Summary of Reliability and Validity Information for Cognitive, Nonverbal IQ, Memory and Learning, Language, Visual-Motor, Attention and Achievement Measures (Cognitive Assessment Workgroup; Campbell, Brown, Cavanagh, Vess, & Segall):

		Reliability Evidence			Validity Evidence			
Domain	**Measure**	**Internal Consistency**	**Test-retest**	**Other (e.g., interrater)**	**Criterion: Concurrent**	**Criterion: Group differences**	**Construct: Structural / Factor Analytic**	**Other (e.g., predictive; divergent)**
Cognitive	Bayley Scales of Infant and Toddler Development, Third Edition (Bayley-III)	Cognitive, Language, and Motor scales, split-half: .91 - .93 General Adaptive Composite (GAC), split-half: 97.	Cognitive, Language, and Motor scales across all ages r=.81 to .87 GAC across all ages =.86 to .92	Inter-rater reliability: r=.82 (GAC); r=.79 (adaptive behavior domains); r=.73 (skill areas) Clinical populations , split-half: .94 - .98 across subtest	Manual reports correlations with WPPSI-III, PLS-4, PDMS-2, and ABAS-II-P r=.71 (Language composite with Mental Index of BSID-II) r=.60 (Motor composite with Motor index of BSID-II)	Sensitive to differences between the normative sample and children at risk for developmental delay (specific language impairment; Cerebral Palsy; Down Syndrome; Pervasive Developmental Disorder)	CFA supports a 3-factor model (Cognitive, Language, and Motor) as the best fit across all age groups	
Cognitive	McCarthy Scales of Children's Abilities (MSCA)	r=.90 - .96 (General Cognitive Index (GCI), across age groups) r=.79 to .88 (Average of Index scales across age groups)	r = .84 - .91 for three age groups for GCI	α=.93 for Spanish speaking sample	r=.81 with WPPSI-R Strong correlation with Binet scales (for K and 1st grade)	Differentiates between learning disabled and non-learning disabled Differentiated developmentally disabled sample	Manual reports support for a 5-factor model (Verbal, Perceptual-performance, Quantitative, Motor and Memory)	Strongly correlated with CAT scores (K, 1st, and 2nd grade) Substantial correlations with a parent-administered test of cognitive abilities
Cognitive	Kaufman Assessment Battery for Children, 2nd Edition (KABC-II)	Global scales (e.g., MPI) Split-half: .90-.97 Scales r=.88 - .93 Subtests r=.69 - .93	Global scale index stability: r=.72 to .94 Scale stability: r=.74 to .95 Subtest stability: r=.50 to .92		Strong correlations with K-ABC, WISC-III, WPPSI-III, KIAT and WJ-III-Cog r=.89 with WISC-IV FSIQ and K-ABC-2 FCI; r=.88 with WISC-IV FSIQ and K-ABC-2 MPI	Differences between clinical samples (SLD, ADHD, and Autism) and normative group, but no distinct pattern between the scales	CFA supports both a 4-factor model (Luria's neuropsychological model) and a 5-factor model (CHC theory)	Moderate correlations with PIAT-R, WJ-III-Ach, and WIAT-II

Cognitive	Stanford-Binet Intelligence Scales, Fifth Edition (SB-5)	IQ (Split-half): r=.98 (Full); r =.95 (Nonverbal); r =.96 (Verbal); r =.91 (Abbreviated) Factor Index: r =.90 to .92 Subtests r=.84 to .89	IQ: r=.92 to .95 (2 to 5 year olds); r=.90 to .93 (6 to 20 year olds) Factor: r=.82 to .92 (2 to 5 year olds); r=.85 to .92 (6 to 20 year olds) Subtests: r=.76-.91 (2-5); r=.76-.91 (6-20)	Median inter-rater reliability: r=.90 (range: .74 to .97) SEM: 2.30 (FSIQ); 3.26 (Nonverbal IQ); 3.05 (Verbal IQ)	r=.90 with SB-4; r=.83 with WPPSI-R; r=.84 with WISC-III; r=.78 with WJ-III Cog	Evidence of distinctive score profiles for Developmental Delay, Autism, ADHD, Motor Delay, and others	5-factor model was best fit for the scale	r=.57 (correlation between SB-5 Nonverbal IQ and UNIT in a hearing-impaired sample) r=.66 to .84 with WJ-III Ach; r=.80 with WIAT-II total score
Cognitive	Wechsler Adult Intelligence Scale, Third Edition (WAIS-III)	IQ (Split-half) r=.98 (Mdn Full); r=.97 (Mdn VIQ); r=.94 (Mdn PIQ) Factor Index: r =.88 to .96 (Mdn = .93) Subtests r=.70 to .93 (Mdn = .86)	IQ (*M* all age groups): r=.96 (FSIQ); r=.96 (VIQ); r=.91 (PIQ) Factor Index: r =.88 to .95 (Mdn = .89) Subtests r=.69 to .94 (Mdn = .82)	Inter-rater reliability: r=.91 to .95 (Vocabulary, Similarities, Comp. subtests)	r=.88 with WISC-III FSIQ r=.65 with Standard Progressive Matrices r=.88 with SB-IV	Manual reports special group studies for MR, ADHD, LD, deaf and hearing-impaired, and other groups. Distinctive score profiles were found for these groups	CFA supports a 4-factor model: VCI, POI, WMI and PSI	r=.42 with WCST and WAIS-III FSIQ
Cognitive	Wechsler Preschool and Primary Scale of Intelligence-Third Edition (WWPSI-III)	IQ (*M* all age groups): r=.96 (FSIQ); r=.95 (VIQ); r=.93 (PIQ) Other Index: r =.89 (Processing Speed) & .93 (General Language) Subtests r=.83 to .95 (Mdn = .88)	IQ (*M* all age groups): r=.92 (FSIQ); r=.91 (VIQ); r=.86 (PIQ) Other Index: r =.90 (Processing Speed) & .91 (General Language) Subtests r=.74 to .90 (Mdn = .83)	Inter-rater agreement: r=.98 to .99 (Vocabulary, Similarities, and Comprehension subtests)	r=.70 to .86 with WPPSI-R; r=.69 to .89 with WISC-III r=.38 to .87 with DAS	Distinctive score profiles were found for Developmental Delays, Autism, and Motor impairment, among others	EFA supports a 2-factor structure for ages 2:6 to 3:11; 3-factor structure for ages 4:0 to 6:0; for ages 6:0 to 7:3, PC subtest loaded on Verbal factor rather than Performance factor In a sample of language impaired children, a 4-factor model was found	r=.31 to .78 with WIAT-II

Cognitive	Woodcock-Johnson III Tests of Cognitive Abilities (WJ-III Cog)	GIA: r=.96-.98 (Mdn = .97); BIA: r=.94-.98 (Mdn=.95) Standard battery cluster scores: r=.90-.95 Subtests (Mdn across ages): r= .74-.97	Thinking Abil r=.68 to .83 (Mdn = .73); Cog. Efficiency r=.60 -.86 (Mdn = .78) Speeded Tests r=.73 -.87 (Mdn = .78 for 7-17 y.o.)		r=.67 with DAS GCA; r=.73 with WPPSI-R	A sample of ADHD children showed ADHD-PI and Combined type had most difficulty on Concept Formation task With the WJ-III Ach, WJ-III Cog differentiates LD from non-LD students	Factor structure is supported across a wide range of age groups (age 6 to 90)	Cognitive clusters and WJ-III Ach clusters correlated in the .70s BIA had moderate correlations with PKRS-II (r=.59)
Cognitive	Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV)	IQ (*M* all age groups): r=.97 (FSIQ) Other Index: r =.88 (PSI) - .94 (VCI) (Mdn = .92) Subtests r=.79 to .90 (Mdn = .86)	IQ (*M* all age groups): r=.93 (FSIQ); Other Index: r =.86 (PSI) - .93 (VCI) (Mdn = .89) Subtests r=.76 to .92 (Mdn = .83)	Interscorer agreement ranged from .95 - .98	r=.89 with WISC-III r=. 89 with WPPSI-III	Group studies support the validity and clinical utility of WISC-IV. Groups include Motor Impairment, Autism, TBI, ADHD, and LD, among others	CFA supports a 4-factor model (VCI, PRI, WMI, PSI)	r=.87 (WISC-IV FSIQ with WIAT-II total achievement) r=.31 (FSIQ with BarOn EQ total score)
Nonverbal IQ	Columbia Mental Maturity Scale (CMMS)	Split-half: r=.85 - .91 (Mdn r=.88)	r=.85 across three age groups		r=.69, .62 with Otis-Lennon Mental Ability Test (two elementary groups) r=.67 with SB-Form LM r=.48 with S-B-IV r=.54 with McCarthy Total and r =.74 with McCarthy Perceptual-Performance Scale			r=.31 to .61 with Stanford Achievement Test (in grade 1); r=.43 to .61 (in grade 2) rs=.14 and .15 with McCarthy Verbal and Memory Scales
Nonverbal IQ	Leiter International Performance Scale-Revised (Leiter-R)	IQ: r =.91-.93 (Full); r =.88-.90 (Brief) Composites: r = .75 to .93 (Mdn = .88) Subtests (α) r=.67-.90 (Mdn = .82)	IQ: r = .90 - .96 (Full); r = .88-.96 (Brief). Composites: r = .83 - .92 (Mdn = .85) Subtests: .61 - .90 (Mdn = .80)		r=.76 - .80 with UNIT for language impaired. r=.72 with UNIT for typical r=.85 with Leiter for typical r=.87 with Leiter for children with autism	Manual: Correctly identified children with intellectual disabilities. Published papers: Children with Down, autism, and Fragile X syndrome performed in intellectually disabled range.	CFA evidence presented in manual in support of multi-factor theoretical model	Regression slopes equal between Caucasian and African-American samples for Leiter-R predicting math achievement.

Nonverbal IQ	Raven Progressive Matrices (RPM)	Split-half exceeds r=.90	r=.85 and higher is reported in most well-conducted studies	r=.54 to .86 with SB and Wechsler tests	No significant differences between RPM and Wechsler PIQ in a sample of children with Autism; RPM and FSIQ differences were significant	r=.62 with Boston Naming Test RPM not correlated with Math & Science grades (South African sample)
Achievement	Peabody Individual Achievement Test – Revised (PIAT-R)	Split-half reliability: r=.83 to .98 (subtests); r=.97 (median Total Reading); r=.99 (median Total Test)	r=.78 to .98 (subtests); r=.95 (median Total Reading); r=.96 (median Total Test)	Correlations in the predicted direction in a learning disabled population were found between the PIAT-R, the Woodcock Johnson Psycho-Educational Battery, and the Diagnostic Achievement Battery PIAT-R subtest and test composite scores correlated with like subtests of the Wide Range Achievement Test-3rd edition (WRAT-3) and the Wide Range Assessment of Memory and Learning (WRAML) in a sample of children with psychiatric and neurological disorders.	PIAT scores in adolescents with substance abuse disorder and conduct disorder did not significantly differ, but both groups scored lower than controls on the subtests and the total test composite	Research involving adolescents with mental retardation demonstrated concurrent validity between the PIAT and the Woodcock Johnson Reading Mastery Tests (WRMT).

Achievement	Wechsler Individual Achievement Test-Second Edition (WIAT-II)	*M* Split-half for Composites: r=.98 (Total) r=.98 (Reading) r=.95 (Math) r=.94 (Written) r=.89 (Oral) *M* Split-half for Subtests r=.80 (Listening Comp.) to .97 (Word Reading & Pseudoword Decoding) Mdn = .92	*M* test-rest for Composites (*N*=291): r=.98 (Total) r=.97 (Reading) r=.95 (Math) r=.94 (Written) r=.92 (Oral) *M* test-retest for Subtests r=.85 (Written Expression) to .98 (Word Reading) Mdn = .93	Interscorer agreement (*M*): r=.94 (Reading Comp.) ICC=.85 (Written Expression) ICC=.96 (Oral Expression)	Similar WIAT-II and WRAT-3 subtests correlated with one another (e.g., Reading, r=.73; Math, r=.77; Spelling, r=.78). WIAT-II subtests correlated with like subtests on Stanford Achievement Tests, 9th ed (rs=.68-.86). WIAT-II scores significantly correlated with school grades (rs for composites: .23-.60).	Children with MR and LD, scored significantly lower on all subtests when compared to matched controls. Children with ADHD and speech impairments scored significantly lower on most subtests when compared to matched controls. Children identified as gifted scored significantly higher when compared to matched controls.	WIAT-II Broad Reading and Broad Math composites supported by SEM analysis of WISC-IV scores predicting WIAT-II composites.	Subtest and composite intercorrelations suggest discriminant validity (e.g., Reading/Math rs= .46 (15 y.o.) -.75 (6 y.o.). WIAT-II composites correlated moderately with Wechsler IQs (e.g., Broad Reading/FSIQ rs = .50 (6 y.o.) - .77 (17-19 y.o.)

Achievement	Wide Range Achievement Test 3 (WRAT-3)	α=.85 to .95 (subtests); α=.92 to .95 (combined tests)	r=.91 to .98 (subtests)	Alternate form correlations:
				r=.98 (Reading);
				r=.98 (Spelling);
				r=.98 (Arithmetic)
					Moderate to high Pearson product-moment correlations between scores on the WRAT-3 and the Weschler Individual Achievement Test
					The manual discusses comparative analyses between the WRAT-3 and two other tests of achievement, the California Test of Basic Skills and the Stanford Achievement Test
						The manual discusses discriminate analysis of children in special education and controls using the WRAT-3.		r=.66 (Reading with WISC-III FSIQ)
								r=.66 (Spelling with WISC-III FSIQ)
								r=.73 (Arithmetic with WISC-III FSIQ)
								r=.59 to .82 between WRAT-3 and the Weschler Intelligence Scale for Children-III.

Achievement	Woodcock-Johnson III Tests of Achievement (WJ-III Ach)	Split-half: r=.98 (total achievement); r=.94 (broad reading); r=.95 (broad math); r=.94 (broad written language); r=.96 (academic skills); r=.93 (academic fluency); r=.95 (academic applications).	Acquired Know.: r=.78-.96. (Mdn = .88); r=.69 to .96 (eight speeded subtests)	Alternate form reliability: r=.80-.96 (Mdn = .85 for 5-17 y.o.) Inter-rater reliability: r=.71 to .85 (handwriting); r=.90 to .99 (writing samples); r=.96 to .99 (writing fluency)	r=.79 with K-TEA composite r=.65 with WIAT total score		CFA provided support for theoretical structure.
Attention	Conners' Continuous Performance Test II (CPT II)	Split-half reliability: r=.66 (variability); r=.83 (commission); r=.94 (omissions); r=.95 (hit reaction time)	r=.05 to .92 for Index scores (N=23) r=.89 (ADHD sample) r=.92 (neurological impairment sample)		CPT II scores showed low correlations with other measures of inattention and hyperactivity/ impulsivity outside of the test setting	Most CPT performance has differentiated ADHD populations from normal population Poor discriminant validity between children with reading disorders (RD) and ADHD. Neurological disorder group made significantly more omission errors, had significantly slower reactions times, and was significantly less consistent than ADHD group. For all CPT II index scores, both clinical groups performed more poorly than the control groups. A study involving the CPT and children with Tourette's Syndrome (TS) found TS children to show significantly slower reaction times than controls.	

Attention	Trail Making Test		r = .41 for Trails A r = .65 for Trails B; both from sample of 100 adolescents (*M* age = 15.9 yr) r = .79 (Trails A) & .89 (Trails B) in adult samples		Part B of the TMT is sensitive to the neuropsychological factors associated with academic problems because children with academic difficulties scored poorly as frequently as children with brain damage. Children with learning disabilities (LD) performed more poorly than controls on the TMT Part B, but differences between these two groups on Part A of the TMT were not significant.		Leon-Carrion (1989) found significant differences in times to complete the TMT between Spanish and American children, thus indicating that the TMT might not be free from cultural influence. r=-.31 between TMT Form A time and WISC-R PIQ
Memory and Learning	California Verbal Learning Test Children's Version (CVLT-C)	List A: r=.84 to .91 (mean: r=.88); Semantic categories: r=.72	r=.31 to .90		Sensitive to memory impairments of TBI in pediatric samples Females with ALL recalled fewer list items and displayed less efficient recall strategies compared to controls Children with dyslexia show difficulty learning novel information, compared to controls	CFA supports a 5-factor model	Predictive of the academic outcome for children with TBI.
Memory and Learning	Wide Range Assessment of Memory and Learning (WRAML)	α=.78 to .90 (subtests); α=.93 (Mdn) Verbal Memory; α=.90 (Mdn) Visual Memory; α=.91 (Mdn) Learning α=.96 (General Memory)	r=.61 to .84	r=.72 with McCarthy Scales (memory index) r=.80 with Stanford-Binet (short-term memory index)	Able to discriminate between children with severe THI and children with mild/moderate THI Among children with ADHD, may be useful in discriminating between those with co-morbid LD and those who do not.	No support for 3-factor model; also, no differentiation between memory and learning Support for a 2-factor model in the standardization sample; 1-factor in a clinical sample	

Visual Motor	Beery-Buktenica Developmental Test of Visual-Motor Integration, 5th Edition (VMI)	α=.82 Odd-even split-half correlation: r=.88	r=.89 (Beery VMI) r=.85 (Visual Perception test) r=.86 (Motor Coordination test)	Inter-scorer reliability: r=.92 (Beery VMI); r=.98 (Visual Perception); r=.93 (Motor Coordination)	r=.52 with WRAVMA Drawing test r=.46 with Bender Gestalt test r=.60 with Geometric Design subtest of WPPSI	Significantly lower VMI scores in a sample of children with Tourette's Syndrome. Preschool children with early onset conduct problems were shown to perform more poorly on the VMI than unaffected children their age		r=.62 with FSIQ of WISC-R; r=.48 with VIQ of WISC-R; r=.66 with PIQ of WISC-R r=.52 with FSIQ of WPPSI; r=.33 with VIQ of WPPSI; r=.61 with PIQ of WPPSI r=.63 with CTBS Overall Total score (elementary school children)
Visual Motor	Bender Visual Motor Gestalt Test (Bender) Bender Visual-Motor Gestalt Test Second Edition (Bender-II)	Bender-II (split-half): r= .86-.94 for 4-16 year olds (*Mdn* = .90)	Bender: r=.50 to .90 (review of over a decade of research) r=.65 (sample of 2nd, 4th, and 6th grade children) Bender-II: r=.83 (39, 4-7 year olds); r=.80 (62, 8-17 year olds)	Bender: Inter-rater reliability: r=.85 (time 1); r=.83 (time 2) Inter-rater: r=.92 (sample of 6 year old children) Bender-II: Inter-rater reliability: Copy: *M* = .90; Recall: *M* = .96.	Bender: r=.74 with VMI (LD sample); r=.36 with VMI (non-LD students) Bender-II: r=.65 with VMI from normative sample (4-17 year olds) r=.80 between Bender-II and Koppitz scoring systems r=.20-.53 on measures of academic achievement r=.31-.40 between WISC-III Block Design, Object Assembly, Picture Completion, and Picture Arrangement subtests	Bender: Less developmental sensitivity than VMI in a sample of EBD adolescents Correlated with IQ, but can not discriminate between high average and superior IQ Scores significantly lower than Koppitz normative data in a sample of children with ADHD Bender-II: Individuals with MR, learning disabilities, ADHD, and autism scores significantly below matched control groups. Intellectually gifted individuals score significantly higher than matched controls.	Bender-II: EFA for standardization sample using Global Scoring System yields a single-factor for all age groups. Scores for each Bender-II figure load significantly on the single factor. In sample of typical children (N=82), Bender-II scores loaded on Perceptual-Organizational factor of WISC-III	Not predictive of brain damage in individuals r=.31, r=.38 with WRAT Visual-motor skills of African American children were less developed than both White and Hispanic children using Bender Bender-II: r=.05-.24 between WISC-III verbal subtests

Visual Motor	Halstead-Reitan Grip Strength Test	r=.79 to .94	r=.91 (males); r=.94 (females)			Grip Strength also identified the lateralization of brain lesions in more instances than both the finger tapping test and the tactual performance test. Grip Strength was found to discriminate between brain-damaged and control groups.	
Visual Motor	Rey-Osterrieth Complex Figure Test (ROCF)		r=.76-.89 (adults)	Intra-rater reliability: r=.96 (Copy); r=.99 (Immediate Recall); r=.96 (Delayed Recall) Inter-rater reliability: r=.88 (Copy); r=.97 (Immediate Recall); r=.96 (Delayed Recall) Inter-rater reliability: r=.94	r=.25-.56 with Beery VMI in 6- to 11-year-olds r=.49 (Copy) and .43 (Recall) with WISC-R PIQ r=-.48 (Copy Distortions) with WISC-R Block Design and -.38 with Object Assembly	Discriminates between adolescents with CNS pathology (e.g., seizure disorders; CP; TBI) versus controls. Error performance scores able to discriminate between girls with ADHD and those without Worse performance on the ROCF in patients with right temporal lobe lesions than those with left temporal lobe lesions	r=.49 with Visual Recall subtest of Wechsler Memory Scale; r=.28 when age and gender were controlled for r=.37 (Copy) and .32 (Recall) with WISC-R VIQ
Visual Motor	Halstead-Reitan Finger Tapping Test (HFTT)		r= .80 (dominant hand scores); r=.82 (non-dominant scores) r=.71 (dominant); r=.76 (non-dominant)	Inter-rater reliability: r=.78 (Dominant); r=.76 (nondominant)		\ HFTT was found to differentiate between patients with traumatic brain injury and normal controls Discriminate validity \ questionable when using standardized “cut-off” scores. In a normal sample of 365 people, 80% of individual scores fell within the impaired range	HFTT made minimal contribution to Processing Speed Index (PSI) Scores on the WAIS-III. HFTT accountable for 3% of variance in PSI scores for 68 adults outpatients with various degrees of traumatic brain injury

Visual Motor	Lafayette Grooved Pegboard Task (GPT)		r=.67-.86 for both dominant and non-dominant hands (ages 15 and older) No test-retest data available for children		r=-.35 for GPT/FTT performance and r=-62 for visual acuity in adult male patients.	The GPT was found to discriminate between groups of children with nonverbal learning disabilities, age-matched peers with reading and spelling difficulties, and from a non-clinical sample. For adult populations, slower scores found for individuals with multiple sclerosis when compared to age-matched controls Individuals with autism have been shown to have significantly lower scores on both hands of the GPT when compared to a control group		
Language	Peabody Picture Vocabulary Test, Third Edition (PPVT-III)	α=.95 (Mdn across age groups) Split-half r = .94 (Mdn across age groups)	r=.92 (Mdn across age groups)	Alternate form: r = .94 (Mdn across age groups)	r=.88 with WISC-III VIQ r=.40 to .87 with measures of expressive and receptive language in a sample of children with Autism			r=.44 to .64 with K-ABC cognitive and achievement scales in young African American children
Neuropsych Battery	NEPSY: A Developmental Neuropsychological Assessment	3-4 year-olds: Domain: r=.70-.91 (Mdn=.88) Subtest: r=.50-.91 (Mdn=.83) 5-12 year-olds: Domain: r=.79-.87 (Mdn=.83) Subtest: r=.68-.91 (Mdn=.78)	3-4 year-olds: Domain: r=.68-.90 (Mdn=.77) Subtest: r=.42-.89 (Mdn=.65) 5-12 year-olds: Domain: r=.67-.76 (Mdn=.70) Subtest: r=.52-.81 (Mdn=.60)	Inter-rater: For qualitative ratings, κ coefficients ranged from .34 – 1.0	WISC-III:.20-.62 WPPSI-R: VIQ/NEPSY Language = .60; PIQ/NEPSY Visuosp. = .42 BSID-II: MDI/NEPSY Language = .61; PDI/NEPSY Sen. Motor = .22 CMS: Gen Mem. NEPSY Mem./Lrn (.57)	Children with ADHD, reading disability, language disorders, autism, fetal alcohol syndrome, TBI, and hearing impairment differed from standardization. Children with neurological impairment and scholastic concerns differed from controls.	5 Domain structure not supported in exploratory factor analysis. NEPSY yielded single factor consisting largely of language subtests. Test manual only reports subscale and domain intercorrelations.	English-speaking Zambian schoolchildren earned NEPSY domain scores within 1 SD of normative mean.