kaplan Executive Function System; NYU = New York University; WAIS-IV = Wechsler Adult Intelligence Scale–Fourth Edition.

mental functions and reflects the ability to sustain focus in order to perform a simple task. Complex attention reflects the ability to control, shift, and divide attentional focus, allowing for the manipulation of information and execution of multiple steps to accomplish a goal. Among tests of attention, all but the 3 continuous performance tests and some versions of the Stroop Test are not copyright-restricted (i.e., in the public domain).

Digit Span Forward[9] and TMT, Part A (TMT-A)[10] are frequently used tests of basic attention. The former assesses maintenance aspects of working memory and TMT-A indexes processing speed. These tests are brief, have robust normative values across the adult lifespan, and have been broadly studied in patients with Alzheimer disease (AD) and other cognitive disorders.[11,12] The Corsi Blocks Test[13] involves imitating sequential tapping patterns on blocks arrayed on a board. The longest sequence (i.e., number of blocks) tapped correctly is a measure of spatially mediated attentional capacity. The test is easy to administer and has been widely studied,[14] but scoring schemes are not well standardized.[15]

TMT, Part B (TMT-B), Months-of-the-Year-Backward, and the Stroop Test are complex attention tasks that are accessible in the public domain and have age-adjusted normative values across the adult lifespan.[16–18] TMT-B is a timed test in which patients are asked to alternately connect circled numbers and letters that are irregularly distributed across a page. It requires planning, sequencing, set-shifting, and response-inhibition. Months-of-the-Year-Backward is a Sequential Operations Series test in which a patient is asked to recite an overlearned sequence in reverse order. It can serve as a brief assessment of complex attention and working memory, and may also be useful for tracking change if administered as a timed test.

Executive domain tests. Tests of executive function range from assessing set-shifting and inhibitory control, to evaluating higher-order conceptual skills and abstract reasoning. Among tests reviewed, only some versions of the Proverbs Test[19] and the Similarities subtest of the Wechsler Adult Intelligence Scale[20] were copyright-restricted. The antisaccade task, which examines the ability to inhibit automatic motor behavior by exerting control over the strong tendency to look toward a visual stimulus, takes 3 minutes or less, has normative values throughout the lifespan,[21] and has been validated in tracking AD progression.[22] It has also been studied in different subtypes of mild cognitive impairment,[23] progressive supranuclear palsy, schizophrenia,[24] and frontotemporal lobar degeneration (FTLD).[25]

The Frontal Assessment Battery (FAB) includes 6 tasks that range from motor control programs to higher-order concept formation and interpretation.[26]

Table 2 Demographics of the 200 survey respondents (members of the AAN-BNS)

Mean age, y (SD)	51 (10.0)
Sex, n (%)	
Male	145 (73)
Female	55 (27)
Certification, n (%)	
Neurology (ABPN or NE)	164 (82)
Psychiatry (ABPN or NE)	7 (4)
Neurology and Psychiatry (ABPN or NE)	9 (5)
Neurology, with special qualification Child Neurology	8 (4)
No reply or other certifications (Internal Medicine, Geriatrics)	12 (6)
Behavioral Neurology & Neuropsychiatry (UCNS)	44 (22)
Time spent for new patient on NBSE (n = 129 responded), n (%)	
21-30 min	28 (22)
>30 min	101 (78)

Abbreviations: AAN-BNS = American Academy of Neurology–Behavioral Neurology Section; ABPN = American Board of Psychiatry and Neurology; NBSE = Neurobehavioral Status Exam; NE = national equivalent; UCNS = United Council for Neurologic Subspecialties.

It has robust age- and educational-based normative data, is not copyright-restricted, and has been used in FTLD, AD, and Parkinson disease.[27] Higher-order executive functions, including abstract thinking and concept formation, can also be assessed with the Similarities subtest of the Wechsler Adult Intelligence Scale–Fourth Edition[28] and the Proverbs Test of the Delis-Kaplan Executive Function System.[19] These tests may be useful as an adjunct to the FAB.

Language domain tests. Language tests were reviewed covering naming, comprehension, reading, writing, speech production, verbal fluency, and semantic-lexical retrieval. Of the 5 tests reviewed, only the Semantic/Category Fluency test is in the public domain.

Verbal fluency is often assessed by the ability to generate words based on semantic category (semantic fluency) or first letter (phonemic/lexical fluency). The semantic fluency test reviewed is part of several language batteries, including the Boston Diagnostic Aphasia Examination[29] and the Western Aphasia Battery–Revised (WAB-R).[30] Patients are asked to list as many exemplars of a specified category (e.g., animals, fruits) as possible in 1 minute. Normative values across the adult lifespan exist for several common categories.[31,32] This test has been studied in longitudinal cohorts of aging and patients with AD at various stages of severity.[33,34] Impaired semantic processing, as indexed by reduced semantic fluency, is a common early sign of underlying AD pathology.[35] The Controlled Oral Word Association Test,[36,37] a phonemic fluency task in which patients generate words from a letter cue (e.g., "F"), also has age- and education-adjusted norms

across the adult lifespan[38] and has been studied in multiple conditions.[33,34,39,40] The WAB-R is a 45- to 60-minute battery that examines the major aspects of language (e.g., comprehension, writing, fluency) and has ample normative data and validity in detecting and tracking language impairment in many disorders.[e1,e2] Although the length of the WAB-R makes it impractical for brief testing, there is a "Bedside WAB-R" tool that screens for major aspects of language dysfunction, takes only 15 to 20 minutes to administer, and has separate normative values.[e3]

The Boston Naming Test (BNT)–15-Item Short Form[e4] is one of several abbreviated versions of the BNT,[e5] a visual confrontation naming test. The BNT has been investigated for use in neurodegenerative diseases (e.g., AD, FTLD), vascular dementia, multiple sclerosis, and epilepsy and may be useful for the longitudinal tracking of aphasia due to multiple etiologies.[e6] Performance on many BNT-15 versions, composed of different subsets of the original BNT stimuli,[e7] is highly correlated with the full 60-item BNT.[e8] In addition to the existence of normative data for older adults (50s to 90s),[e8] this feature makes the BNT-15 useful for brief evaluation like the NBSE.

Memory domain tests. All tests reviewed were verbally based auditory memory tests. Several tests (e.g., Rey-Osterrieth Complex Figures Test [ROCFT])[e9] discussed in the attention and spatial cognition sections have visual-based memory components. Among the 7 tests reviewed, only the Consortium to Establish a Registry in Alzheimer's Disease (CERAD) Word-List Memory Test[e10] and the Rey Auditory Verbal Learning Test[e9] are in the public domain. The CERAD is a 10-word list of unrelated words that includes 3 learning trials, followed by a delayed recall and forced-choice recognition trial. The test has ample normative values for ages 50 to 90, is validated in several longitudinal AD trials, and has a relatively brief administration time.[e8–e10] The Hopkins Verbal Learning Test–Revised (HVLT-R)[e11] is another relatively brief 12-word list that contains 4 words from each of 3 categories (e.g., gems). This "semantic clustering" provides an opportunity to use associative learning as a strategy for encoding and retrieving words. The HVLT-R has been studied in a number of conditions, including vascular dementia[e12] and traumatic brain injury (TBI).[e13]

Another approach to assessing verbal memory is the recall of prose passages. Learning and storing information that is part of a colorful narrative text is often less demanding than memorizing a list of unrelated words. Assessing both can provide multiple perspectives on the severity of memory deficits. The Logical Memory I and II subtests of the Wechsler Memory Scale[e12] have robust normative values for ages 16 to 90[e12] and,

Table 3	Example of a summary table from the evidenced-based review
Test name	Trail Making Test
Description	A brief, 2-part test of basic attention and working memory. The original Trail Making Test (TMT) was developed by a team of US Army psychologists as part of the Army Individual Test Battery (1944).[1] Different administration and scoring systems were subsequently introduced. It has developed into a widely used instrument, with versions incorporated into the Delis-Kaplan Executive Function System (D-KEFS)[2] and the Halstead-Reitan Battery (HRB).[3]
	In Part A (TMT-A), the patient draws lines connecting consecutively numbered circles. In Part B (TMT-B), the patient draws lines alternating between circles containing numbers and circles containing letters, fulfilling a specific, alternating sequence (i.e., 1-A, 2-B, 3-C). The measured value is the time to completion for each part. In some adapted versions, such as the HRB, the examiner notes errors and points them out to the patient during the test. Other factors that might affect performance are visual scanning and motor speed.
Specific functions assessed	Attention/visual scanning, motor speed, working memory
Subscales	Of the common versions, the D-KEFS involves 5 subscales:
	Visual Scanning. This is a cancellation task in which the patient is asked to identify and cross out a specific target on an $11'' \times 17''$ page. This test would also serve to identify hemispatial neglect. Omission/commission errors are included in the scoring as well as time to complete.
	Number Sequencing. This is a number sequencing task (similar to TMT-A).
	Letter Sequencing. This involves connecting letters of the alphabet in sequence.
	Number-Letter Switching. The patient is asked to draw lines, switching between numbers and letters in sequence (e.g., 1-A, 2-B). This is the primary executive function task, similar to TMT-B.
	Motor Speed. The patient traces over a line connecting a series of circles.
Number of items/scoring	On the 2-part TMT, the times to completion for A and B are recorded and the difference of time to completion on B minus A is used as the variable and compared with norms. A large B minus A difference reflects attentional difficulties and/or problems with working memory. An alternative computation involves the ratio B/A, but this is thought to be less reliable.[4]
Copyright status	Copyrighted[a] and public domain[b]
Administration time	Variable depending on the version and type of patient, typically 5–20 min.
Normative psychometric data	Norms are specific to the version. Ample norms are available for ages 18–89 years.[5] The D-KEFS version provides age-based norms for each of the 5 subtests (listed above). Norms for ages 8–89 years are provided for each subscale task and error analysis. In addition to time-completion scores, error scores for sequencing errors vs set-loss error scores are provided. Contrasts between performances on the various subscale tasks (with norms) are also provided.
Sensitivity and specificity	Vary across versions. Performance on the TMT, particularly Part B, correlates with reduced driving performance[6,7] and mobility impairment with aging.[8] The TMT has also been used in studies of Huntington disease[9] and traumatic brain injury,[10] among other disorders.
Advantages	Sensitive measure of sustained attention, attentional control, and set-shifting, which has been widely applied to various disorders (e.g., attention deficit disorder). Careful observation of the way patients approach the task, the type of errors they make, and analysis of the type of errors contribute to the test's value.
Limitations	Task performance is affected by a number of factors (e.g., age, educational status) in addition to neurocognitive issues. These need to be factored into the analysis.[11,12]
References	1. Army Individual Test Battery. In: War Department, editor. *Manual of Directions and Scoring*. Washington, DC: Adjutant General's Office; 1944.
	2. Delis DC. *Delis-Kaplan Executive Function System (D-KEFS)*. San Antonio: The Psychological Corporation; 2001.
	3. Lezak MD. *Neuropsychological Assessment*, 3rd ed. New York: Oxford University Press; 1995.
	4. Martin TA, Hoffman NM, Donders J. Clinical utility of the Trail Making Test ratio score. *Appl Neuropsychol* 2003;10:163–169.
	5. Tombaugh TN. Trail Making Test A and B: normative data stratified by age and education. *Arch Clin Neuropsychol* 2004;19:203–214.
	6. Emerson JL, Johnson AM, Dawson JD, Uc EY, Anderson SW, Rizzo M. Predictors of driving outcomes in advancing age. *Psychol Aging* 2012;27:550–559.
	7. Hargrave DD, Nupp JM, Erickson RJ. Two brief measures of executive function in the prediction of driving ability after acquired brain injury. *Neuropsychol Rehabil* 2012;22:489–500.
	8. Vazzana R, Bandinelli S, Lauretani F, et al. Trail Making Test predicts physical impairment and mortality in older persons. *J Am Geriatr Soc* 2010;58:719–723.
	9. O'Rourke JJ, Beglinger LJ, Smith MM, et al. The Trail Making Test in prodromal Huntington disease: contributions of disease progression to test performance. *J Clin Exp Neuropsychol* 2011;33:567–579.

Continued

Table 3	Continued
Test name	Trail Making Test
	10. Perianez JA, Rios-Lago M, Rodriguez-Sanchez JM, et al. Trail Making Test in traumatic brain injury, schizophrenia, and normal ageing: sample comparisons and normative data. *Arch Clin Neuropsychol* 2007;22:433–447.
	11. Bornstein RA, Suga LJ. Educational level and neuropsychological performance in healthy elderly subjects. *Dev Neuropsychol* 1988;4:17–22.
	12. Stuss DT, Stethem LL, Pelchat G. Three tests of attention and rapid information processing: an extension. *Clin Neuropsychol* 1988;2:246–250.

[a] The D-KEFS is copyrighted and commercially available.
[b] The original Trail Making Test[5] and Halstead-Reitan adapted version are in the public domain.

together, take about 10 minutes to complete, with a suggested 20-minute delay. Performance on Logical Memory correlates with outcomes in several neuropsychiatric settings, including recovery from TBI,[e13] social functioning in schizophrenia,[e14] and the ability to comprehend and reason about treatment decisions in mild to moderate AD.[e15]

Various 3-word and 5-word recall tests were not reviewed in this study. There are limited normative data for these tests and recall performance varies widely depending on the words used and whether patients are instructed to remember the words. The recall of 0 or 1 of 3 words presented may be a "red flag" for memory impairment in adults aged 65 to 90

Table 4	Summary of results: Selected cognitive tests			
Test name	Administration time, min	Normative data (adults)	Psychometric data for dementia	Public domain
Attention				
Trail Making Test, A and B	5–10	√	√	√
Digit Span Forward	3–5	√	√	√
Digit Span Backward	3–5	√	√	√
Sequential Operation Series (e.g., Months-of-the-Year-Backward)	3 (per task)	√		√
Stroop	7–10	√	√	√
Executive				
Antisaccade	3	√	√	√
Similarities subtest (WAIS-IV)	3–5	√	√	
Frontal Assessment Battery	10	√	√	√
Clock-Drawing Test	3–7	√	√	√
Language				
Boston Naming Test–15-Item Short Form	3–5	√		
Controlled Oral Word Association	5	√	√	
Semantic Category Fluency	5	√	√	√
Memory				
CERAD Word-List Learning	5–10		√	√
Hopkins Verbal Learning Test	5–10	√	√	
Rey Auditory Verbal Learning Test	15	√	√	√
Spatial cognition				
Short-forms JLO	10	√	√	
Cube Copying Test	2–3		√	√
Cancellation Test	3–5		√	√
Line bisection test	1–2			√

Abbreviations: CERAD = Consortium to Establish a Registry in Alzheimer's Disease; JLO = Judgment of Line Orientation; WAIS-IV = Wechsler Adult Intelligence Scale–Fourth Edition.

years, as suggested by the Mini-Cog screening test.[e16] However, up to 19% of older adults who recall only 1 word on a 3-word test perform normally on more detailed memory testing, suggesting that these simple tests may not accurately reflect a patient's memory function.[e16,e17]

Spatial cognition domain tests. Nine tests in this domain were reviewed, including tests of visuoperceptual processing; visuoconstructive ability (which depends on organization and planning); and spatial bias processing associated with spatial inattention, neglect syndromes, or simultanagnosia. Most of the spatial cognition tests reviewed are in the public domain, except the Behavioral Inattention Test (BIT)[e18] and the Benton Judgment of Line Orientation (JLO).[e19]

Among visuoconstructive tests, the Clock-Drawing Test is well-suited for an NBSE. It is brief, has abundant normative data[e20,e21] (although based on several different scoring criteria), and is validated in stroke, AD, and a range of psychiatric disorders.[e22–e24] It has been used for decades to both detect and serially grade disease severity in degenerative dementias.[e25–e27] The Cube Copying Test, also a part of the Short Test of Mental Status[e28] and the MoCA, involves copying a 3-dimensional cube. Although typically scored as correct/incorrect, several quantitative scoring methods are available.[e29,e30] The JLO[e19] is useful for identifying visuoperceptual deficits in judging spatial relationships. There are several short forms of the JLO that utilize a subset of the stimuli and significantly reduce administration time but are also copyright-protected.[e31] Performance on short forms of the JLO generally correlates well with full JLO scores.[e31] Normative values for both the JLO[e32] and some frequently used short forms[e33] are available, particularly for ages 50 to 90.

Visual cancellation tests[e34] assess visual scanning strategies and spatial bias and may be a useful component of an NBSE. Typically, patients are asked to mark each occurrence of target stimuli that are scattered across a page containing distracter items. Many versions of this brief test are in the public domain, and some have normative values across the adult lifespan.[e35] Line bisection tests are easy to administer and also provide a straightforward assessment of spatial bias. Both cancellation and line bisection tests have been well studied and validated in patients with hemispheric lesions due to stroke or tumors.

The BIT[e18] and ROCFT[e9] examine spatial cognitive processes, but may be challenging to include in an office-based NBSE. The BIT is lengthy, has limited normative values across the adult lifespan, and is copyright-protected. The ROCFT complements verbal memory tests by assessing the acquisition and storage of visual (nonverbal) information, and has abundant normative values.[e36] Drawbacks include lengthy administration time and potential confounding by organizational and planning deficits.

DISCUSSION Cognitive testing by behavioral neurologists often entails a mixture of relatively brief cognitive tests, such as 3-word memory recall or a motor Go/No Go test, and abbreviated versions of neuropsychological assessments or subtests of test batteries. Whereas the former may be informative in some clinical contexts, tests that lack standardized administration and scoring guidelines have less utility for an NBSE. The tests reviewed varied considerably in terms of available normative data and information on their use in specific patient populations (e.g., AD, TBI). Some tests (e.g., cancellation, Luria Hand-Sequencing) do not have available normative values to use for office-based reference and scoring.

Relevance to clinical practice: Toward evidence-based NBSE. The availability of age- and education-appropriate normative values may be the most important test characteristic to help accurately interpret a patient's performance. Given the wide age range of adults seeking neurology evaluations, tests with normative values across the adult lifespan are often most helpful. Many normative values are published and some relevant citations for each test are in appendix e-1. Although test choice should ideally be based on psychometric validity for the symptoms being assessed, both the cost and limited access of copyrighted tests may be nontrivial barriers for many practices. Tests used in an NBSE, especially those that are subtests of neuropsychological batteries, generally require formal training in test administration, scoring, and interpretation, with provision for expert guidance and feedback during the learning phase.

As a reference guide for the NBSE, this review is intended to help neurologists better assess major cognitive domains with tests that are most appropriate for evaluating hypotheses about a given patient. The pattern of impairments can suggest which neural networks may be disrupted. Along with the available history, neurologic examination, and laboratory or imaging data, this can help to identify the likely underlying pathologic process. For example, a profile of deficits in memory storage, semantic fluency, and complex attention suggests disruption of temporolimbic structures (e.g., hippocampus), temporal neocortex, and frontal networks. In a 75-year-old patient with progressive deficits but no functional decline, this pattern suggests a diagnosis of mild cognitive impairment, amnestic subtype, which often reflects underlying AD pathology.

Our review suggests some incongruence between a few widely used tests and those that we suggest may be most optimal for an NBSE. For example, the Luria Hand-Sequencing Test was reported as either often used or occasionally used by more than two-thirds of respondents. Although this test is brief and does not require test materials, it has limited published norms across the adult lifespan. The Luria test is a component of several batteries (e.g., the FAB) for which norms are available. Although a patient's inability to learn a Luria sequence or the tendency to exhibit sustained, perseverative responses would clearly represent impaired performance, judging whether more subtle sequencing errors is abnormal for age is much more challenging in the absence of norms. Also, various 3- or 5-word memory recall tests were reported as used often by more than half of respondents. Although brief and often available in the public domain, these frequently used bedside tests have limited normative data and have been less studied in dementia populations than the word-list memory tests reviewed here.

Neuropsychological evaluation. An NBSE conducted by neurologists may bring to light issues that would benefit from a comprehensive neuropsychological assessment. Such evaluations are much more extensive (2–6 hours) than the typical 30- to 60-minute NBSE and provide a more detailed, quantitative assessment of cognitive function, particularly regarding premorbid estimates of cognitive abilities. For example, neuropsychological assessment may corroborate or challenge results from an NBSE and help track disease progression over time, provide feedback to families about changes in the patient's status, and allow for the planning of cognitive rehabilitation programs. Neuropsychological evaluation may also be helpful in patients who show borderline impairment on an NBSE, are at either end of the educational spectrum, have prominent psychiatric comorbidities, or who may be involved in medicolegal action. In our experience, collaboration between neurologists and neuropsychologists often facilitates optimal diagnostic and treatment plans for patients. It is essential that clinicians be cognizant of practice effects if both an NBSE and neuropsychological evaluation are completed in close succession.

Limitations. We reviewed a small number of the hundreds of cognitive tests that have been developed. Although our survey data helped identify the most frequently used tests for review, our findings do not preclude other available tests from being appropriate for an NBSE. This may be particularly true for tests that were not initially, or are not widely, used in English. We did not review the tests "written-in" by respondents as being used in addition to the surveyed tests. Some of the tests evaluated have limited normative data across certain patient populations and do not account for medication-related or practice effects. In addition, the literature on cognitive testing is extensive, and important findings for the reviewed tests may have been inadvertently omitted.

In an effort to maintain brevity, we only summarily reviewed the psychometric properties of individual tests, and we urge clinicians to refer to comprehensive neuropsychology texts as needed for additional information about individual tests. We also did not specifically address general cognitive tests such as the MMSE or MoCA or how to integrate them into an NBSE, both issues which are beyond the scope of this report. Our survey sample rate was relatively low (200/713, 28%) and only 44 of 200 respondents (22%) were UCNS certified in Behavioral Neurology & Neuropsychiatry. For the current report, the survey data were primarily used as a guide, along with the authors' consensus, for determining which tests would be reviewed. This study provides a glimpse into the current practice of formal and informal cognitive testing among behavioral neurologists and provides an evidence-based review and a test selection resource to help neurologists make informed choices about testing. A large-scale and broader survey of other subspecialty neurologists who perform cognitive assessment in various settings would complement this study. We also envision future studies that incorporate general screening tests and longitudinal follow-up. This review of single-domain cognitive tests serves as a starting point for quality improvement in clinical cognitive assessment.

REFERENCES

1. Nasreddine ZS, Phillips NA, Bedirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005;53:695–699.

2. Mioshi E, Dawson K, Mitchell J, Arnold R, Hodges JR. The Addenbrooke's Cognitive Examination Revised (ACE-R): a brief cognitive test battery for dementia screening. Int J Geriatr Psychiatry 2006;21:1078–1085.

3. Folstein MF, Folstein SE, McHugh PR. "Mini-Mental State": a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975;12:189–198.

4. American Medical Association. Current Procedural Terminology (CPT). Chicago: American Medical Association; 2013.

5. Cummings JL, Mega M, Gray K, Rosenberg-Thompson S, Carusi DA, Gornbein J. The Neuropsychiatric Inventory: comprehensive assessment of psychopathology in dementia. Neurology 1994;44:2308–2314.

6. Pfeffer RI, Kurosaki TT, Harrah CH Jr, Chance JM, Filos S. Measurement of functional activities in older adults in the community. J Gerontol 1982;37:323–329.

7. Lezak M. Neuropsychological Assessment, 5th ed. New York: Oxford University Press; 2012.

8. Strauss E, Elizabeth MS, Spreen O. A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary, 3rd ed. New York: Oxford University Press; 2006.

9. Lezak MD. Neuropsychological Assessment, 3rd ed. New York: Oxford University Press; 1995.

10. Army Individual Test Battery. In: War Department, editor. Manual of Directions and Scoring. Washington, DC: Adjutant General's Office; 1944.

11. Ahn HJ, Seo SW, Chin J, et al. The cortical neuroanatomy of neuropsychological deficits in mild cognitive impairment and Alzheimer's disease: a surface-based morphometric analysis. Neuropsychologia 2011;49:3931–3945.

12. Martyr A, Clare L. Executive function and activities of daily living in Alzheimer's disease: a correlational meta-analysis. Dement Geriatr Cogn Disord 2012;33:189–203.

13. Corsi PM. Human memory and the medial temporal region of the brain. Diss Abstr Int 1972;34:891B.

14. Berch DB, Krikorian R, Huha EM. The Corsi block-tapping task: methodological and theoretical considerations. Brain Cogn 1998;38:317–338.

15. Kessels RP, van Zandvoort MJ, Postma A, Kappelle LJ, de Haan EH. The Corsi block-tapping task: standardization and normative data. Appl Neuropsychol 2000;7:252–258.

16. Troyer AK, Leach L, Strauss E. Aging and response inhibition: normative data for the Victoria Stroop Test. Neuropsychol Dev Cogn B Aging Neuropsychol Cogn 2006;13:20–35.

17. Ostberg P, Hansson V, Haagg S. Adult norms and test-retest reliability for the Months Backward test: durational and response accuracy measures. Logoped Phoniatr Vocol 2012;37:11–17.

18. Tombaugh TN. Trail Making Test A and B: normative data stratified by age and education. Arch Clin Neuropsychol 2004;19:203–214.

19. Delis DC. Delis-Kaplan Executive Function System (D-KEFS). San Antonio: The Psychological Corporation; 2001.

20. Wechsler D. Wechsler Adult Intelligence Scale, 4th ed. San Antonio: Pearson; 2008.

21. Klein C, Fischer B, Hartnegg K, Heiss WH, Roth M. Optomotor and neuropsychological performance in old age. Exp Brain Res 2000;135:141–154.

22. Crawford TJ, Higham S, Renvoize T, et al. Inhibitory control of saccadic eye movements and cognitive impairment in Alzheimer's disease. Biol Psychiatry 2005;57:1052–1060.

23. Hellmuth J, Mirsky J, Heuer HW, et al. Multicenter validation of a bedside antisaccade task as a measure of executive function. Neurology 2012;78:1824–1831.

24. Gooding DC, Tallent KA. The association between anti-saccade task and working memory task performance in schizophrenia and bipolar disorder. J Nerv Ment Dis 2001;189:8–16.

25. Burrell JR, Hornberger M, Carpenter RH, Kiernan MC, Hodges JR. Saccadic abnormalities in frontotemporal dementia. Neurology 2012;78:1816–1823.

26. Dubois B, Slachevsky A, Litvan I, Pillon B. The FAB: a Frontal Assessment Battery at bedside. Neurology 2000;55:1621–1626.

27. Oshima E, Terada S, Sato S, et al. Frontal Assessment Battery and brain perfusion imaging in Alzheimer's disease. Int Psychogeriatr 2012;24:994–1001.

28. Wechsler D. Wechsler Memory Scale, 4th ed. San Antonio: Pearson; 2008.

29. Goodglass H, Kaplan E, Barresi B. The Assessment of Aphasia and Related Disorders, 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2001.

30. Kertesz A. Western Aphasia Battery–Revised. San Antonio: PsychCorp; 2007.

31. Gladsjo JA, Schuman CC, Evans JD, Peavy GM, Miller SW, Heaton RK. Norms for letter and category fluency: demographic corrections for age, education, and ethnicity. Assessment 1999;6:147–178.

32. Holtzer R, Goldin Y, Zimmerman M, Katz M, Buschke H, Lipton RB. Robust norms for selected neuropsychological tests in older adults. Arch Clin Neuropsychol 2008;23:531–541.

33. Clark LJ, Gatz M, Zheng L, Chen YL, McCleary C, Mack WJ. Longitudinal verbal fluency in normal aging, preclinical, and prevalent Alzheimer's disease. Am J Alzheimers Dis Other Demen 2009;24:461–468.

34. Teng E, Leone-Friedman J, Lee GJ, et al. Similar verbal fluency patterns in amnestic mild cognitive impairment and Alzheimer's disease. Arch Clin Neuropsychol 2013;28:400–410.

35. Chertkow H, Whatmough C, Saumier D, Duong A. Cognitive neuroscience studies of semantic memory in Alzheimer's disease. Prog Brain Res 2008;169:393–407.

36. Benton AL, Hamsher K. Multilingual Aphasia Examination. Iowa City: AJA Associates; 1989.

37. Benton AL, Sivan AB, Hamsher K, Varney NR, Spreen O. Contributions to Neuropsychological Assessment: A Clinical Manual, 2nd ed. New York: Oxford University Press; 1994.

38. Loonstra AS, Tarlow AR, Sellers AH. COWAT meta-norms across age, education, and gender. Appl Neuropsychol 2001;8:161–166.

39. Kremen WS, Seidman LJ, Faraone SV, Tsuang MT. Is there disproportionate impairment in semantic or phonemic fluency in schizophrenia? J Int Neuropsychol Soc 2003;9:79–88.

40. Roca M, Manes F, Gleichgerrcht E, et al. Intelligence and executive functions in frontotemporal dementia. Neuropsychologia 2013;51:725–730.