

**Table 4**  
Detailed review of tests that include traumatic brain injury in the test purpose, applications, or standardization sample. If not stated, test/subtest did not meet criteria. See Appendix B for definitions for evaluation categories. DNT = Did not test. Criteria for Inter-Rater Reliability = Strict  $r \geq .90$ , Kappa  $\geq .80$ ; Relaxed  $r \geq .80$ , Kappa  $\geq .70$ . Criteria for Internal Consistency (split half) = Strict Cronbach Alpha  $\geq .90$ , Relaxed  $\geq .80$ . Criteria for Test-Retest reliability = Strict  $r \geq .90$ , Relaxed  $r \geq .80$ .

<u>Test</u>	<u>Inter-rater</u>	<u>Internal consistency</u>	<u>Test-retest reliability</u>	<u>Content/Face validity</u>	<u>Construct validity</u>	<u>Criterion-related validity: Concurrent</u>	<u>Criterion-related validity: Predictive</u>
Aphasia Diagnostic Profiles (ADP; Helm-Estabrooks, 1992)	DNT Suggested training in teams.	7/9 subtests met strict criterion, 1/9 subtests met relaxed criterion.	1/12 subtests met strict criterion, 3/12 subtests met relaxed criterion.	Author is an expert in aphasiology. Based on the premise that individuals with acquired brain damage will have a wide range of linguistic problems and a wide range of severity. "Includes tasks that have a range of difficulty to allow clinician to identify severely impaired and relatively spared language skills and those in between." Designed to classify/profile aphasia type based on preserved language skills. Includes measures of non-verbal characteristics of aphasia, e.g. behavioral profile and alternate communication profile.	19 subjects with aphasia improved scores on 9 subtests at re-testing at 11 days after initial test, as expected during recovery from stroke. Intercorrelations among scores on subtests and coefficients are mostly in the moderate to moderately high range (.40-.80). Four-factor solution generated on 217 subjects; 54% of variance accounted for by first factor, described as general language expression factor.	DNT	DNT
American Speech Language Hearing Association Functional Assessment of Communication Skills for Adults (ASHA FACS; Frattali, Thompson, Holland, Wohl, & Ferketic, 1995)	Completed by 2 examiners for same subject within 48 hours after a minimum of 3 subject-examiner contacts, for 35 subjects. Overall scores $r=.95$ ; Social and Basic Needs $r=.92$ , Reading/writing/#s and Planning $r=.89$ . For Qualitative	Majority of items (all but 3) covered full 7-pt range of scale – including in mild group (supports sensitivity). Inter-item correlations consistently high; also between items and total score except for 4 items; mean correlation $=.82$ . Analyzed item-	Calculated intra-rater reliability (i.e. test-retest) on 38 subjects within 2-5 days: all $r$ 's $\geq .94$ . For Qualitative Dimensions overall $r=.99$ . All 4 dimensions $\geq .94$ .	Effects of race, education and sex considered, though generalizability limited by "insufficient or unbalanced samples". Education and sex were significant factors for Reading/writing, Daily Planning, Overall independence score, and for Appropriateness, Promptness and Adequacy in qualitative domains. Sign differences by sex for qualitative adequacy. Consumer survey used to	Factor analysis: all 44 items loaded on one factor that accounted for 59% of the total variance; 4 factor solution accounted for "more than" 73% of total variance: factors appeared to be 4 assessment domains – Social communication, Communication of basic needs, Reading/writing and # concepts, Daily planning. Overall communication independence scores	Compared to: 1) FIM scales for comprehension, expression, problem solving, memory, social interaction: a) for aphasia - all significant but modest; range $=.53-.87$ ; b) for cognitive-communication disorders – all significant; range $=.61-.87$ .	DNT

<u>Test</u>	<u>Inter-rater</u>	<u>Internal consistency</u>	<u>Test-retest reliability</u>	<u>Content/Face validity</u>	<u>Construct validity</u>	<u>Criterion-related validity: Concurrent</u>	<u>Criterion-related validity: Predictive</u>
	<p>Dimensions, overall <math>r=.88</math>, Adequacy <math>.82</math>, Comm. Sharing <math>.84</math>, Appropriateness <math>.77</math>, Promptness <math>.72</math>.</p>	<p>total correlations for each diagnostic group and all but 4 (aphasia) and 3 (TBI) items had moderate-high correlations. Items with low correlations kept anyway because they were considered pragmatically important and important to communication with family and friends. Communication independence mean scores for domains correlated with overall score <math>.89-.99</math>. Qualitative domain scores correlated with overall score <math>.91-.96</math></p>		<p>rate each item in re: its "importance to functional communication" – 69% or more of 179 respondents rated each item as "somewhat important" or "very important". 31 clinicians rated usability and most had positive feedback (detailed results provided). 7 peer reviewers reviewed the test and made largely positive comments and suggestions for improvement (details provided). Published version has some changes based on this – 3 deletions and one wording change in instructions.</p>	<p>correlated <math>.93</math> with overall qualitative scores. Significant difference between cognitive-communication and aphasia groups for Social communication, and Reading/writing and # concepts. Significant difference between fluent and nonfluent aphasia for all 4 communication independence domains. Qualitative domains also significantly better both in cognitive-communication disorders than aphasia and in less severely impaired subjects, except Appropriateness (explained more by severity)</p>	<p>2) WAB subtest and AQ scores for subjects with aphasia: all significant; range <math>=.38-.81</math>. 3) Scales of Cognitive Ability for Traumatic Brain Injury for subjects with cognitive-communication disorders: all significant; range <math>=.53-.79</math> 4) Correlated domain and overall independence scores with Rating of Communication Abilities from families and clinicians. Correlations ranged from <math>.43-.64</math>. Families did not report higher measures than clinicians (range of correlations <math>= .49-.73</math>).</p>	

<u>Test</u>	<u>Inter-rater</u>	<u>Internal consistency</u>	<u>Test-retest reliability</u>	<u>Content/Face validity</u>	<u>Construct validity</u>	<u>Criterion-related validity: Concurrent</u>	<u>Criterion-related validity: Predictive</u>
<p>Behavior Rating Inventory of Executive Function (BRIEF; Gioia, Isquith, Guy, &amp; Kenworthy, 1996)</p>	<p>Parent vs. teacher ratings had low correlation (.30-.50), consistent with the literature and probably related to differences in context in which ratings were made.</p>	<p>All index scores and 19/32 subscale scores for Parent and Teacher version for Normative and combined clinical samples (i.e., 8 subscales per group) met strict criterion. Remaining 13/32 subscales met relaxed criterion.</p>	<p>Re-tested over 2-6 week period. Teacher form: Normative sample met strict criterion for index scores, relaxed criterion for all subscales. Clinical sample DNT. Parent form: met relaxed criterion for all index scores and 4/8 subscale scores for both Normative and Clinical samples.</p>	<p>Developed based on review of literature and identification of domains of executive function. Domains then reviewed by colleagues in neuropsychology and were the basis for the BRIEF normative edition. Revised based on factor analysis from first stage of standardization. Designed to be easy to administer and score and yield clinically useful information. Each scale "built from items whose content reflects the intended domain." (p. 35) Items collected from high-frequency issues raised in clinical interviews rather than generated by authors. Inter-rater agreement on item clusters determined by calculating responses of neuropsychologists with expertise in the area. Tryouts conducted. Inconsistency and negativity scales added to address response bias. No significant effect of ethnic group membership. Low SES scores associated with lower BRIEF scores.</p>	<p>Other tests reviewed to ensure completeness of scales and divergence from existing measures. Age groupings determined by examination of data (although developmental trends not evident on visual inspection of those data). Factor analysis on normative sample, combined clinical samples showed 2 factor solution; inclusion of related measures (i.e. Child Behavior Checklist/Teacher Rating Form, Behavior Assessment System for Children, ADHD Rating Scale-IV) for normative and clinical samples showed discriminant validity (i.e. added new factors where constructs were hypothesized to differ, loaded on same factors where they were hypothesized to be similar). Studied on children with Attention Deficit Hyperactivity Disorder (ADHD), brain injury, Tourette's Disorder (with and without ADHD), Reading Disorder, low birth weight, high functioning autism, pervasive developmental disorders, mental retardation, and children with documented frontal vs. extrafrontal lesions. Overall, discriminated</p>	<p>No other executive function tests of this type existed, so examined patterns of relationships by correlating individual subscales with related measures e.g., ADHD Rating Scale-IV, Child Behavior Checklist, Teacher's Report Form, Behavior Assessment System for Children, Connors' Ratings Scale. Correlations were high where constructs were hypothesized to converge (e.g., Inhibit cluster vs. Hyperactivity-Impulsivity on the ADHD-IV scale) and low where constructs were hypothesized to differ (e.g., between Inhibit and Somatic Complaints on Child Behavior Checklist).</p>	<p>Parent and Teacher ratings on Working Memory and Inhibit subscales predicted diagnostic group with 70%-85% accuracy, for ADHD inattentive type or combined type diagnosis. Sensitivity and specificity calculated on extant samples with known prevalence. Overall, this appears to be modest – e.g., Parent form T-score of 75 classifies 56% of ADHD inattentive type correctly and misclassifies 10% of normal sample. Teacher</p>

<u>Test</u>	<u>Inter-rater</u>	<u>Internal consistency</u>	<u>Test-retest reliability</u>	<u>Content/Face validity</u>	<u>Construct validity</u>	<u>Criterion-related validity: Concurrent</u>	<u>Criterion-related validity: Predictive</u>
					between controls and other groups, except for children with Phenylketenuria and Mental retardation – groups differed on only 2 and 1 score respectively.		form is marginally better. Authors suggested that BRIEF be used to supplement other tools rather than for screening or diagnosis.
Brief Test of Head Injury (BTHI; Helm-Estabrooks & Hotz, 1991)	DNT	Severity score and total score met strict criterion; 5/7 subtests met relaxed criterion.	19/53 patients with TBI retested within 57 days of injury (mean time between testing = 19.1 days): met relaxed criterion for BTHI total raw score. 18 patients with TBI in standardization sample were retested within 86 days after injury (mean time between testing = 16.6 days). Did not meet criterion: $r=.46$ for BTHI total raw score (may be consistent with early stage of recovery).	Constructed by experts based on observed clinical need for quick probes of a variety of skills in a wide range of patients. Targeted areas were chosen based on literature on sequelae from TBI. Designed to be used with patients who are evolving quickly and have limited tolerance for test administration.	Raw score accurately classified 100% of 41 uninjured subjects and 85% of 52 patients with TBI when testing within 30 days of injury with a cutoff score of 43. Within 7 days of testing, test accurately classified 100% of uninjured subjects and 64% of subjects with TBI, using a cutoff score of 38. No significant correlation with age and time post-onset.	Total raw scores differed significantly between patients classified via Glasgow Coma Scale (GCS) scores as severe vs. mild, and moderate vs. mild, but not moderate vs. severe. Mean scores differed across levels of Rancho Los Amigos Levels of Cognitive Functioning (LCF) except IV and V in 193 subjects in standardization sample. Significant correlation between total score and GCS at admission, GCS at test administration, LCF at test administration.	DNT

<u>Test</u>	<u>Inter-rater</u>	<u>Internal consistency</u>	<u>Test-retest reliability</u>	<u>Content/Face validity</u>	<u>Construct validity</u>	<u>Criterion-related validity: Concurrent</u>	<u>Criterion-related validity: Predictive</u>
Communication Activities of Daily Living, Second Edition (CADL-2; Holland, Frattali, & Fromm, 1999)	Stanine scores calculated by 2 ProEd Inc. research staff for 30 scored protocols, met strict criterion (though they didn't have two people actually score one examinee).	Met strict criterion. Also calculated SEM=sqr(1-alpha) for 68% and 95% confidence intervals.	Met relaxed criterion for adults with chronic aphasia at 2-32 week test-retest interval.	Test designed to answer specific questions regarding functional communication (e.g. What do normal readers read? What are some elementary adaptive strategies for clarifying miscommunications?). 10 certified speech language pathologists answered questionnaires about scope and appropriateness of items, categories, stimulus pictures. Item discrimination (i.e. extent to which each item correlates with total score) analyzed using point-biserial correlation technique and 49/50 items met .2 criterion for significant coefficient.	Significant differences between people with aphasia and those without. Item discriminant validity high (see point bi-serial correlations).	Moderate negative relationship between Global Communication Disorder ratings (1=mild, 2=moderate, 3=severe) and test scores of 127 participants in standardization sample (r=-.50)	DNT
Comprehensive Assessment of Spoken Language (CASL; Carrow-Woolfolk, 1999)	DNT	Internal consistency for 15 subtests and 6 index scores for the 12 age groups range = .64-.94; 37/252 (i.e., across all ages, subtests, and index scores) met strict criterion; 57/252 met relaxed criterion; core composite met strict criterion across all ages; all index scores met strict criterion except 3 receptive indexes (age 7, 8, 9) that	Test administered twice to 148 randomly selected examinees in 3 age groups: 5.0-6.11 (n=41), 8.0-10.11 (n=38), and 14.0-16.11 (n=69); interval between tests 7-109 days, median=6 weeks. All Core Composites met strict criterion; 4/5 Index Scores at age 8-10 and 3/3 at age 14-16 met strict criterion, 1/5 at age 8-10 met relaxed criterion.	Test purports to measure oral language; subtest content and structure is consistent with test model.	Developmental progression of scores: a table of means and standard deviations of test raw scores for the standardization sample by age "reveals steady increases in mean raw scores throughout the age range, with greater increases in the early years and more gradual changes in later years". One discrepancy noted in table: Nonliteral language subtest score for 19.0-21.11 was 4.3 (typo in manual?). Intercorrelations among subtests: coefficients range from .30 to .79. Factor structure of	Test of Auditory Comprehension of Language - Revised (ages 5.0-5.11) and CASL correlations ranged from .43 on syntax construction to .76 on pragmatic judgment. Oral and Written Language Scales (ages 7.10-10.11) vs. CASL correlations ranged from .34 on nonliteral subtest to .89 on the syntactic and expressive indices; Peabody	DNT

<u>Test</u>	<u>Inter-rater</u>	<u>Internal consistency</u>	<u>Test-retest reliability</u>	<u>Content/Face validity</u>	<u>Construct validity</u>	<u>Criterion-related validity:</u> <u>Concurrent</u>	<u>Criterion-related validity:</u> <u>Predictive</u>
		met relaxed criterion.			<p>Category and Processing Indexes Subtests developed and organized on the basis of the Integrative Language Theory.</p> <p>Factor analysis results consistent with design – e.g., 1 factor model for 3-6 and no index scores for this age, receptive vs. expressive factors for 7-21.</p> <p>The following groups scored significantly lower than age-matched controls: 50 children previously diagnosed with language delay, 42 children with language impairment, 44 children and adolescents with Mental Retardation, 30 12-18 year olds with a learning disability in reading, and a group of children with a variety of learning disabilities</p> <p>Scores of 50 8-10 year old children with speech impairments were not significantly lower than the control group.</p>	<p>Picture Vocabulary Test (Third Edition) and Expressive One Word Vocabulary Test (ages 7.1-10.11) vs. CASL correlations ranged from .45 on syntax construction to .85 on lexical/semantic index.</p> <p>Kaufman Brief Intelligence Test (ages 14.0-17.9) vs. CASL correlations ranged from .40 on pragmatic judgment to .81 on lexical/semantic index.</p>	

<u>Test</u>	<u>Inter-rater</u>	<u>Internal consistency</u>	<u>Test-retest reliability</u>	<u>Content/Face validity</u>	<u>Construct validity</u>	<u>Criterion-related validity: Concurrent</u>	<u>Criterion-related validity: Predictive</u>
Clinical Evaluation of Language Fundamentals – Third Edition (CELF-3; Semel, Wiig, & Secord, 1995)	Compared 2 randomly-selected raters on 590 Formulated Sentences and 599 Word Associations record forms at ages 6, 11 and 16 years. Formulated sentences: met strict criterion for age 6, relaxed criterion for age 11, $r=.70$ for age 16. Word Associations: met strict criterion for all three ages.	Receptive language: met relaxed criterion for 4/12 age groups, strict criterion for 8/12 age groups. Expressive language: met relaxed criterion for 3/12 age groups, strict criterion for 9/12 age groups. Total language: met strict criterion for all 12 age groups. Individual subtests variable: range from .54-.91 across subtests and age groups.	152 examinees ages 7, 10, and 13 years re-tested after 1-4 weeks. Receptive language: met relaxed criterion for 2/3 age groups; $r=.77$ for age 13. Expressive language: met strict criterion for 2/3 age groups, $r=.77$ for age 13. Total language: met strict criterion for 2/3 age groups, relaxed criterion for 1/3 age groups. Individual subtests variable: range from .45-.94 across subtests and age groups.	Choice of subtests is based on a model of language (form and content), and expressive vs. receptive language. Includes subtests that “focus on memory of spoken language,” although, purpose is not to test memory. In third edition, “gender and ethnic biases have been eliminated.” Each subtest based on relevant literature.	Factor structure supports single factor measured by test (i.e., receptive and expressive language are not differentiated in factor structure). Compared children with vs. without diagnosed language disorders, matched for age, gender, demographic characteristics ( $n=272$ ). Test correctly classified 71.3% of students as LD vs. non-LD.	Correlations between CELF-3 composite standard scores and Wechsler Intelligence Scales for Children Verbal, Performance and Full-Scale IQ ranged from .56-.75. CELF-3 classified students with TBI who were receiving speech and language services as LD but not students with communication disorders who were not receiving services; subtest score variability not significantly greater in students with TBI (Turkstra, 1999).	DNT
Cognitive Linguistic Quick Test (CLQT; Helm-Estabrooks, 2001)	Clock Drawing: 2 raters, 170 tests, met relaxed criterion. Generative Naming: 2 raters, 81 tests, met strict criterion.	DNT	Normal sample ( $n=46$ ) re-tested 80-140 days from first test: Across 10 tasks, test-retest correlations ranged from .03-.81 with 1/10 tasks meeting relaxed criterion. Across 5 cognitive domains correlations ranged from .61-.90: 1/5 (Language) met relaxed criterion, 1/5 (Executive	Based on a framework that cognitive functions are needed in daily life; 5 functions identified and tasks designed to test them. Items based on a review of the literature on cognitive function and other tests. Author’s expert opinion determined percent each task contributed to each cognitive domain; author reviewed standardization sample to confirm this. Spanish version developed concurrently and included only items that would be	Confirmatory factor analysis completed and goodness-of-fit statistics $\geq .97$ across items from 5 cognitive domains. Clock drawing correlated .74 with total composite. Age and severity cutoffs determined by author based on examination of normal subjects and subjects with neurological impairments. Significant difference between clinical and normal subjects matched by age group, education, sex.	Author reviewed patient charts to confirm diagnosis. Data presented by etiology but diagnostic utility not directly tested.	DNT

<u>Test</u>	<u>Inter-rater</u>	<u>Internal consistency</u>	<u>Test-retest reliability</u>	<u>Content/Face validity</u>	<u>Construct validity</u>	<u>Criterion-related validity: Concurrent</u>	<u>Criterion-related validity: Predictive</u>
			Functions) met strict criterion. Average absolute score differences were less than 1 point for 8/10 tasks – i.e., statistics limited in normal subjects because of ceiling effects and relatively small number of points available for some questions. Absolute score differences on cognitive domains ranged from 1.22-16.57. Did not test in clinical samples.	familiar to both English and Spanish speakers. Stories designed to use everyday vocabulary and minimize cultural or socioeconomic bias, and reviewed by Spanish speakers with modifications based on pilot study feedback.			
Children's Orientation and Amnesia Test (COAT; Ewing-Cobbs, Levin, Fletcher, Miner, & Eisenberg, 1990)	During the sub-acute stage of recovery in 11 children with TBI, compared scores obtained by 2 examiners: met strict criterion.	DNT	DNT	Developed by experts in pediatric TBI. Based on an adult test (Galveston Orientation and Amnesia Test) that has established reliability and validity. No effect of sex on scores. Test based on current concepts of anterograde and retrograde amnesia.	Scores increase with age from 3 to 15 years.	Glasgow Coma Scale (GCS) scores significantly higher in subjects with post-traumatic amnesia (PTA) duration $\leq 7$ days (Group 1, n=21) than in subjects with PTA duration 8-14 days (Group 2, n=7) or PTA duration $> 14$ days (Group 3, n=9). Fewer focal lesions in Group 1 (52%) than in Group 2 (57%) and Group 3 (66%). PTA as determined by the COAT correlated - .61 with admission	PTA duration significantly correlated with nonverbal CLTR at 6 and 12 months. Distribution of Glasgow Outcome Scale scores at 1 year post-TBI differed significantly according to PTA duration.



<u>Test</u>	<u>Inter-rater</u>	<u>Internal consistency</u>	<u>Test-retest reliability</u>	<u>Content/Face validity</u>	<u>Construct validity</u>	<u>Criterion-related validity: Concurrent</u>	<u>Criterion-related validity: Predictive</u>
						GCS score and .48 with duration of unconsciousness. PTA as determined by the COAT correlated significantly with verbal consistent long-term retrieval (CLTR) at baseline.	
Discourse Comprehension Test (DCT; Brookshire & Nicholas, 1993)	DNT	Performance of 40 control subjects was "virtually identical" across 2 sets of 5 stories. No statistics calculated.	Met relaxed criterion for aphasic subjects and strict criterion for subjects with right hemisphere damage.	Design based on test authors' belief that a test of discourse comprehension must assess comprehension and retention of both main ideas and details, as well as comprehension of both directly stated and implied info. 10 judges (graduate students and speech-language pathologists) verified that questions actually tested stated and implied main ideas and details. Passage Dependency Index (PDI) calculated for each story based on the performance of 10 adults without brain damage. 9/10 PDI scores were at or above the level "generally considered acceptable" (.40).	Analysis of word frequency and word types showed that DCT content words are more like adult-to-adult conversations than newspapers.	Pearson product-moment correlation coefficients calculated for DCT total scores vs. auditory comprehension subtests of the Boston Diagnostic Examination of Aphasia ( $r=.76$ ), the 4 subtest version of the Porch Inventory of Communicative Abilities ( $r=.64$ ), and the Sentence Completion Test ( $r=.53$ ) for 20 aphasic adults. Authors cite Katsuki-Nakamura, Brookshire, & Nicolas (1988) as evidence that DCT predicts comprehension in daily life: Purpose was to compare paragraph	

<u>Test</u>	<u>Inter-rater</u>	<u>Internal consistency</u>	<u>Test-retest reliability</u>	<u>Content/Face validity</u>	<u>Construct validity</u>	<u>Criterion-related validity:</u> <u>Concurrent</u>	<u>Criterion-related validity:</u> <u>Predictive</u>
						comprehension for stimuli read as monologues vs. dialogues. Study stimuli were DCT precursors. For aphasic subjects, no significant correlation of total errors between stories read as monologues and stories read as dialogues. No significant correlation of main ideas recalled for monologues vs. dialogues, significant correlation of details recalled in monologues vs. dialogues; both for aphasic and control subjects combined.	
Galveston Orientation and Amnesia Test (GOAT; Levin, O'Donnell, & Grossman, 1979)	Ratings by 2 examiners who alternated interviewing patients demonstrated a "significant level" of inter-observer agreement (no statistics reported).	DNT	DNT	Developed by experts based on current concepts of memory and stages of recovery after injury.	DNT	In 32 patients "association was highly significant to duration of post-traumatic amnesia, as measured by the GOAT, and acute neurologic impairment, measured by eye, motor, and verbal components of the Glasgow Coma Scale." Serial GOAT scores strongly correlated with the	Most patients with post-traumatic amnesia (n=18) of greater than 2 weeks achieved good recovery. Longer periods of post-traumatic amnesia (n=18) were

<u>Test</u>	<u>Inter-rater</u>	<u>Internal consistency</u>	<u>Test-retest reliability</u>	<u>Content/Face validity</u>	<u>Construct validity</u>	<u>Criterion-related validity: Concurrent</u>	<u>Criterion-related validity: Predictive</u>
						severity of acute cerebral disturbance after TBI.	frequently followed by prolonged disability.
LaTrobe Communication Questionnaire (LCQ; Douglas, O'Flaherty, & Snow, 2000)	DNT	Item analysis results met relaxed criterion.	Tested 24 primary subjects and 24 close others (self-selected conversation partners) over an 8-week period: Did not meet criteria. For close others, authors cited inconsistency of subject's choice of close other as reason for low reliability.	Developed for use with individuals with brain damage and close others. In primary subjects (i.e. vs. close others), age and education did not account for significant percent of variance in self-perceived communication ability. Females reported significantly fewer difficulties than males with 2 items: keeping track of main details of conversation, changing speech style according to situation. No effect of age in adults tested.	Completed using self-report data as sample size for close others was insufficient. Factor analysis based on 147 responses yielded 6 factors accounting for 49% of variance: conversational tone, conversational effectiveness, conversational flow, conversational engagement, conversational (partner) sensitivity, and conversational attention/focus.	Age and social desirability poorly correlated with LCQ scores.	DNT

<u>Test</u>	<u>Inter-rater</u>	<u>Internal consistency</u>	<u>Test-retest reliability</u>	<u>Content/Face validity</u>	<u>Construct validity</u>	<u>Criterion-related validity: Concurrent</u>	<u>Criterion-related validity: Predictive</u>
Mount Wilga High Level Language Test (Christie, Clark, & Mortensen, 1986)	DNT	DNT	DNT	Developed by individuals working with TBI clients with mild language problems. Designed for assessment of linguistic skills without ceiling effects of standard aphasia tests, and with consideration of communication efficacy and appropriateness (i.e. impairments that are common after TBI). Revisions based on a normative study (Beverley & Scott, 1986), but current version not normed; authors stated need for collection of normative data.	DNT	DNT	DNT
Repeatable Battery for the Assessment of Neuropsychological Status (RBANS; Randolph, 2001)	3 scorers on 20 random Figure Copy/Recall subtests from standardization sample. Met relaxed criterion.	Booklet A vs. B: repeated measures ANOVA of Index scores showed no significant difference for any age except Language Raw Score – Semantic Fluency increased 4 points on second test and scores for Form B adjusted to compensate. Split half: Total scale scores met strict criterion for all 6 age groups; 3/5 index scores met relaxed criterion for all age groups; 2/5 index scores met	Total score met relaxed criterion. Index scores did not meet criteria.	Designed to meet 3 test goals: 1) alternative forms developed for <u>repeated testing</u> , 2) overall length permits use as <u>screening</u> and portable test, 3) range of item difficulty so it can be used as a <u>core test</u> for dementia. Tests cognitive areas known to be affected in dementia. Existing tests are excessively difficult for older subjects – designed to meet needs of that population and be sensitive enough to detect mild impairment. Tasks used based on a review of the literature on dementia. Standardized on younger subjects because it became apparent that it could also be used to	Calculated statistical probability of differences between index scores at $p > .15$ and $p > .05$ , and base rates at which differences occurred. Intercorrelations among subtests range from .28-.44, and among index scores range from .20-.40 – i.e., measuring different constructs. Preliminary comparison to data from Wechsler Adult Intelligence Scales (WAIS) Revised, Wechsler Memory Scales Revised, Benton Judgment of Line Orientation, Complex Figure Test, Wide Range Achievement Test Third Edition, Boston Naming Test, Controlled Oral Word Association – but sample sizes small and	DNT	DNT

<u>Test</u>	<u>Inter-rater</u>	<u>Internal consistency</u>	<u>Test-retest reliability</u>	<u>Content/Face validity</u>	<u>Construct validity</u>	<u>Criterion-related validity: Concurrent</u>	<u>Criterion-related validity: Predictive</u>
		relaxed criterion for 4/6 age groups.		screen cognitive function in younger patients e.g. acute care, tracking recovery after injury, neuropsychological screening for non-neuropsychologists.	patient groups had mixed etiologies: largest group was for WAIS-Full Scale IQ vs. RBANS and correlation between these was .78.		
Rivermead Behavioural Memory Test (RBMT; Wilson, Baddeley, Cockburn, & Hiorns, 1985)	2 raters scoring 40 subjects met strict criterion.	DNT	Tested by comparing Version A vs. B, C, D. For Screening scores: A vs. B and A vs. C met strict criterion. For Standardized profile scores: all met relaxed criterion.	Test developed by experts and based on the idea that patients with brain have substantially more memory problems than controls. Designed to represent everyday memory demands.	Overall level of performance between patients and controls revealed that patients have substantially lower scores than the control subjects. Performance on RBMT compared to performance on standardized memory tests: Warrington's (1984) Recognition Memory Test for words and faces, Digit Span, Spatial Span, and the Paired-Associate Learning subtest taken from Randt, Brown and Osborne (1980). Results showed that the RBMT had moderate but significant correlations with all of the tests and can be regarded as measuring overall memory performance when assessed in terms of existing tests. Designed to reflect everyday memory, and performance on the RBMT correlated with observation made by therapists on memory	DNT	DNT

<u>Test</u>	<u>Inter-rater</u>	<u>Internal consistency</u>	<u>Test-retest reliability</u>	<u>Content/Face validity</u>	<u>Construct validity</u>	<u>Criterion-related validity: Concurrent</u>	<u>Criterion-related validity: Predictive</u>
					lapses of 80 patients for a mean of 35 hours of observation per patient. Correlation was -0.71 for the screening score and -0.75 for the standardized profile score. Results suggest that the RBMT offers a good measure of everyday memory performance		
Ross Information Processing Ability, Second Edition (RIPA-2; Ross-Swain, 1996)	3 clinicians independently scored a videotaped test administration. Inter-scorer agreement met strict criterion.	Analyzed using 126 protocols. 4/10 subtests met strict criterion, 5/10 met relaxed criterion.	DNT	Test construction based on theoretical framework of cognitive-linguistic function. Rationale for test format and item/subtest selection based on the literature and expert opinion. Panel of three professionals assessed face validity. Item analysis "satisfied requirements."	Inter-subtest correlations relatively high ( $r$ 's = .56-.89). Factor analysis reveals that RIPA-2 is a one-factor test, with subtest factor loadings of .56-.89. Does not support reporting of individual subtest scores. Test stated purpose is to "quantify and qualify cognitive-linguistic deficits" and permit clinician to analyze "task-by-task performance." Not supported by the test's single factor structure.	RIPA-2 subtests compared to selected subtests of the Woodcock Johnson Tests of Cognitive Ability. Coefficients range from .02 – .78.	DNT
Scales of Cognitive Ability for Traumatic Brain Injury (SCATBI; Adamovich & Henderson, 1992)	DNT	Based on 164 tests. Raw scores on the five subscales met strict criterion.	33 subjects with TBI retested within 8 months. Statistically significant increase in mean scores on all five subscales. 1/5 subtests met strict criterion, 4/5 did not meet relaxed criterion.	Not based on any single model of cognitive function. Instead, constructed to have hierarchy of cognitive processes that typically are impaired after TBI. Based on TBI literature and child cognitive development research on progression of cognitive task difficulty. Use of principles from child development based on premise that cognitive development and cognitive	Significant but low correlation between 5/5 subscale scores and age in TBI subjects; 1/5 subtest scores in non-injured subjects. Factor analysis yielded 4-factor solution that supports subtest structure. Discriminant analysis on 164 TBI subjects vs. 78 demographically matched non-injured subjects,	Correlations between raw scores on five subscales and Rancho Levels = .50-.60, all statistically significant. Significant increase in 4/5 subscale scores between Rancho Levels of Cognitive	DNT

<u>Test</u>	<u>Inter-rater</u>	<u>Internal consistency</u>	<u>Test-retest reliability</u>	<u>Content/Face validity</u>	<u>Construct validity</u>	<u>Criterion-related validity: Concurrent</u>	<u>Criterion-related validity: Predictive</u>
				<p>recovery progress in similar hierarchy over time. Task difficulty increases in multiple domains – e.g. stimulus complexity, time requirement. Target areas (e.g. attention, memory) are those that must improve through treatment, per the authors. Items based on authors' detailed study of changes on 39 standardized and nonstandardized measures, in 14 patients with TBI over six months of treatment vs. 9 uninjured controls. Scales organized around common cognitive and perceptual functions with similar items. Included both auditory and visual stimuli, with "reasonable balance where possible".</p>	<p>using subscale raw scores: 2 subscales accounted for 50% of variance in group membership; remaining 3/5 add nothing; classification rate = 95.7 accuracy for non-injured, 79.2 for TBI subjects</p>	<p>Functioning (LCF) V and VI, and VI and VII. No significant differences between LCF IV vs. V, and VII vs. VIII. Low but significant correlation between 4/5 subscale scores and days post-injury.</p>	
<p>The Speed and Capacity of Language Processing Test (SCOLP; Baddeley, Emslie, &amp; Nimmo-Smith, 1992)</p>	<p>DNT</p>	<p>Internal consistency not measured. In 25 normal adults, correlation between two halves of the Speed of Comprehension test met relaxed criterion at test and at re-test.</p>	<p>25 subjects re-administered Speed of Comprehension an average of 19.4 days on Version A vs. Version B. Met strict criterion. 50 subjects completed Spot-the-Word Version A and Version B. Correlation did not meet relaxed criterion. 224 normal adults in standardization sample: correlation</p>	<p>Based on the premise that brain damage slows the rate of information processing. Developed by experts in cognitive psychology and rehabilitation. Based on a measure originally designed to test retrieval from semantic memory. Vocabulary is subtracted from speeded processing, to "differentiate between a subject who has always been slow and a subject whose performance has been impaired as a result of brain damage or some</p>	<p>In 76 normal subjects, significant correlation between Speed of Comprehension Test and other measures of speeded verbal fluency: category generation <math>r=.52</math>, color naming <math>r=.56</math>; letter-matching tasks <math>r=.40</math> (name match), <math>r=.34</math> (physical match); vocabulary <math>r=.51</math>. In 50 normal subjects, Spot-the-Word Test correlated with Mill Hill Vocabulary test .60 for Version A, .71 for Version B. In 224 adults in</p>	<p>In 76 normal subjects, correlation between Speed of Comprehension and Raven's Standard Progressive Matrices <math>r=.20</math>; age <math>r=.03</math>. Higher correlation with Vocabulary than measure of general intelligence, in normal elderly subjects (n not stated). Interpreted as</p>	<p>DNT</p>

<u>Test</u>	<u>Inter-rater</u>	<u>Internal consistency</u>	<u>Test-retest reliability</u>	<u>Content/Face validity</u>	<u>Construct validity</u>	<u>Criterion-related validity: Concurrent</u>	<u>Criterion-related validity: Predictive</u>
			between Speed of Comprehension Version A and Version B met relaxed criterion; between Spot-the-Word Version A and Version B met relaxed criterion.	other stressor.” Lexical decision (Spot-the-Word) task designed to be taken without the need to read aloud (e.g. pronounce) words, given in a group, adapted to any language, minimize embarrassment (via silent reading). Significant effects of age, social class, and sex on Speed, and of age and social class on Spot.	standardization sample, correlation between Speed of Comprehension and vocabulary (.60) and grammatical reasoning (.60). In 224 normal adults, correlation between Speed and Spot = .57 – i.e., measuring a different but related construct.	evidence that test is not measuring general intelligence.	
The Awareness of Social Inference Test (TASIT; McDonald, Flanagan, Kinch, & Rollins, 2002)  <u>Subtest abbreviations used:</u> ER = Emotion recognition  SI-M = Social Inference –Minimal  SI-E = Social Inference – Enriched	DNT	Correlations not calculated. ER: Form A and Form B have no significant difference in normal subjects; scores for individual emotions differed between tests. SI-M and SI-E: Form A and Form B have no significant difference in normal subjects; scores for individual inference items differed between tests.	Practice effects studied by re-administration in a single setting to subgroups of 127 normal subjects from development sample, and 6 months later to 6 adults from community. Results analyzed with sign test; correlations not calculated. Immediate re-test: all parts except Form B Emotions improved, range of point increase = 1-6 (max possible score = 28 for emotions, 20 for social inference both subtests combined); no significant change 6 months later.	Test developed based on literature on sequelae of TBI and existing measures of social cognition and emotion. Items developed with feedback from three consecutive cohorts of normal viewers, with revisions based on each. 7 viewers with TBI gave feedback re: clarity of instructions. Significant correlation between education (-.41) and age (.53) on ER Form B, no significant correlation on Form A or with intelligence and scores on either Form A or B. No significant correlation of age or education on SI subtests, but significant correlation with intelligence and scores on both SI-M Form A (.37) and B (.49) and SI-E Form A (.36) and Form B (.38).	12 adults with severe TBI <u>M</u> =4 weeks post emergence from post-traumatic amnesia, all with normal performance on language, facial recognition, immediate memory and information processing speed; compared to 12 age- and education-matched controls: significant differences in total scores for Emotion and SI-E, not SI-M.	DNT	DNT
Test of Everyday Attention for Children (TEA-Ch; Manly, Robertson, Anderson,	DNT	DNT	Correlation coefficients reported for 9 subtests based on	Based on: 1) published research in adults suggesting that attention is a multi-process construct	Factor analysis performed with three factor model determined a prior (selective, sustained,	24 boys with DSM IV-based diagnosis of attention deficit	DNT



<u>Test</u>	<u>Inter-rater</u>	<u>Internal consistency</u>	<u>Test-retest reliability</u>	<u>Content/Face validity</u>	<u>Construct validity</u>	<u>Criterion-related validity:</u> <u>Concurrent</u>	<u>Criterion-related validity:</u> <u>Predictive</u>
& Nimmo-Smith, 1999)			55 children. Test-retest interval not stated. Correlation coefficients met relaxed criterion for Sky Search time-per-target, Sky Search DT decrement, Same World time, Opposite World Time. Percent agreement reported for 3 subtests with ceiling effects. Range = 71.0-76.2%. Reliable change scores calculated for practice effects from Version A to Version B on children retested between 5-20 days (n not stated) and presented in Table 13.	(selective, sustained, and spatial); 2) analysis of the Test of Everyday Attention Adult version suggesting at least three different attentional factors (selective, sustained, switching), 3) research on development of attention in children; and 4) well-validated measures of attention in adults. Bias addressed indirectly in factor analysis (i.e. factor analysis should be affected by confounds)	switching attention). Comparative fit index for this model = .97, Normed Fit index = .91 and Non-Normed Fit Index = .96 (all indicate good fit). Regression coefficients for individual subtests on each factor ranged from .44-.79. Creature Counting, Sky Search time-per-target and Same World time not included in the factor analysis (no reason given). 96 children also received Stroop Test, Trails, Matching Familiar Figures Test. Correlations are modest and in directions predicted by test model. As predicted, no significant correlation with Wechsler Intelligence Scales for Children Third Edition (WISC-III). Small but significant correlations between some subtests and aspects of achievement.	hyperactivity disorder had significantly lower scores than vocabulary- and age-matched controls from standardization sample on all subtests except Sky Search. Same with standardization sample children matched on Block Design Scores. Significant differences between 18 children with TBI and 18 age and sex-matched controls on all subtests but Score and Creature Counting time, with no significant difference in scores on WISC-III Block Design or Object Assembly (groups differed significantly on Vocabulary and Similarities scores).	

<u>Test</u>	<u>Inter-rater</u>	<u>Internal consistency</u>	<u>Test-retest reliability</u>	<u>Content/Face validity</u>	<u>Construct validity</u>	<u>Criterion-related validity: Concurrent</u>	<u>Criterion-related validity: Predictive</u>
Test of Language competence (TLC; Wiig & Secord, 1989)	Based on 16 raters for Level 1 and 8 raters for Level 2, rating 2 protocols each. Subtests 3 and 4 met strict criterion for Levels 1 and 2.	Level 1 composite score met strict criterion for all 8 age groups except 8-0 to 8-5 and 9-6 to 9-11, which met relaxed criterion. Level 2 composite score reliabilities met relaxed criterion for oldest 3/8 age groups and did not meet criterion for 5/8 age groups.	For Level 1, based on 88 students ages 6 and 8 years with there was 4-8 week test-retest interval. 1/7 subtests and composite score met strict criterion; 5/7 subtests met relaxed criterion. For Level 2, based on 92 students ages 9-18 years with 3 week test-retest interval. None met criteria.	Designed to measure a single construct termed "language competence". Subtests viewed as a series of tasks designed to discriminate between competent and non-competent language use. Subtests may be categorized according to a model of language competence. Model focuses on the levels of semantics, semantic-syntactic interfaces, and pragmatics. Within these levels, the model identifies the content of each of the subtests as propositions in narrow contexts or propositions in communication-like contexts.	Most subtest intercorrelations moderate (.17-.57). Factor analysis performed on subtest correlation matrix for each age interval. For Level 2, more than 86% of common variance explained by the first unrotated factor. For Level 1, all common variance explained by the first factor at each age. Should be viewed as a one-factor test. With factor analysis of item intercorrelation matrix, Level 1 range from .36-.55 and Level 2 range .27-.49. Classification of Language Learning Disability (LLD) compared between TLC-E vs. Test of Adolescent Language (TOAL): LLD = 96% for TLC-E vs. 89% for TOAL; non-LLD = 89% for TLC-E vs. 86% for TOAL. Discriminated children with vs. without TBI in subsequent study (Towne & Entwisle, 1993).	Compared to a priori classification of students by school districts as LLD or normally achieving. Accuracy of TLC = 96% LLD, 93% non-LLD. Correlations between the TLC subtests and Wechsler Intelligence Scales for Children Revised (WISC-R), Test of Language Development Second Edition, Peabody Picture Vocabulary Test Revised (PPVT-R) as predicted by content – i.e. relatively high with verbal measures (e.g. .39-.55 with PPVT) and low with nonverbal measures (e.g., .07-.12 with WISC-R Performance IQ).	DNT

<u>Test</u>	<u>Inter-rater</u>	<u>Internal consistency</u>	<u>Test-retest reliability</u>	<u>Content/Face validity</u>	<u>Construct validity</u>	<u>Criterion-related validity: Concurrent</u>	<u>Criterion-related validity: Predictive</u>
Western Aphasia Battery (WAB; Kertesz, 1982)	Based on 3 raters rating 10 videotaped patients at two time points 6 months apart. 9/9 subtests met strict criterion.	Met strict criterion.	7/9 subtests and Aphasia Quotient met strict criterion; 1/9 subtest and Cognitive Quotient met relaxed criterion; Praxis subtest did not meet criterion.	Used clinical and neurolinguistic principles, concepts, and some of the subtests developed by Goodglass and Kaplan. WAB "tests all language modalities and is comparable in content with other batteries assessing language impairment in aphasia; praxis and construction subtests, while not included in all standard aphasia batteries, provide an assessment of more general higher-cortical functioning". Assesses language content areas common to all aphasia batteries. Items within subtests similar to those of the BDAE.	15 subjects administered WAB and NCCEA within 2-week interval. Correlations ranged from .82 to .92. 2 summary comparisons computed for comparable WAB and Neurosensory Center Comprehensive Examination for Aphasia subtests, correlation = .97. Normal control subjects (n=31) differed significantly from 70 individuals with non-dominant-hemisphere lesions, 17 subjects with diffuse brain lesions, and 117 subjects with acute aphasia except anomic group. Non-left hemisphere groups did not differ significantly among themselves. Across aphasia subtypes, all groups differed significantly except Broca's vs. Wernicke's aphasic types.	Scores related to site of lesion – i.e. subjects with right hemisphere lesions did not test in aphasic range. AQ correlation with Raven's Standard Progressive Matrices scores = .55 for 140 patients.	DNT

## References

- Adamovich, B., & Henderson, J. (1992). *Scales of Cognitive Ability for Traumatic Brain Injury* (Normed ed.): The Riverside Publishing Company.
- Baddeley, A., Emslie, & Nimmo-Smith, I. (1992). *Speed and Capacity of Language Processing Test* (First ed.). Suffolk, England: Thames Valley Test Company.
- Brookshire, R., & Nicholas, L. (1993). *Discourse Comprehension Test* (First ed.): BRK Publishers.
- Carrow-Woolfolk, E. (1999). *Comprehensive Assessment of Spoken Language* (First ed.). Circle Pines, MN: American Guidance Service, Inc.
- Christie, J., Clark, W., & Mortensen, L. (1986). *Mount Wilga High Level Language Test* (First ed.). Sidney, NSW: Mt. Wilga Rehabilitation Centre: Unpublished Manuscript.
- Douglas, J., O'Flaherty, C., & Snow, P. (2000). Measuring perception of communication ability: The development and evaluation of the La Trobe communication questionnaire. *Aphasiology*, 14, 251-268.

- Ewing-Cobbs, L., Levin, H., Fletcher, J., Miner, M., & Eisenberg, H. (1990). The Children's Orientation and Amnesia Test: Relationship to severity of acute head injury and to recovery of memory. *Neurosurgery*, 27, 683-691.
- Frattali, C., Thompson, C., Holland, A., Wohl, C., & Ferketic, M. (1995). *American Speech Language Hearing Association Functional Assessment of Communication Skills for Adults* (First ed.). Rockville, MD: American Speech Language Hearing Association.
- Gioia, G., Isquith, P., Guy, S., & Kenworthy, L. (1996). *Behavior Rating Inventory of Executive Function* (First ed.). Odessa, FL: Psychological Assessment Resources, Inc.
- Helm-Estabrooks, N. (1992). *Aphasia Diagnostic Profiles* (First ed.): Applied Symbolix.
- Helm-Estabrooks, N. (2001). *Cognitive Linguistic Quick Test* (First ed.). San Antonio, TX: Psychological Corporation.
- Helm-Estabrooks, N., & Hotz, G. (1991). *Brief Test of Head Injury* (First ed.): Riverside Publishing Company.
- Holland, A., Frattali, C., & Fromm, D. (1999). *Communication Activities of Daily Living* (Second ed.). Austin, TX: Pro-Ed, Inc.
- Kertesz, A. (1982). *Western Aphasia Battery* (First ed.). San Antonio, TX: The Psychological Corporation.
- Levin, H., O'Donnell, V., & Grossman, R. (1979). The Galveston Orientation Amnesia Test: A practical scale to assess cognition after head injury. *Journal of Nervous and Mental Disease*, 167, 675-684.
- Manly, T., Robertson, I., Anderson, V., & Nimmo-Smith, I. (1999). *The Test of Everyday Attention for Children* (First ed.). Suffolk, England: Thames Valley Test Company.
- McDonald, S., Flanagan, S., Kinch, J., & Rollins, J. (2002). *The Awareness of Social Inference Test* (First ed.). Suffolk, England: Thames Valley Test Company.
- Randolph, C. (2001). *Repeatable Battery for the Assessment of Neuropsychological Status* (First ed.). San Antonio, TX: Psychological Corporation.
- Ross-Swain, D. (1996). *Ross Information Processing Assessment* (Second ed.). Austin, TX: Pro-Ed.
- Semel, E., Wiig, E., & Secord, W. (1995). *Clinical Evaluation of Language Fundamentals* (Third ed.). San Antonio, TX: The Psychological Corporation.
- Towne, R. L., & Entwisle, L. M. (1993). Metaphoric comprehension in adolescents with traumatic brain injury and in adolescents with language learning disability. *Language, Speech and Hearing Services in Schools*, 24, 100-107.
- Turkstra, L. (1999). Language testing in adolescents with brain injury: A consideration of the CELF-3. *Language, Speech, and Hearing Services in Schools*, 30, 132-140.
- Wiig, E., & Secord, W. (1989). *Test of Language Competence-Expanded Edition* (Expanded ed.). San Antonio, TX: Psychological Corporation.
- Wilson, B., Baddeley, A., Cockburn, J., & Hiorns, R. (1985). *The Rivermead Behavioural Memory Test* (Second Supplement ed.): Western Psychological Services.