



## A Guide to Child Nonverbal IQ Measures

Laura S. DeThorne

Barbara A. Schaefer

*Pennsylvania State University, University Park*

This guide provides a basic overview of 16 child nonverbal IQ measures and uses a set of specified criteria to evaluate them in terms of their psychometric properties. In doing so, the goal is neither to validate nor to criticize current uses of IQ but to (a) familiarize clinicians and investigators with the variety of nonverbal IQ

measures currently available, (b) highlight some of the important distinctions among them, and (c) provide recommendations for the selection and interpretation of nonverbal IQ measures.

**Key Words:** assessment, cognition, language

The use of IQ scores has become relatively commonplace within the clinical and research practices of speech-language pathology. From a clinical standpoint, speech-language pathologists in the public schools are often encouraged to consider IQ when making decisions regarding treatment eligibility (Casby, 1992; Whitmire, 2000). From a research perspective, IQ measures are often administered to define the population of interest, as in the case of specific language impairment (Plante, 1998; Stark & Tallal, 1981). One distinction across IQ measures that appears particularly relevant to the field of speech-language pathology is whether the measure of interest is considered verbal or nonverbal in nature. Nonverbal intelligence tests were designed to measure general cognition without the confound of language ability. The major impetus for the development of nonverbal IQ measures appeared to come from the U.S. military during World War I. As summarized in McCallum, Bracken, and Wasserman (2001), the military used group-administered IQ measures to place incoming recruits and consequently needed a means to assess individuals that either were illiterate or demonstrated limited English proficiency. Once nonverbal assessment tools were developed, their use expanded to include additional populations, such as individuals with hearing loss, neurological damage, psychiatric conditions, learning disabilities, and speech-language impairments. Because the practice of speech-language pathology centers on such special populations, the present article focuses exclusively on measures of *nonverbal IQ*.

The distinction between verbal and nonverbal IQ was not originally based on empirically driven theory; rather, Wechsler's original IQ scales differentiated verbal from performance (nonverbal) primarily for practical reasons (McGrew & Flanagan, 1998, p. 25). Subsequent theorists have applied factor analytic methods to a variety of

cognitive measures to derive empirically based models of intelligence. For example, Carroll (1993) conducted factor analyses of 477 data sets to develop his three-stratum theory of intelligence. Stratum III includes a general factor, often referred to as *g*, which is thought to contribute in varying degrees to all intellectual activities. The second stratum contains 8 broad abilities, such as fluid intelligence or visual perception. These broad abilities are then further divided into 70 narrow cognitive abilities that contribute to Stratum I. McGrew and Flanagan (1998) have integrated the three-stratum theory with the theory of fluid and crystallized intelligences, associated with Horn and Cattell (1966), to create the Cattell–Horn–Carroll theory of cognitive abilities (McCallum, 2003, p. 64). Based on this theory, McGrew and Flanagan (1998) have suggested that verbal IQ represents the construct of crystallized intelligence but that “nonverbal IQ” is not a valid construct in and of itself: “There is no such thing as ‘nonverbal’ ability—only abilities that are expressed nonverbally” (p. 25).

Although questions regarding the nature of intelligence and how it is best measured are beyond the scope of this article, the Cattell–Horn–Carroll theory of cognitive abilities (McGrew & Flanagan, 1998) is offered as a theoretical framework within which to speculate about the nature of individual IQ measures. Previous published work has provided thorough discussion of the psychometric properties and general issues related to standardized assessment (e.g., McCauley, 2001; McCauley & Swisher, 1984a, 1984b). The present review intends to expand on such work by applying the information to specific nonverbal IQ measures and by including a discussion of how individual tests may relate to broad cognitive abilities. The specific aims of the present article are to (a) familiarize clinicians and investigators with the variety of nonverbal IQ measures currently available, (b) highlight some of the

important distinctions among them, and (c) provide recommendations for the selection and interpretation of nonverbal IQ measures.

## Overview

We undertook a review of all mainstream IQ measures currently on the market and selected for this review those that met five specific criteria. First, each measure had to be marketed or commonly used as a measure of general cognitive functioning. Second, each included measure had to provide a standardized score of nonverbal ability. By nonverbal, we mean that the tests rely heavily on visuo-spatial skills and do not require examinees to provide verbal responses. Some IQ measures, such as the Cognitive Assessment System (CAS; Naglieri & Das, 1997) and the Woodcock–Johnson Tests of Cognitive Abilities—III (Woodcock, McGrew, & Mather, 2001), include nonverbal components but do not organize them into a composite or scaled score. Such measures were not included given that interpretation at the subscale level is generally not advised due to psychometric concerns (McDermott & Glutting, 1997).

Third, in addition to providing a nonverbal measure of general cognitive functioning, each included IQ measure was required to be relatively recent, meaning each was developed or revised within the last 15 years. One exception was made in the case of the Columbia Mental Maturity Scale—Third Edition (CMMS–3; Burgemeister, Hollander, & Lorge, 1972). Although this measure was published over 30 years ago, it was included here given its frequent use within the field of speech-language pathology. The fourth criterion for test inclusion was suitability of the measure for preschool or school-age children. Although a number of the included measures can be administered well into adulthood, IQ measures that focused exclusively on adults were omitted. Fifth, each measure had to be developed for the population at large as opposed to specific groups, such as individuals with visual impairment or deafness. Table 1 contains a list of 16 IQ measures that met all five specified criteria. The remainder of the article is devoted to (a) elaborating on the test information presented in Table 1, (b) providing an evaluation of each test's psychometric properties, and (c) offering recommendations for the selection and interpretation of nonverbal IQ measures.

## Test Information

### Age Range

Column 1 lists the ages at which normative data were collected for the test as a whole or for the nonverbal section specifically if it differed from the whole.

### Verbal Subscore

Column 2 in Table 1 denotes whether each measure provides a verbal subscore in addition to the nonverbal components. Although the focus of the present article is on the nonverbal components of each measure, we thought it might be useful for readers to know whether the same test

provided a verbal subscore as well. Many verbal subscores are composed of subtests that are analogous to language measures. For example, the Wechsler Intelligence Scale for Children—Fourth Edition (WISC–IV; Wechsler, 2003) includes a Similarities subtest that requires children to explain how two items are alike, and the Differential Ability Scales (DAS; C. D. Elliott, 1990a) include a Naming Vocabulary subtest that requires children to name objects/pictures. To the extent they reflect language abilities specifically, verbal subscores may help validate and/or identify language concerns in particular children. Verbal subscores can be derived from 5 of the 16 measures listed in Table 1. Separate from the verbal subscores noted in column 2, a number of measures presented in Table 1 offer additional batteries, supplemental subtests, or alternative forms. Such additional components are noted in column 6, with the subtest descriptions.

### Form of Instructions

Column 5 in Table 1 refers to the form of instructions, either verbal or nonverbal, that is used during test administration. It is not uncommon for measures that are promoted as nonverbal to rely to varying degrees on verbal instruction. In fact, the nonverbal components of almost all tests listed in Table 1 include verbal instructions. For example, even though the child is allowed to respond nonverbally by pointing, administration of the Picture Completion subtest from the WISC–IV (Wechsler, 2003) is accompanied by the instruction, "I am going to show you some pictures. In each picture there is a part missing. Look at each picture carefully and tell me what is missing." Given the use of verbal instructions, McCallum et al. (2001) have argued that most "nonverbal tests" are better described as "language-reduced instruments" (p. 8).

Some measures, such as the Kaufman Assessment Battery for Children (K–ABC; Kaufman & Kaufman, 1983) and the Comprehensive Test of Nonverbal Intelligence (CTONI; Hammill, Pearson, & Wiederholt, 1997), provide verbal instructions within the administration procedures, but note that nonverbal instructions could be used. Such measures are identified by V/NV in column 5 of Table 1. In contrast to measures that simply offer nonverbal instructions as an administration option, three measures in Table 1 were specifically normed with nonverbal instructions: the Leiter International Performance Scale—Revised (Leiter–R; Roid & Miller, 1997), the Universal Nonverbal Intelligence Test (UNIT; Bracken & McCallum, 1998), and the Test of Nonverbal Intelligence—Third Edition (TONI–3; Brown, Sherbenou, & Johnsen, 1997). In such cases, instruction and feedback are provided nonverbally via gestures, facial expressions, modeling, and so on. Of course, the use of nonverbal instructions does not guarantee that language ability will not influence test performance. Verbal problem-solving strategies can be used to solve nonverbal problems. In other words, children can talk themselves through problems that do not require a verbal response. For example, on recent administration (by the second author) of the UNIT, a 7-year-old girl labeled the figures *girl*, *man*, *baby*, *boy* on

**TABLE 1 (Page 1 of 3). Summary of nonverbal IQ measures.**

IQ test	Age range (years;months)	Verbal subscore	Time <sup>a</sup>	Manipulatives <sup>b</sup>	Instructions	Description of nonverbal subtests	Cost <sup>c</sup>
Columbia Mental Maturity Scale—Third Edition (CMMS-3; Burgemeister, Hollander, & Lorge, 1972)	3;6–9;11	No	15–20	No	V	Complete a series by selecting the picture or figure that does not belong with the others	\$725
Comprehensive Test of Nonverbal Intelligence (CTONI; Hammill, Pearson, & Wiederholt, 1997)	6;0–90;11	No	40–60	No	V/NV	<p>Pictorial Analogies—identify relationships in a 2 x 2 picture matrix and select a response to complete a similar set of relationship pictures</p> <p>Geometric Analogies—same as Pictorial Analogies using geometric designs</p> <p>Pictorial Categories—select 1 out of 5 picture figure(s) that shares a relationship with 2 stimulus pictures</p> <p>Geometric Categories—same as Pictorial Categories using geometric designs</p> <p>Pictorial Sequences—identify the rule in a progression of picture stimuli and select a picture to complete the sequence</p> <p>Geometric Sequences—same as Pictorial Sequences using geometric designs</p>	\$378
Differential Ability Scales (DAS; C. D. Elliott, 1990a)	2;6–17;11	Yes: Verbal Cluster at age 3;6–17;11	15–30	Yes: wooden blocks, foam squares, plastic cubes, picture cards, and pencil	V/NV	<p>Block Building—replicate constructions made by the examiner using wooden blocks (age 2;6–3;5)</p> <p>Picture Similarities—place individual picture cards next to similar or related items within illustrated arrays (age 2;6–5;11)</p> <p>Pattern Construction—(timed, with untimed alternative) reproduce pictured designs using foam squares or plastic cubes (age 3;6–17;11)</p> <p>Copying—copy line drawings made by the examiner or shown in a picture (age 3;6–5;11)</p> <p>Matrices—select pictured items to complete rule-based patterns (ages 6–17;11)</p> <p>Sequential/Quantitative Reasoning—view pairs of figures/numbers and determine the missing items (age 6–17;11)</p> <p>Recall of Designs—draw individual abstract designs after viewing each briefly (age 6–17;11)</p>	\$799
Kaufman Assessment Battery for Children <sup>d</sup> (K-ABC; Kaufman & Kaufman, 1983)	4;0–12;6	No	35–50	Yes: foam triangles, photo cards, & color form squares	V/NV	<p>Face Recognition—select matching faces after seeing each for 5 s (age 4)</p> <p>Hand Movement—imitate hand movements made by the examiner</p> <p>Triangles—(timed) replicate pictured geometric configurations using foam triangles</p> <p>Matrix Analogies—select pictures or color form squares that complete the illustrated analogies (age 5–12)</p> <p>Spatial Memory—identify the positions of pictured items after seeing them on a target page for 5 s (age 5–12)</p> <p>Photo Series—put photo cards of a story in chronological order (age 6–12)</p> <p><i>Note.</i> Additional subtests combine to form the Sequential and Simultaneous Processing Scales, which together form the Mental Processing Composite; Achievement subtests are also available. A Spanish version of the test is also available.</p>	\$470
Leiter International Performance Scale—Revised (Leiter-R; Roid & Miller, 1997)	2;0–20;11	No	40	Yes: foam shapes,	NV	<p>Figure Ground—find a variety of pictured objects within increasingly complex illustrations</p> <p>Design Analogies—(bonus points on later items) use picture cards to complete visual analogies (age 6–20)</p>	\$939

**TABLE 1 (Page 2 of 3). Summary of nonverbal IQ measures.**

IQ test	Age range (years;months)	Verbal subscore	Time <sup>a</sup>	Manipulatives <sup>b</sup>	Instructions	Description of nonverbal subtests	Cost <sup>c</sup>
Leiter International Performance Scale—Revised (continued)				picture cards		Form Completion—assemble foam shapes to match an illustration or pair pictures of “dissected” objects with pictures of them in whole form Matching—match foam shapes or picture cards to an array of objects pictured on an easel (age 2–10) Sequential Order—select the picture card that best completes the illustrated array Repeated Patterns—determine which foam shape or picture card best completes an illustrated pattern of shapes or objects Classification—associate foam shapes or picture cards with illustrated items according to similarities in color, size, function, or shape (age 2–5) Paper Folding—(bonus points on later items) look at an object pictured on a card and select the illustration of what that object would look like if folded (age 6–20) <i>Note.</i> Also offers an Attention and Memory Battery with 10 different nonverbal subtests. A Spanish version of the test is also available.	
Naglieri Nonverbal Ability Test (NNAT; Naglieri, 2003)	5;0–17;11	No	25–30	No	V	Solve figural matrix items by identifying a pattern and selecting an answer to complete the pattern. <i>Note.</i> Parallel forms are available.	\$233
Pictorial Test of Intelligence—Second Edition (PTI–2; French, 2001)	3;0–8;11	No	15–30	No	V	Verbal Abstractions—Find the pictured item that is being labeled or described Form Discrimination—look at the pictured target item and identify the associated item from a pictured display Quantitative Concepts—find the pictured item that is associated with the verbal description <i>Note.</i> Test was designed to accommodate eye gaze responses.	\$143
Raven's Standard Progressive Matrices—Parallel form (SPM–P; Raven, Raven, & Court, 1998)	6–adult	No	45	Yes: pencils	V/NV	Complete 5 sets of 12 illustrated puzzles by selecting from an array of potential answers (e.g., select the piece that completes the illustrated pattern) <i>Note.</i> Additional forms of the Standard Progressive Matrices comprise the Coloured Matrices and the Advanced Progressive Matrices.	£129
Reynolds Intellectual Assessment Scales (RIAS; Reynolds & Kamphaus, 2003)	3–94	Yes	10–15	No	V	Odd Item Out—(timed) Select picture from array of 5–7 pictures that makes it different than the others What's Missing—(timed) Identify important part of pictured object that is missing	\$319
Stanford–Binet Intelligence Scale—Fifth Edition (SB5; Roid, 2003)	2–89;11	No	30	Yes: plastic form board shapes, counting rods, 1-in. plastic blocks, pencil	V	Object Series/Matrices—solve novel figural problems, sequences of pictured objects, or matrix-type patterns by pointing to pictured objects or picking up target objects, such as form board shape, counting rod, or block (serves as the routing subtest) Procedural Knowledge—identify common signals and actions Picture Absurdities—select missing or absurd details in pictures Quantitative Reasoning—solve premath, arithmetic, algebraic or functional concepts depicted by selecting blocks, counting rods, or choosing from illustrations Form Board—visualize and solve spatial and figural puzzlelike problems using form board shapes Form Patterns—(timed) complete patterns by moving plastic form board shapes into place	\$858

**TABLE 1 (Page 3 of 3). Summary of nonverbal IQ measures.**

IQ test	Age range (years;months)	Verbal subscore	Time <sup>a</sup>	Manipulatives <sup>b</sup>	Instructions	Description of nonverbal subtests	Cost <sup>c</sup>
Stanford-Binet Intelligence Scale—Fifth Edition (continued)						Delayed Response—find an object that has been placed by the examiner under 1 of 3 cups Block Span—tap sequences of blocks in the same order as modeled by the examiner <i>Note.</i> Depending on examinee's skill level (as determined via the Object Series/Matrices routing subtest), not all subtests are administered.	
Test of Nonverbal Intelligence—Third Edition (TONI-3; Brown, Sherbenou, & Johnsen, 1997)	6;0–89;11	No	15–20	No	NV	Abstract Figural Problem-Solving—select from an array of pictured items to complete the pictured rule-based matrix or figure series <i>Note.</i> Two parallel forms are available.	\$299
Universal Nonverbal Intelligence Test (UNIT; Bracken & McCallum, 1998)	5;0–17;11	No	30	Yes: plastic chips, 1-in. cubes, circular response chips, pencil	NV	Symbolic Memory—model a pictured series of symbols using plastic chips Spatial Memory—use circular response chips to replicate a pattern of dots after a brief exposure to a pictured target page Cube Design—(timed) use cubes to construct a three-dimensional design that matches a pictured target Analogic Reasoning—complete a matrix analogy by selecting from four possible pictured responses <i>Note.</i> Two additional subtests are available for the test's extended battery.	\$579
Wechsler Abbreviated Scale of Intelligence (WASI; Psychological Corporation, 1999)	6–89	Yes	15	Yes: 1-in. blocks	V	Block Design—(timed + bonus points) construct a design with blocks to match a given target stimulus Matrix Reasoning—(timed) choose the stimulus to complete the target matrix or pattern sequence	\$235
Wechsler Intelligence Scale for Children—Fourth Edition (WISC-IV; Wechsler, 2003)	6–16;11	Yes	25	Yes: 1-in. blocks	V	Block Design—(timed + bonus pts.) construct a design with blocks to match a given target stimulus Picture Concepts—identify two to three pictures for different picture arrays that best go together Matrix Reasoning—select stimulus to complete target matrix <i>Note.</i> Supplemental nonverbal subtests are available. A Spanish version of the test is due for release in the fall of 2004.	\$799
Wechsler Preschool and Primary Scale of Intelligence—Third Edition (WPPSI-III; Wechsler, 2002)	2;6–7;3	Yes	25	Yes: 1-in. blocks, cardboard puzzle pieces	V	Object Assembly—(timed) construct a figure given cardboard puzzle pieces (age 2;6–3;11) Block Design—(timed) construct a design with blocks to match a model presented in the stimulus book Matrix Reasoning—select from an array of pictured items to complete the pictured rule-based matrix (age 4–7;3) Picture Concepts—select the two pictures that go best together from 2 rows of pictured items (age 4–7;3)	\$799
Wide Range Intelligence Test (WRIT; Glutting, Adams, & Sheslow, 2000)	4–85	Yes	20–30	Yes: diamond chips	V/NV	Matrices—(timed) select picture that completes rule-based design series Diamonds—(timed) construct specific designs using single or multiple diamond-shaped chips	\$250

*Note.* V = verbal; NV = nonverbal.

<sup>a</sup>Time estimate for nonverbal components only. <sup>b</sup>Refers to the objects that the examinee is required to manipulate during test administration. <sup>c</sup>Cost as of January 6, 2004, taken from Web sites or test publisher catalogs in January 2004. <sup>d</sup>Currently undergoing revision.

the Symbolic Memory subtest to help herself remember the order of presentation. Another example comes from the Leiter–R Figure Ground subtest (administered by the first author) as a 6-year-old boy was scanning an illustration of clowns in search of a small plus-like design when he remarked, “I can’t see no band-aid.” Labeling the plus-like design as a *band-aid* likely served as a heuristic of sorts (albeit an unsuccessful one in this case), which could serve to focus his visual scan on areas within the illustration where a bandage might occur (i.e., on one of the clown’s bodies). As these examples illustrate, linguistic ability is likely to influence all nonverbal IQ measures to some extent. However, the influence is minimized when both the instructions and the responses are provided nonverbally.

### *Description of Nonverbal Components*

Column 6 within Table 1 contains a brief description of the nonverbal tasks or subtests included within each IQ measure. For most IQ measures, column 6 does not provide an exhaustive list of all the available subtests. As previously mentioned, many measures in Table 1 offer a verbal scale as well. Additional components, such as associated batteries and subtests that do not qualify as a core component of either the verbal or nonverbal subscore, have been noted in column 6.

Embedded within the subtest descriptions in column 6 is information regarding (a) applicable ages and (b) time constraints. If the age range for a specific subtest does not differ from the age range provided in column 1, then no additional age information is provided for that subtest in column 6. For example, the Leiter–R Visualization and Reasoning Battery (Roid & Miller, 1997) includes a total of 10 nonverbal subtests that combine to form the IQ measure. However, only 7 are administered to any particular child depending on that child’s age. Note that the age ranges provided in column 6 for each subtest refer to the use of that subtest as a core component of the IQ score, not for its use as a supplemental subtest. The age at which individual subtests can be administered as supplemental subtests may vary from the information presented in Table 1.

In addition to information regarding age, it is noted within the subtest description if the examinee’s response is subject to time constraints. For example, the Triangles subtest of the K–ABC (Kaufman & Kaufman, 1983) allots 2 min per item. If the examinee replicates the desired model but exceeds this time allotment, the item is scored as incorrect. In addition, some subtests assign bonus points, depending on the speed with which the task is completed. Bonus points can exist regardless of whether the subtest is subject to time constraints. For example, the Paper Folding subtest of the Leiter–R (Roid & Miller, 1997) does not require examinees to select the correct response within a set time frame to receive credit. However, on later items the examinees can receive additional points for timely performance. When the examinee’s response has to be completed within a set time period, the word *timed* appears in parentheses at the beginning of the subtest description in Table 1. Similarly, any subtest that awards bonus points for

quick responses is marked by *bonus points* at the beginning of the subtest description.

### **Psychometric Properties**

Any overview of nonverbal IQ measures would be incomplete without consideration of the tests’ normative construction and psychometric properties. The American Educational Research Association, American Psychological Association, and National Council on Measurement in Education worked together to revise their standards for educational and psychological testing in 1999. The resulting standards summarize both the general principles and specific standards underlying sound test development and use. Test developers hold the responsibility for sound instrument development, including (a) collection of an appropriate normative sample and (b) documentation of adequate reliability and validity. Based in large part on these professional standards, we have evaluated each nonverbal measure in terms of its normative sample, reliability, and forms of validity evidence. A summary of our evaluation is provided in Table 2 with related information offered in the accompanying text. Our review of psychometrics is admittedly cursory, and readers are referred to additional publications for more in-depth information (e.g., Athanasiou, 2000; McCauley, 2001).

### *Normative Sample*

Developers of norm-referenced tests must identify and select appropriate normative samples. Normative samples consist of the participants to whom individual examinees’ performances will be compared. The normative sample of each measure listed in Table 2 was evaluated in terms of its (a) size, (b) representativeness, and (c) recency. We rated each aspect of the normative sample as positive (+) or negative (–) when information was available. When adequate information was not provided in the test manual to evaluate a particular aspect of the normative sample, 0 appears in the relevant column of Table 2.

*Size.* Reasonably large numbers of children are preferred to minimize error and achieve a representative sample. To take an extreme example, imagine trying to find ten 6-year-olds that are representative of all 6-year-olds currently living within the United States! To evaluate each measure in terms of sample size, we referred to Sattler’s (2001) recommendation of a 100-participant minimum for each child age group included in the normative sample. An age group was defined by a 1-year interval. Of the 16 tests listed in Table 2, 10 successfully met the criteria for subsample size. Of the 5 measures that failed, the Leiter–R (Roid & Miller, 1997) included less than 100 children in groups age 8, 10, and 11–20, with numbers reaching as low as 45 for some of the 12–17-year-old age groups. Within the normative sample for the Naglieri Nonverbal Ability Test (NNAT; Naglieri, 2003), the 16- and 17-year-olds were combined for a subsample size of 100. Similarly, the 16- and 17-year-olds were combined in the normative data from the UNIT (Bracken & McCallum, 1998) for a subsample total of 175, and the

TABLE 2 (Page 1 of 2). Evaluation of each measure based on specific psychometric criteria.

IQ test	Normative sample				Reliability				Validity				Additional reviews
	Size	Rep	Rec	Int	T-R	Inter	SEM	Dev	Test comp	FA	Group comp	Pred ev	
Columbia Mental Maturity Scale—Third Edition (CMMS-3; Burgemeister, Hollander, & Lorge, 1972)	+	-	-	-	-	np	+	np	√	np	np	np	Egeland (1978), Kaufman (1978)
Comprehensive Test of Nonverbal Intelligence (CTONI; Hammill, Pearson, & Wiederholt, 1997)	+	+	+	+	-	+	+	√	√	√	√	np	Athanasiou (2000), Aylward (1998), Bracken & Naglieri (2003), Bradley-Johnson (1997), Drossman & Mailer (2000), McCallum (2003), Van Lingen (1998)
Differential Ability Scales (DAS; C. D. Elliott, 1990a)	+	+	+	-	-	np	-	np	√	√	√	np	Aylward (1992), C. D. Elliott (1990b), S. N. Elliott (1990), Flanagan & Alfonso (1995), Platt, Kamphaus, Keltgen, & Gilliland (1991), Reinehr (1992)
Kaufman Assessment Battery for Children (K-ABC; Kaufman & Kaufman, 1983 <sup>a</sup> )	+	+	-	-	-	np	+	√	√	√	np	√	Anastasi (1985), Bracken (1985), Castellanos, Kline, & Snyder (1996), Coffman (1985), Conoley (1990), Hessler (1985), Hopkins & Hodge (1984), Kamphaus (1990), Kamphaus & Reynolds (1984), Keith (1985), Mehrens (1984), Page (1985), Sternberg (1983, 1984)
Leiter International Performance Scale—Revised (Leiter-R; Roid & Miller, 1997)	-	0	+	0	0	np	-	√	√	√	√	np	Athanasiou (2000), Bracken & Naglieri (2003), Farrell & Phelps (2000), Marco (2001), McCallum (2003), Stinnett (2001)
Naglieri Nonverbal Ability Test (NNAT; Naglieri, 2003)	-	+	+	-	-	np	+	np	√	np	√	np	McCallum (2003)
Pictorial Test of Intelligence—Second Edition (PTI-2; French, 2001)	+	0	+	+	0	+	+	√	√	√	√	np	Athanasiou (2003), Flanagan & Caltabiano (2003)
Raven's Standard Progressive Matrices—Parallel form (SPM-P; Raven, Raven, & Court, 1998)	-	-	-	-	-	np	-	np	√	√	np	√	Bortner (1965), Bracken & Naglieri (2003), McCallum (2003)
Reynolds Intellectual Assessment Scales (RIAS; Reynolds & Kamphaus, 2003)	-	+	+	0	-	+	-	√	√	√	√	np	
Stanford-Binet Intelligence Scale—Fifth Edition (SB5; Roid, 2003)	+	-	+	+	-	np	+	√	√	√	√	np	

TABLE 2 (Page 2 of 2). Evaluation of each measure based on specific psychometric criteria.

IQ test	Normative sample			Reliability			Validity				Additional reviews		
	Size	Rep	Rec	Int	T-R	Inter	SEM	Dev	Test comp	FA		Group comp	Pred ev
Test of Nonverbal Intelligence—Third Edition (TONI-3; Brown, Sherbenou, & Johnsen, 1997)	+	+	+	-	-	+	+	√	√	√	√	np	Athanasiou (2000), Atlas (2001), DeMauro (2001), McCallum (2003)
Universal Nonverbal Intelligence Test (UNIT; Bracken & McCallum, 1998)	-	+	+	-	-	np	+	np	√	√	√	np	Athanasiou (2000), Bandalos (2001), Bracken & Naglieri (2003), Farrell & Phelps (2000), Fives & Flanagan (2002), McCallum (2003), Young & Assing (2000)
Wechsler Abbreviated Scale of Intelligence (WASI; Psychological Corporation, 1999)	+	+	+	+	-	0	+	np	√	√	√	np	Keith (2001), Linksog & Smith (2001)
Wechsler Intelligence Scale for Children—Fourth Edition (WISC-IV; Wechsler, 2003)	+	+	+	+	-	+	+	np	√	√	√	np	
Wechsler Preschool and Primary Scale of Intelligence—Third Edition (WPPSI-III; Wechsler, 2002)	+	+	+	-	-	+	+	np	√	√	√	np	Hamilton & Burns (2003)
Wide Range Intelligence Test (WRIT; Glutting, Adams, & Sheslow, 2000)	-	0	+	-	0	np	-	√	√	√	np	np	Stinnett (2003), Widaman (2003)

Note. Rep = representativeness; Rec = recency; Int = internal; T-R = test-retest; Inter = interrater; Dev = developmental; Test comp = test comparison; FA = factor analysis; Group comp = group comparison; Pred ev = predictive evidence; + = specified criteria met; - = specified criteria not met; np = such evidence not provided within the text manual; √ = present; 0 = inadequate information provided by the test manual.



17- and 18-year-olds were combined within the Wide Range Intelligence Test (WRIT; Glutting, Adams, & Sheslow, 2000) for a subsample size of 125. Within the normative data from the Reynolds Intellectual Assessment Scales (RIAS; Reynolds & Kamphaus, 2003), children age 11 to 19 years have been combined to form subsamples ranging in size from 156 to 167 per 2–3-year grouping. The Raven's Standard Progressive Matrices—Parallel Form (SPM-P; Raven, Raven, & Court, 1998) also failed to meet the sample-size criteria given that its normative sample of American children was derived by combining smaller local norms for which individual sample sizes are not reported.

*Representativeness.* Related but distinct from the issue of sample size is the property of representativeness. The representativeness of a normative sample is critical in evaluating the usefulness of an instrument for any specific population. Since these tests were not designed for any particular group, normative samples should be collected and stratified to approximate the population percentages from the U.S. Bureau of the Census (2001). To receive a positive rating for representativeness in Table 2, a measure had to meet two criteria. First, the measure's normative sample needed to be expressly stratified on the following variables: race/ethnicity, geographic region, parent educational level/socioeconomic status, and sex. Second, the test's normative sample needed to include children with disabilities as appropriate given the test's administration procedures. For example, it made sense to exclude children whose visual or motor impairments would have limited their performance. However, a sample that excludes children with disabilities as a general rule was not considered representative within our review (see McFadden, 1996, for rationale).

All reviewed measures except for the SPM-P (Raven et al., 1998) attempted to derive a nationally representative sample in regard to race/ethnicity, geographic region, parent educational level/socioeconomic status, and sex. The measures in Table 2 were more variable in regard to whether children with disabilities were included. Two measures received a negative rating for representativeness because they excluded children with disabilities altogether: the CMMS-3 (Burgemeister et al., 1972) and the Stanford-Binet Intelligence Scale (5th ed.; SB5; Roid, 2003). Similarly, the Leiter-R (Roid & Miller, 1997) and the WRIT (Glutting et al., 2000) received an uncertain rating, identified as 0 in Table 2, because they did not specify within the manual whether children with disabilities were included in the normative sample. Although the second edition of the Pictorial Test of Intelligence (PTI-2; French, 2001) included children with disabilities, it received an uncertain rating as well due to the relatively sparse information it provided on its normative sample and concerns regarding the extent to which the sample characteristics differed from U.S. Bureau of the Census data (see reviews by Athanasiou, 2003; Flanagan & Caltabiano, 2003).

*Recency.* In addition to being large and representative, normative samples should also be relatively recent. The Flynn (1999) effect has revealed trends in IQ scores across time such that IQs have increased by 3 points per decade.

Consequently, reliance on instruments with outdated norms may overestimate youths' IQ (see Neisser et al., 1996, for additional discussion). In keeping with Bracken (2000), we considered the normative data to be acceptably recent if it had been collected within the last 15 years. Given this criterion, normative samples from the CMMS-3 (Burgemeister et al., 1972) and the K-ABC (Kaufman & Kaufman, 1983) were considered out-of-date. Note, however, that a revised version of the K-ABC is currently under way.

### **Reliability**

In addition to identifying and selecting appropriate normative samples, test developers must demonstrate that their tests have adequate reliability and validity to justify their proposed use. In classical test theory, reliability, as measured via correlation coefficients, represents the degree to which test scores are consistent across items (i.e., internal consistency), across time (i.e., test-retest), and across examiners (i.e., interrater). Accordingly, we evaluated each measure in terms of internal reliability, test-retest reliability, and interrater reliability. A fourth factor, the availability of standard error of measurement for each summary score, was also considered. Given the focus of the present article on nonverbal summary scores rather than subtests, reliability evidence on subtest scores was not evaluated. Identical to our rating system for the normative sample, each form of reliability evidence was evaluated as positive or negative in Table 2. When the test manual did not provide adequate information for us to determine whether a specific criterion was met, 0 appears in the relevant column of Table 2. The reader is referred to McCauley (2001) for additional information regarding the constructs of reliability and validity.

*Internal reliability.* Internal reliability was rated positively within Table 2 if uncorrected internal reliability coefficients reached .90 or above for the nonverbal summary score within each child age group. This criterion for reliability coefficients is consistent with general recommendations for test construction (see Bracken, 1988; McCauley & Swisher, 1984a; Salvia & Ysseldyke, 1998). Only 5 of the 16 measures met the established criterion for internal reliability. Of those that did not, the DAS (C. D. Elliott, 1990a); Leiter-R (Roid & Miller, 1997); RIAS (Reynolds & Kamphaus, 2003); and WRIT (Glutting et al., 2000) received uncertain marks because they failed to provide separate reliability coefficients for each age group. Although these measures reported reliability coefficients on combined age groups, such coefficients can mask substantial variability. Of the measures that provided coefficients at each age group, 7 failed to consistently meet .90, although none fell too far from the mark.

*Test-retest reliability.* The second criterion for test reliability, in keeping with Bracken (1987), was a test-retest coefficient of .90 or greater for the nonverbal summary score at each child age group. In addition, the retesting had to occur at least 2 weeks, on average, from the initial test administration. Not a single nonverbal IQ measure from Table 2 met this criterion. Three measures,

the WRIT (Glutting et al., 2000), Leiter-R (Roid & Miller, 1997), and PTI-2 (French, 2001), received uncertain marks for test-retest reliability because they reported coefficients for collapsed age groups, thereby making reliability estimates for individual age groups impossible to determine. Most measures from Table 2 reported coefficients from collapsed or select age groups but received a failed mark (as opposed to uncertain) because the reported coefficients did not reach .90. Only three measures presented individual test-retest coefficients for each age group: the CTONI (Hammill et al., 1997), which reported coefficients of .79 to .94; the UNIT (Bracken & McCallum, 1998), which reported coefficients of .83 to .90; and the WISC-IV (Wechsler, 2002), which reported coefficients of .85 to .93 across age groups.

**Interrater reliability.** The third criterion for evaluation of test reliability was an interrater reliability of .90 or greater for the nonverbal summary score. We want to be clear that this form of reliability refers only to the consistency of scoring across examiners and does not evaluate the consistency of administration across examiners. The latter is a source of error that should be assessed but was not reported by any of the reviewed test manuals. Interrater reliability of scoring procedures for the nonverbal composite was reported for 7 of the 16 measures: the CTONI (Hammill et al., 1997); PTI-2 (French, 2001); RIAS (Reynolds & Kamphaus, 2003); TONI-3 (Brown et al., 1997); Wechsler Abbreviated Scale of Intelligence (WASI; Psychological Corporation, 1999); WISC-IV (Wechsler, 2003); and Wechsler Preschool and Primary Intelligence Scale—Third Edition (WPPSI-III; Wechsler, 2002). All of these received a positive rating for meeting the proposed criterion except the WASI, which received an uncertain mark due to its relatively vague statement that interscorer agreement for the nonverbal subtests “tends to be in the high .90s” (Psychological Corporation, 1999, p. 129). The DAS (C. D. Elliott, 1990a) presented interrater reliability information, but only for particular subtests.

**Standard error of measurement.** The fourth criterion for evaluating test reliability refers to the test manual’s presentation of standard error of measurement (*SEM*). Whereas the reliability coefficients can inform us about group-level consistency, such correlations are not directly applicable to individual scores. Instead, the *SEM* is needed. *SEM* is a derivative of the reliability coefficient that takes into account the scale of the measurement used, the standard deviation, and the internal consistency of the score. As such, the *SEM* provides some sense of how much error is likely encompassed within the obtained score (see McCauley & Swisher, 1984b, for additional detail). For example, on the UNIT (Bracken & McCallum, 1998), the Standard Battery Full-Scale score *SEM* is approximately 4 points. Consequently if a child received a score of 87, one could feel reasonably confident that the child’s true score would be captured within the range of 87  $\pm$  4 points (i.e., 83–91 is the 68% confidence interval). Given that the derivation of confidence intervals relies on *SEM*, the final criterion for test reliability was that the *SEM* be presented for the nonverbal summary score at each child age group. As glimpsed from Table 2, 11 of the 16 tests presented

*SEM* by age for the nonverbal summary score. Of the 5 measures that failed this criterion, the Leiter-R (Roid & Miller, 1997); RIAS (Reynolds & Kamphaus, 2003); and WRIT (Glutting et al., 2000) provided *SEMs* for combined age groups only. The DAS (C. D. Elliott, 1990a) presented *SEMs* for individual age groups, but only for children age 3;6–7;11. The SPM-P (Raven et al., 1998) did not provide *SEM* values at all.

## Validity

Although evidence for test validity can take many forms, validity as an overall concept represents the utility of a test score in measuring the construct it was designed to measure (McCauley, 2001; Messick, 1989). For IQ tests, developers should clearly articulate the theoretical grounding of their instrument and document evidence for validity from a variety of sources. Table 2 illustrates the validity evidence presented within each test manual. The presence of each form of validity evidence within the test manual is indicated by a  $\checkmark$  in Table 2. The letters *np* appear in the relevant columns of Table 2 to indicate which forms of validity evidence were not provided for any particular measure.

Table 2 reviews each measure according to whether it provided the following most common forms of construct validity evidence: (a) developmental trends, (b) correlation with similar tests, (c) factor analyses, (d) group comparisons, and (e) predictive ability. Although demonstration of internal consistency is another commonly used form of validity evidence, this evidence was already covered under the discussion and evaluation of each test’s internal reliability.

**Developmental trends.** Demonstrating that a test’s raw scores increase with age provides necessary, but by no means sufficient, evidence of a test’s validity. Although cognitive abilities increase with age, so do many other qualities—attention, language use, and height, to name just a few. Perhaps in part because it represents a relatively weak form of validity evidence, only 8 of the 16 measures explicitly mention evidence of developmental change in their discussions of validity (see Table 2).

**Correlation with similar tests.** In contrast to the relatively sparse evidence of developmental trends, the most common form of validity evidence appeared to be information regarding correlations with other measures. All 16 tests from Table 2 presented positive correlations with other measures of general cognition. In addition, there was a general tendency to report positive correlations with achievement tests as a form of validity evidence. Of course, the strength of such intertest correlations in demonstrating test validity is dependent on the validity of the test being used for comparison. For example, two tests developed to assess general cognition might be highly correlated if both reflect the same confounding variable (e.g., language skill or test-taking abilities). Given that the same tests tend to benchmark with each other, intertest correlation evidence is limited by an air of circularity.

**Factor analyses.** Another commonly used form of validity evidence is the presentation of factor analyses, either confirmatory, exploratory, or both. Without getting

into much detail, factor analysis is a statistical procedure based on item or subscale intercorrelations that is used to determine whether the test components load on the abilities of interest (see McCauley, 2001, for additional information). Such evidence was presented by 14 of the 16 measures. Of course, the limitation of such evidence is that just because a test taps a relatively unitary construct or constructs does not necessarily mean it's the *right* construct.

**Group comparisons.** Given the limitations of the forms of evidence previously mentioned, additional validity support is contributed through comparison studies of specific subgroups—specifically, subgroups with clear associations with IQ (e.g., children identified as gifted or with diagnosed cognitive deficits). In other words, the measure of interest is administered to subgroups of children with previously established diagnoses to determine how the descriptive data from such subgroups compare with the test's normative data. Children with previously identified cognitive deficits are expected to score below the normative mean, whereas children previously identified as gifted would be expected to score above the normative mean. Of course, the extent to which subsample means are expected to deviate from the normative mean is less clear, and even clear mean differences can mask a substantial amount of individual variability. Nonetheless, 12 of 16 measures provided validity evidence based on such subgroup comparisons (see Table 2). For example, the manual from the SB5 (Roid, 2003) reported that children previously diagnosed with mental retardation scored 2–3 *SD* below the normative mean on the SB5, whereas children previously classified as gifted scored 1–2 *SD* above the normative mean.

**Predictive ability.** Predictive evidence, perhaps one of the strongest forms of validity evidence, is reported least often. Predictive validity evidence encompasses the effectiveness of a test in predicting the performance of individuals in related activities. For example, a measure might report the extent to which its scores predict children's educational placement or later academic achievement. Obviously, collecting such evidence requires laborious longitudinal investigation with potentially unclear results. As such, predictive evidence is provided by only 2 of the 16 measures, the K-ABC (Kaufman & Kaufman, 1983) and the SPM-P (Raven et al., 1998). The K-ABC presents results from six studies that used the K-ABC to predict scores on achievement tests administered 6–12 months later. Similarly the SPM-P reports correlations “ranging up to about .70” between scores from the SPM-P and scores from achievement tests administered “some time” later (Raven et al., 1998, p. 29). Additional measures, such as the SB5 (Roid, 2003) and UNIT (Bracken & McCallum, 1998), presented correlation data with concurrently administered achievement tests as evidence of predictive validity. However, the *concurrent* administration of the tests disqualified it as predictive evidence under our criteria.

### **Additional Resources**

Given that for many measures our review represents but one of many, we have provided the final column in Table 2

as a gateway to the research literature on these tests. Although our list is not exhaustive, we have included all reviews that we were able to access via PsycINFO as well as those that were available through the Buros's *Mental Measurements Yearbook* series and the *Tests in Print* series. In addition, we have included reviews from Athanasiou (2000), Bracken and Naglieri (2003), and McCallum (2003), as well as those cited within McGrew and Flanagan (1998). Note that a limited number of published reviews are available for the recently developed or revised measures. Readers are referred to the Buros Institute *Mental Measurements Yearbook* Test Reviews On-Line at <http://buros.unl.edu/buros/jsp/search.jsp> and the Educational Testing Service's Test Link database available at <http://www.ets.org/testcoll/index.html> for current information.

## **Recommendations**

### **Selection of an Appropriate Measure**

Despite the responsibility incumbent on test developers, “the ultimate responsibility for appropriate test use and interpretation lies predominately with the test user” (American Educational Research Association et al., 1999, p. 111). We recognize that the administration of IQ measures is primarily the responsibility of psychologists rather than speech-language pathologists. However, we provide suggestions for test selection for two primary reasons. First, both clinicians and investigators sometimes select and administer IQ measures within the scope of their assessment protocols. Second, it is our hope that speech-language professionals will use the information presented within this article to engage in meaningful dialogue with collaborating psychologists regarding how IQ scores should (and should not) be interpreted and which measures are most appropriate for children with communication difficulties. To this end, we offer three considerations to guide the selection of a nonverbal IQ measure. First, the measure should be psychometrically sound, as previously discussed. Although none of the measures presented in Table 2 are psychometrically ideal, four emerge as particularly strong: the TONI-3 (Brown et al., 1997), UNIT (Bracken & McCallum, 1998), WASI (Psychological Corporation, 1999), and WISC-IV (Wechsler, 2003). Although all four measures received negative marks in terms of test-retest reliability, all fell close to the .90 criterion (although note that the WASI and TONI-3 represented information on collapsed age groups only). Similarly, the TONI-3 and UNIT received negative marks under internal reliability, but their values fell just below the .90 criterion, at .89 for certain age groups. In addition, the UNIT received a negative mark under sample size, but that was due specifically to the 16–17-year-old age grouping.

The second consideration for test selection is potential special needs of the population or individual being evaluated. Given our particular interest in children with language difficulties, it is important to consider the linguistic demands of individual IQ tests. Standard 7.7 (American Educational Research Association et al., 1999) specifically indicates, “in testing applications where the level of

linguistic or reading ability is not part of the construct of interest, the linguistic or reading demands of the test should be kept to the minimum necessary for the valid assessment of the intended construct” (p. 82). Clearly, the nonverbal instruments listed in Table 1 all attempt to address this standard but with differing levels of success. If concerns exist regarding language abilities—particularly receptive language skills—test users should select a test that involves nonverbal instructions only, such as the UNIT (Bracken & McCallum, 1998) or the TONI-3 (Brown et al., 1997). Although the Leiter-R (Roid & Miller, 1997) also relies on nonverbal instruction, its psychometric properties do not appear as favorable as the UNIT and the TONI-3. In terms of special needs, note also that the color scheme of the UNIT (green and black) was specifically designed to avoid most problems associated with color blindness. For children with limited motor control, examiners might want to give special consideration to the TONI-3, because it is untimed and does not depend heavily on manipulatives.

The third factor to consider in selecting a nonverbal IQ measure is the purpose of assessment. High-stakes assessments that will result in diagnoses, treatment eligibility, and/or educational placement should use multidimensional batteries, such as the UNIT (Bracken & McCallum, 1998) or the WISC-IV (Wechsler, 2003). In contrast, if one is conducting a low-stakes assessment, such as a screening for research purposes, a one-dimensional battery may suffice, such as the TONI-3 (Brown et al., 1997) or the WASI (Psychological Corporation, 1999). McCallum et al. (2001) describe one-dimensional measures as assessing “a narrow aspect of intelligence through the use of progressive matrices,” whereas multidimensional measures “assess multiple facets of examinees’ intelligence,” such as memory, reasoning, and attention (pp. 9–10). In sum, given all three aforementioned considerations, if one is selecting a nonverbal IQ measure for a child with language impairment, we recommend the UNIT for cases of high-stakes assessment and the TONI-3 for cases of low-stakes assessment.

### *Interpretation of Nonverbal IQ Measures*

Once an IQ measure has been selected and administered, one’s focus turns naturally to interpretation. IQ scores as a whole are interpreted as estimates of an individual’s current level of general cognitive functioning. IQ scores tend to correlate positively with academic achievement ( $r = .50-.75$ ; Anastasi, 1988; Neisser et al., 1996) and have proved to be one of the best predictors of academic and job-related outcomes (Neisser et al., 1996; Sattler, 2001). However, given our focus on nonverbal IQ measures, it seems important to note that nonverbal IQ has proven to be less effective in predicting academic outcomes than verbal IQ (Sattler, 2001). This difference in predictive power is potentially due to two explanations, neither of which precludes the other: (a) Verbal subtests have a higher percentage of variance attributable to general intelligence, or  $g$ , than do nonverbal subtests, and (b) language plays an important role in the standard academic

learning environment. In regard to the latter point, it is difficult for children with poor language skills in general to succeed in school regardless of their level of general cognitive functioning (Records, Tomblin, & Freese, 1992; Tomblin, Zhang, Buckwalter, & Catts, 2000).

Regardless of predictive power, the most accurate interpretation of an individual’s performance on any given measure comes from an understanding of the specific ability or abilities reflected by that measure. As previously mentioned, factor analytic techniques have given shape to a number of broad cognitive abilities (see Carroll, 1993; Horn & Cattell, 1966). Each broad cognitive ability can be further dissected into numerous narrow cognitive abilities, a discussion of which is beyond the scope of this article (see Carroll, 1993; McCallum, 2003; McGrew & Flanagan, 1998). McGrew and Flanagan have proposed the Cattell-Horn-Carroll theory, which includes 10 broad cognitive abilities: crystallized intelligence, fluid intelligence, quantitative knowledge, reading and writing ability, short-term memory, visual processing, auditory processing, long-term storage and retrieval, processing speed, and decision/reaction time.

McGrew and Flanagan (1998) used the broad intelligences from the Cattell-Horn-Carroll theory as a framework for reviewing a number of IQ measures. They concluded that nonverbal IQ measures as a whole draw heavily on visual processing abilities and fluid intelligence (see also analyses by Johnston, 1982, and Kamhi, Minor, & Mauer, 1990). Fluid intelligence, typified by inductive and deductive reasoning, refers to “mental operations that an individual may use when faced with a relatively novel task that cannot be performed automatically” (McGrew & Flanagan, 1998, p. 14). Select examples of such tasks include the Picture Similarities subtest from the DAS (C. D. Elliott, 1990a); the Photo Series subtest from the K-ABC (Kaufman & Kaufman, 1983); the Analogic Reasoning subtest from the UNIT (Bracken & McCallum, 1998); and the matrices tasks that essentially define the NNAT (Naglieri, 2003); CMMS-3 (Burgemeister et al., 1972); CTONI (Hammill et al., 1997); SPM-P (Raven et al., 1998); and TONI-3 (Brown et al., 1997).

Visual processing is defined by McGrew and Flanagan (1998) as “the ability to generate, perceive, analyze, synthesize, manipulate, transform, and think with visual patterns and stimuli” (p. 23). Example measures of visual processing taken from McGrew and Flanagan include all the subsets from the K-ABC (Kaufman & Kaufman, 1983) as well as the Block Building, Pattern Construction, and Copying subtests from the DAS (C. D. Elliott, 1990a). One would expect similar subtests across nonverbal IQ measures to be associated with visual processing ability as well, such as the Design Analogies subtest from the Leiter-R (Roid & Miller, 1997) and the Block Design subtest from the WASI (Psychological Corporation, 1999). Note that any single subtest can and likely does reflect more than one broad cognitive ability. For example, in addition to tapping visual processing skills, it seems likely that the Cube Design subtest from the UNIT (Bracken & McCallum, 1998) would be influenced by decision/reaction time. Similarly, the Symbolic Memory and Spatial

Memory subtests from the UNIT are likely to be influenced by both visual processing and short-term memory.

Table 3 contains proposed links between broad cognitive abilities from the Cattell–Horn–Carroll theory and subtests from the four measures that emerged from our review as psychometrically strong: the TONI–3 (Brown et al., 1997); UNIT (Bracken & McCallum, 1998); WASI (Psychological Corporation, 1999); and WISC–IV (Wechsler, 2003). Such proposed links are based largely on the work of McCallum (2003) and McGrew and Flanagan (1998), with some additional speculation on our part. Note that although individual subtests may tap into different nonverbal abilities, subtests’ often inferior psychometric properties make interpretation in isolation or in comparison with other subtests largely speculative (see McDermott & Glutting, 1997). Consequently, reliability coefficients for the individual subtests are provided in Table 3 for reference purposes.

The published test reviews cited in Table 2 revealed little additional information on the broad cognitive requirements of individual IQ measures, although we summarize here what we found for our two recommended measures: the TONI–3 and the UNIT. Both have been criticized for providing a rather limited assessment of more specific cognitive abilities. For example, the TONI–3 has been criticized for its one-dimensional nature (Athanasίου, 2000; see also Kamhi et al., 1990). Along similar lines, Atlas (2001, p. 1260) noted that the TONI–3’s correlation with the widely accepted and multidimensional WISC–III was “at best moderate” ( $r = .53-.63$ ) and based on a small sample (i.e., 34 students). In contrast, however, DeMauro (2001) attributed this attenuated correlation to the higher

verbal demands of the Weschler scales and, as such, presented it as rather positive evidence of the TONI–3’s validity. In regard to the UNIT (Bracken & McCallum, 1998), Young and Assing (2000) credited its manual with a thorough description of the test’s theoretical foundation, but Fives and Flanagan (2002) criticized the measure for not providing a clear and comprehensive association with more specific cognitive abilities. As an example, Fives and Flanagan (2002) suggested that an examinee’s low scores on the Symbolic Memory and Spatial Memory subtests could be attributed to “limitations in memory, reasoning, or both of these constructs” (p. 430). In closing, regardless of whether one is interpreting individual subtest or composite scores, it is important to bear in mind that valid assessment of cognition, as with any developmental domain, must combine information from multiple sources, including observation and interview.

## Summary

In sum, the present article presented information regarding 16 commonly used child nonverbal IQ measures, highlighted important distinctions among them, and provided recommendations for test selection and interpretation. When selecting a nonverbal IQ measure for children with language difficulties, we currently recommend the UNIT (Bracken & McCallum, 1998) for cases of high-stakes assessment and the TONI–3 (Brown et al., 1997) for cases of low-stakes assessment. Regardless of which specific test is administered, professionals are encouraged to be cognizant of each test’s strengths and limitations, including its psychometric properties and its potential

**TABLE 3. Broad cognitive abilities and reliability coefficients associated with individual subtests from the recommended IQ measures.**

IQ test	Nonverbal subtest	Cognitive ability <sup>a</sup>	$r^b$
Test of Nonverbal Intelligence—Third Edition (TONI–3; Brown, Sherbenou, & Johnsen, 1997)	Abstract Figural Problem-Solving	Fluid intelligence	.89–94 <sup>c</sup>
Universal Nonverbal Intelligence Test (UNIT; Bracken & McCallum, 1998)	Symbolic Memory	Visual processing Short-term memory	.80–87
	Spatial Memory	Visual processing Short-term memory	.74–.84
	Cube Design	Visual processing Fluid intelligence Decision/reaction time	.81–.95
	Analogic Reasoning	Fluid intelligence	.70–.83
Wechsler Abbreviated Scale of Intelligence (WASI; Psychological Corporation, 1999)	Block Design	Visual processing Fluid intelligence Decision/reaction time	.84–.93
	Matrix Reasoning	Fluid intelligence Decision/reaction time	.86–.96
Wechsler Intelligence Scale for Children—Fourth Edition (WISC–IV; Wechsler, 2003)	Block Design	Visual processing Fluid intelligence Decision/reaction time	.83–.88
	Picture Concepts	Fluid intelligence	.76–.85
	Matrix Reasoning	Visual processing	.86–.92

<sup>a</sup>Broad cognitive abilities from the Cattell–Horn–Carroll theory, as proposed by McGrew and Flanagan (1998). <sup>b</sup>Range of subtest reliability coefficients across ages for that particular subtest. <sup>c</sup>Subtest reliability for the TONI–3 is the same as for the entire test.

relation to different cognitive abilities. We hope that this guide of nonverbal intelligence measures will help clinicians and investigators to better interpret available IQ scores and, when applicable, select the most appropriate instrument for their specific needs.

## Acknowledgments

We wish to thank John McCarthy, Joanna Burton, Patricia Ruby, and Colleen Curran for their assistance with this project.

## References

- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education.** (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Anastasi, A.** (1985). Review of the Kaufman Assessment Battery for Children. In J. V. Mitchell (Ed.), *Mental measurements yearbook* (9th ed., pp. 769–771). Lincoln, NE: Buros Institute of Mental Measurements.
- Anastasi, A.** (1988). *Psychological testing* (6th ed.). New York: Macmillan.
- Athanasίου, M. S.** (2000). Current nonverbal assessment instruments: A comparison of psychometric integrity and test fairness. *Journal of Psychoeducational Assessment, 18*, 211–229.
- Athanasίου, M.** (2003). Review of the Pictorial Test of Intelligence, Second edition. In B. S. Plake, J. C. Impara, & R. A. Spies (Eds.), *Mental measurements yearbook* (15th ed., pp. 676–678). Lincoln, NE: Buros Institute of Mental Measurements.
- Atlas, J. A.** (2001). Review of the Test of Nonverbal Intelligence, Third edition. In B. S. Plake, & J. C. Impara (Eds.), *Mental measurements yearbook* (14th ed., pp. 1259–1260). Lincoln, NE: Buros Institute of Mental Measurements.
- Aylward, G. P.** (1992). Review of the Differential Ability Scales. In J. J. Kramer & J. C. Conoley (Eds.), *Mental measurements yearbook* (11th ed., pp. 281–282). Lincoln, NE: Buros Institute of Mental Measurements.
- Aylward, G. P.** (1998). Review of the Comprehensive Test of Nonverbal Intelligence. In J. C. Impara & B. S. Plake (Eds.), *The mental measurements yearbook* (13th ed., pp. 310–312). Lincoln, NE: Buros Institute of Mental Measurements.
- Bandalos, D. L.** (2001). Review of the Universal Nonverbal Intelligence Test. In B. S. Plake & J. C. Impara (Eds.), *Mental measurements yearbook* (14th ed., pp. 1296–1298). Lincoln, NE: Buros Institute of Mental Measurements.
- Bortner, M.** (1965). Review of the Raven Standard Progressive Matrices. In O. K. Buros (Ed.), *Mental measurements yearbook* (6th ed., pp. 489–491). Highland Park, NJ: Gryphon Press.
- Bracken, B. A.** (1985). A critical review of the Kaufman Battery for Children (K–ABC). *School Psychology Review, 14*, 21–36.
- Bracken, B. A.** (1987). Limitations of preschool instruments and standards for minimal levels of technical adequacy. *Journal of Psychoeducational Assessment, 4*, 313–326.
- Bracken, B. A.** (1988). Ten psychometric reasons why similar tests produce dissimilar results. *Journal of School Psychology, 26*, 155–166.
- Bracken, B. A.** (Ed.). (2000). *The psychoeducational assessment of preschool children* (3rd ed.). Boston: Allyn & Bacon.
- Bracken, B. A., & McCallum, R. S.** (1998). *Universal Nonverbal Intelligence Test*. Itasca, IL: Riverside Publishing.
- Bracken, B. A., & Naglieri, J. A.** (2003). Assessing diverse populations with nonverbal tests of general intelligence. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children: Intelligence, aptitude, and achievement* (2nd ed., pp. 243–274). New York: Guilford Press.
- Bradley-Johnson, S.** (1997). Review of the Comprehensive Test of Nonverbal Intelligence. *Psychology in the Schools, 34*, 289–292.
- Brown, L., Sherbenou, R. J., & Johnsen, S. K.** (1997). *Test of Nonverbal Intelligence—Third Edition*. Austin, TX: Pro-Ed.
- Burgemeister, B. B., Hollander, L. H., & Lorge, I.** (1972). *Columbia Mental Maturity Scale—Third Edition*. San Antonio, TX: Psychological Corporation.
- Carroll, J. B.** (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, England: Cambridge University Press.
- Casby, M. W.** (1992). The cognitive hypothesis and its influence on speech–language services in schools. *Language, Speech, and Hearing Services in Schools, 23*, 198–202.
- Castellanos, M., Kline, R. B., & Snyder, J.** (1996). Lessons from the Kaufman Assessment Battery for Children (K–ABC): Toward a new cognitive assessment model. *Psychological Assessment, 8*, 7–17.
- Coffman, W. E.** (1985). Review of the Kaufman Assessment Battery for Children. In J. V. Mitchell (Ed.), *Mental measurements yearbook* (9th ed., pp. 771–773). Lincoln, NE: Buros Institute of Mental Measurements.
- Conoley, J. C.** (1990). Review of the K–ABC: Reflecting the unobservable. *Journal of Psychoeducational Assessment, 8*, 369–375.
- DeMauro, G. E.** (2001). Review of the Test of Nonverbal Intelligence, Third edition. In B. S. Plake & J. C. Impara (Eds.), *Mental measurements yearbook* (14th ed., pp. 1260–1262). Lincoln, NE: Buros Institute of Mental Measurements.
- Drossman, E. R., & Maller, S. J.** (2000). Review of the Comprehensive Test of Nonverbal Intelligence. *Journal of Psychoeducational Assessment, 18*, 293–301.
- Egeland, B. R.** (1978). Review of the Columbia Mental Maturity Scale. In O. K. Buros (Ed.), *Mental measurements yearbook* (8th ed., pp. 298–299). Highland Park, NJ: Gryphon Press.
- Elliott, C. D.** (1990a). *Differential Ability Scales*. San Antonio, TX: Psychological Corporation.
- Elliott, C. D.** (1990b). The nature and structure of children’s abilities: Evidence from the Differential Ability Scales. *Journal of Psychoeducational Assessment, 8*, 376–390.
- Elliott, S. N.** (1990). The nature and structure of the DAS: Questioning the test’s organizing model and use. *Journal of Psychoeducational Assessment, 8*, 406–411.
- Farrell, M., & Phelps, L.** (2000). A comparison of the Leiter–R and the Universal Nonverbal Intelligence Test (UNIT) with children classified as language impaired. *Journal of Psychoeducational Assessment, 18*, 268–274.
- Fives, C. J., & Flanagan, R.** (2002). A review of the Universal Nonverbal Intelligence Test (UNIT): An advance for evaluating youngsters with diverse needs. *School Psychology International, 23*, 425–449.
- Flanagan, D., & Alfonso, V.** (1995). A critical review of the technical characteristics of new and recently revised intelligence tests for preschool children. *Journal of Psychoeducational Assessment, 13*, 66–90.
- Flanagan, D. P., & Caltabiano, L. F.** (2003). Review of the Pictorial Test of Intelligence, Second edition. In B. S. Plake, J. C. Impara, & R. A. Spies (Eds.), *Mental measurements yearbook* (15th ed., pp. 678–681). Lincoln, NE: Buros Institute of Mental Measurements.

- Flynn, J. R.** (1999). Searching for justice: The discovery of IQ gains over time. *American Psychologist*, *54*, 5–20.
- French, J. L.** (2001). *Pictorial Test of Intelligence—Second Edition*. Austin, TX: Pro-Ed.
- Glutting, J., Adams, W., & Sheslow, D.** (2000). *Wide Range Intelligence Test*. Wilmington, DE: Wide Range.
- Hamilton, W., & Burns, T. G.** (2003). WPPSI-III: Wechsler Preschool and Primary Scale for Intelligence (3rd ed.). *Applied Neuropsychology*, *10*, 188–190.
- Hammill, D. D., Pearson, N. A., & Wiederholt, J. E.** (1997). *Comprehensive Test of Nonverbal Intelligence*. Austin, TX: Pro-Ed.
- Hessler, G.** (1985). Review of the Kaufman Assessment Battery for Children: Implications for assessment of the gifted. *Journal for the Education of the Gifted*, *8*, 133–147.
- Hopkins, K. D., & Hodge, S. E.** (1984). Review of the Kaufman Assessment Battery (K-ABC) for Children. *Journal of Counseling and Development*, *63*, 105–107.
- Horn, J. L., & Cattell, R. B.** (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, *57*, 253–270.
- Johnston, J. R.** (1982). Interpreting the Leiter IQ: Performance profiles of young normal and language-disordered children. *Journal of Speech and Hearing Research*, *25*, 291–296.
- Kamhi, A. G., Minor, J. S., & Mauer, D.** (1990). Content analysis and intratest performance profiles on the Columbia and the TONI. *Journal of Speech and Hearing Research*, *33*, 375–379.
- Kamphaus, R. W.** (1990). K-ABC theory in historical and current contexts. *Journal of Psychoeducational Assessment*, *8*, 356–368.
- Kamphaus, R. W., & Reynolds, C. R.** (1984). Development and structure of the Kaufman Assessment Battery for Children (K-ABC). *Journal of Special Education*, *18*, 213–228.
- Kaufman, A. S.** (1978). Review of the Columbia Mental Maturity Scale. In O. K. Buros (Ed.), *Mental measurements yearbook* (8th ed., pp. 299–301). Highland Park, NJ: Gryphon Press.
- Kaufman, A. S., & Kaufman, N. L.** (1983). *K-ABC: Kaufman Assessment Battery for Children*. Circle Pines, MN: American Guidance Service.
- Keith, T. Z.** (1985). Questioning the K-ABC: What does it measure? *School Psychology Review*, *14*, 9–20.
- Keith, T. Z.** (2001). Review of the Wechsler Abbreviated Scale of Intelligence. In B. S. Plake & J. C. Impara (Eds.), *Mental measurements yearbook* (14th ed., pp. 1329–1331). Lincoln, NE: Buros Institute of Mental Measurements.
- Linkskog, C. O., & Smith, J. V.** (2001). Review of the Wechsler Abbreviated Scale of Intelligence. In B. S. Plake & J. C. Impara (Eds.), *Mental measurements yearbook* (14th ed., pp. 1331–1332). Lincoln, NE: Buros Institute of Mental Measurements.
- Marco, G. L.** (2001). Review of the Leiter International Performance Scale—Revised. In B. S. Plake & J. C. Impara (Eds.), *Mental measurements yearbook* (14th ed., pp. 683–687). Lincoln, NE: Buros Institute of Mental Measurements.
- McCallum, R. S.** (2003). *Handbook of nonverbal assessment*. New York: Kluwer Academic/Plenum.
- McCallum, R. S., Bracken, B., & Wasserman, J.** (2001). *Essentials of nonverbal assessment*. New York: Wiley.
- McCauley, R. J.** (2001). *Assessment of language disorders in children*. Mahwah, NJ: Erlbaum.
- McCauley, R. J., & Swisher, L.** (1984a). Psychometric review of language and articulation tests for preschool children. *Journal of Speech and Hearing Disorders*, *49*, 34–42.
- McCauley, R. J., & Swisher, L.** (1984b) Use and misuse of norm-referenced tests in clinical assessment: A hypothetical case. *Journal of Speech and Hearing Disorders*, *49*, 338–348.
- McDermott, P. A., & Glutting, J. J.** (1997). Informing stylistic learning behavior, disposition, and achievement through ability subtests—Or, more illusions of meaning? *School Psychology Review*, *26*, 163–175.
- McFadden, T. U.** (1996). Creating language impairments in typically achieving children: The pitfalls of “normal” normative sampling. *Language, Speech, and Hearing Services in Schools*, *27*, 3–9.
- McGrew, K. S., & Flanagan, D. P.** (1998). *The intelligence test desk reference: Gf-Gc cross-battery assessment*. Boston: Allyn & Bacon.
- Mehrens, W. A.** (1984). A critical analysis of the psychometric properties of the K-ABC. *Journal of Special Education*, *18*, 297–310.
- Messick, S.** (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–109). New York: American Council on Education and Macmillan.
- Naglieri, J. A.** (2003). *Naglieri Nonverbal Ability Test*. San Antonio, TX: Psychological Corporation.
- Naglieri, J. A., & Das, J. P.** (1997). *Cognitive Assessment System*. Itasca, IL: Riverside.
- Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., et al.** (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*, 77–101.
- Page, E. B.** (1985). Review of the Kaufman Assessment Battery for Children. In J. V. Mitchell (Ed.), *Mental measurements yearbook* (9th ed., pp. 773–777). Lincoln, NE: Buros Institute of Mental Measurements.
- Plante, E.** (1998). Criteria for SLI: The Stark and Tallal legacy and beyond. *Journal of Speech, Language, and Hearing Research*, *41*, 951–957.
- Platt, L. O., Kamphaus, R. W., Keltgen, J., & Gilliland, F.** (1991). An overview and review of the Differential Ability Scales: Initial and current research findings. *Journal of School Psychology*, *29*, 271–277.
- Psychological Corporation.** (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: Author.
- Raven, J., Raven, J. C., & Court, J. H.** (1998). *Standard Progressive Matrices—Parallel form*. Oxford, England: Oxford Psychologists Press.
- Records, N. L., Tomblin, J. B., & Freese, P. R.** (1992). The quality of life of young adults with histories of specific language impairment. *American Journal of Speech-Language Pathology*, *1*, 44–53.
- Reinehr, R. C.** (1992). Review of the Differential Ability Scales. In J. J. Kramer & J. C. Conoley (Eds.), *Mental measurements yearbook* (11th ed., pp. 282–283). Lincoln, NE: Buros Institute of Mental Measurements.
- Reynolds, C. R., & Kamphaus, R. W.** (2003). *Reynolds Intellectual Assessment Scales*. Lutz, FL: Psychological Assessment Resources.
- Roid, G. H.** (2003). *Stanford-Binet Intelligence Scale—Fifth Edition*. Itasca, IL: Riverside.
- Roid, G. H., & Miller, L. J.** (1997). *Leiter International Performance Scale—Revised*. Wood Dale, IL: Stoelting.
- Salvia, J., & Ysseldyke, J. E.** (1998). *Assessment* (7th ed.). Boston: Houghton Mifflin.
- Sattler, J. M.** (2001). *Assessment of children: Cognitive applications* (4th ed.). San Diego, CA: Author.
- Stark, R. E., & Tallal, P.** (1981). Selection of children with specific language deficits. *Journal of Speech and Hearing Disorders*, *46*, 114–122.
- Sternberg, R. J.** (1983). Should K come before A, B, and C? A review of the Kaufman Assessment Battery for Children

- 
- (K-ABC). *Contemporary Education Review*, 2, 199-207.
- Sternberg, R. J.** (1984). The Kaufman Assessment Battery for Children: An information-processing analysis and critique. *Journal of Special Education*, 18, 269-279.
- Stinnett, T. A.** (2001). Review of the Leiter International Performance Scale-Revised. In B. S. Plake & J. C. Impara (Eds.), *Mental measurements yearbook* (14th ed., pp. 687-692). Lincoln, NE: Buros Institute of Mental Measurements.
- Stinnett, T. A.** (2003). Review of the Wide Range Intelligence Test. In B. S. Plake, J. C. Impara, & R. A. Spies (Eds.), *Mental measurements yearbook* (15th ed., pp. 1011-1015). Lincoln, NE: Buros Institute of Mental Measurements.
- Tomblin, J. B., Zhang, X., Buckwalter, P., & Catts, H.** (2000). The association of reading disability, behavioral disorders, and language impairment among second-grade children. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 41, 473-482.
- U.S. Bureau of the Census.** (2001). Census 2000 Summary File 1—United States [Machine-readable computer file]. Washington, DC: Author.
- Van Lingen, G.** (1998). Review of the Comprehensive Test of Nonverbal Intelligence. In J. C. Impara & B. S. Plake (Eds.), *The mental measurements yearbook* (13th ed., pp. 312-314). Lincoln, NE: Buros Institute of Mental Measurements.
- Wechsler, D.** (2002). *Wechsler Preschool and Primary Intelligence Scale—Third Edition*. San Antonio, TX: Psychological Corporation.
- Wechsler, D.** (2003). *Wechsler Intelligence Scale for Children—Fourth Edition*. San Antonio, TX: Psychological Corporation.
- Whitmire, K.** (2000). Cognitive referencing and discrepancy formulae: Comments from ASHA resources. *ASHA Special Interest Division 1, Language Learning and Education*, 7, 13-16.
- Widaman, K. F.** (2003). Review of the Wide Range Intelligence Test. In B. S. Plake, J. C. Impara, & R. A. Spies (Eds.), *Mental measurements yearbook* (15th ed., pp. 1015-1017). Lincoln, NE: Buros Institute of Mental Measurements.
- Woodcock, R. W., McGrew, K. S., & Mather, N.** (2001). *Woodcock-Johnson Tests of Cognitive Abilities—III*. Itasca, IL: Riverside.
- Young, E. L., & Assing, R.** (2000). Review of the Universal Nonverbal Intelligence Test. *Journal of Psychoeducational Assessment*, 18, 280-288.

Received April 5, 2004

Revision received June 30, 2004

Accepted October 7, 2004

DOI: 10.1044/1058-0360(2004/029)

Contact author: Laura S. DeThorne, PhD, Pennsylvania State University, 110 Moore Building, University Park, PA 16802.  
E-mail: lsd12@psu.edu