

---

# Leiter-3: Nonverbal Cognitive and Neuropsychological Assessment

8

Gale H. Roid and Christopher Koch

The need for nonverbal measures in psychology and education is more urgent than ever before. The multicultural nature of schools and society in Western countries is expanding faster in this decade than ever before due to significant immigration. And, the number of languages spoken by students in schools and universities has increased greatly causing the need for non-English measures of ability, achievement, and behavior. A recent report from the U.S. Department of Education (2016) showed 19 prominent languages spoken by students in English Language Learner (ELL) programs. Spanish was the most prominent (71% of ELL students nationally). Other prominent languages included Chinese, Arabic, Vietnamese, Haitian, Russian, Navajo, and 12 other languages ranked second or third in frequency among the 21 U.S. States with 45,000 or more ELL students. Furthermore, cognitive abilities are often listed in the clinical criteria for disorders in the DSM-5 (American Psychiatric Association 2013), frequently as “rule out” conditions. For example, for language, communication, and learning disorders (315.00, 315.1, 315.2, 315.39), the con-

dition cannot be due to low intellectual ability to satisfy the criteria. So, using an estimate of IQ, found to be near average or above, would rule out intellectual deficiency as a reason for the disorder. Given the known probability of learning and communication disabilities in schoolchildren and adults (U.S. Department of Education 2007), combined with the frequency of non-English proficiency, the need for nonverbal cognitive assessment is well established. The recently published Leiter International Performance Scale, Third Edition (Leiter-3) provides a completely nonverbal, comprehensive measure of both cognitive and neuropsychological processes, and is the subject of this chapter.

---

## Goals and Rationale for Leiter-3 Development

The Leiter International Performance Scale (Leiter 1938, 1979; Roid and Miller 1997) has a long history of use in special education and psychology (Levine 1982; Roid et al. 2009). The validity and usefulness of the current edition (Roid et al. 2013) rests in part on this long history of research and development. This chapter details the development of the 3rd Edition.

Although the theoretical background and nonverbal nature of the Leiter have been highly praised, it was generally felt that the original Leiter “lacked the necessary technical characteristics to make it psychometrically adequate” (Salvia and Ysseldyke 1991, p. 208). For these reasons, the goals of the Leiter-R and Leiter-3

---

G.H. Roid (✉)  
Testing Research, 1838 Mistwood Drive NE, Keizer,  
OR 97303, USA  
e-mail: galeroide@gmail.com

C. Koch  
George Fox University, 414 N, Meridian Street,  
Newberg, OR 97132, USA  
e-mail: ckoch@georgefox.edu

standardizations included a full range of psychometric studies and a nationally representative norm sample.

The Leiter-3 is an individually administered, nonverbal battery of 10 subtests that measure three major dimensions of cognitive ability—General Intellectual Ability (IQ), Nonverbal Memory, and Processing Speed. The Leiter-3 was designed for children, adolescents, and adults, ages 3 years, 0 months to 75+ years. Each subtest comes with pantomime or other nonvocal (unspoken, ‘nonvocal’) instructions so that neither the examiner nor the examinee needs to speak aloud during the administration of each subtest. Thus, the Leiter-3 provides a fully nonverbal, nonvocal set of subtests. For hearing/speaking examinees, the examiner is encouraged to build rapport by speaking with the examinee between subtests. The Cognitive Battery subtests (4 subtests and one alternative) can be administered in approximately 30–40 min. These general cognitive subtests provide a nationally standardized estimate of nonverbal IQ. The remaining five subtests measure various memory, attention, and cognitive interference processes. This Attention Memory (A/M) Battery can be completed in another 20–30 min and provides a supplement for measuring cognitive processes associated with disorders. The A/M subtests allow examiners to identify strengths and weaknesses in neurocognitive processes suspected of affecting the IQ estimate or providing evidence for additional neuropsychological testing.

## History of the Leiter

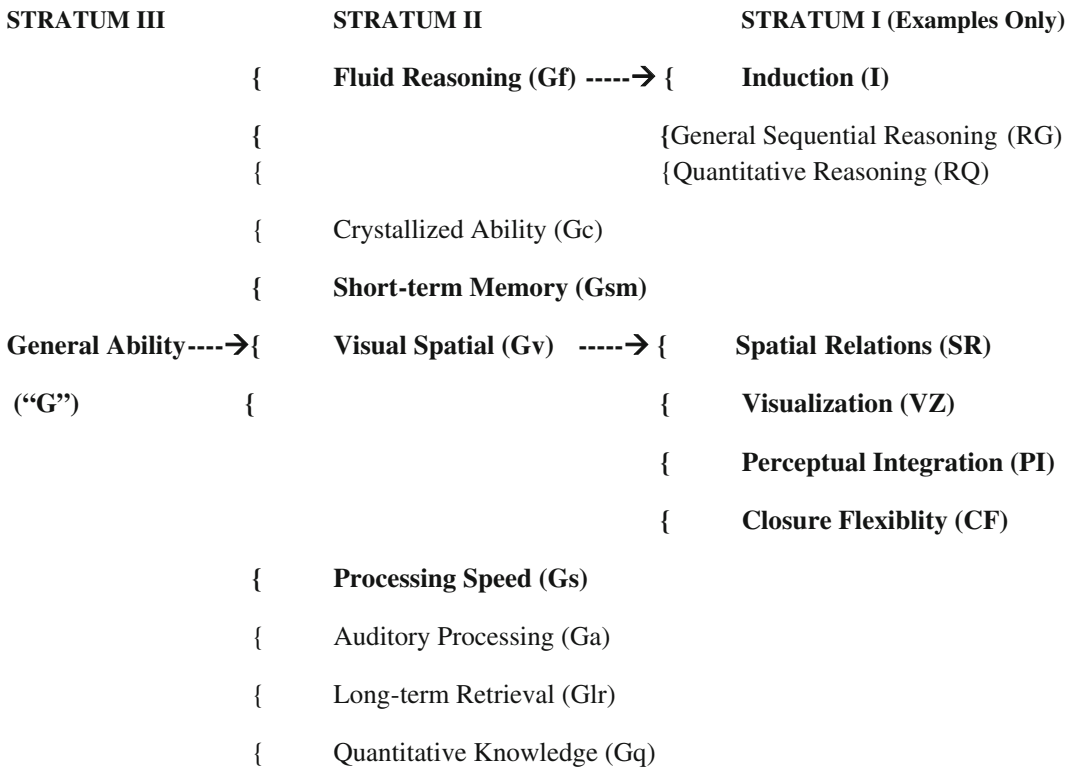
Leiter (1938) developed the test for children and adolescents with multi-ethnic backgrounds (in Hawaii and California), using a unique “block and frame” response method, which required examinees to move wooden blocks into slots in a wooden frame to complete puzzles, figure completion, numerical series, visual matching, and sequences of geometric or pictorial objects. The test did not require spoken directions from either the examiner or examinee; instead administration relied on pantomime directions and the obvious

movement of blocks for responding. The test was totally revised by Roid and Miller (1997), who created a modern battery of 20 subtests with nationally standardized scale scores, for ages 2–20 (Roid et al. 2009). The test used a series of response cards or pointing responses to measure various aspects of cognitive ability (10 subtests) and 10 subtests measuring memory and attention factors. Leiter-3 is the most recent iteration.

## Theoretical Rationale

**Cognitive Model.** A unified cognitive ability model has emerged from a number of independent researchers over the last 50 years. Carroll (1993) proposed a three-stratum theory of cognitive abilities based on factor analysis of more than 460 data sets, including special education and multi-ethnic samples. Carroll’s model included aspects of the fluid–crystallized theory of Horn and Cattell (1966) and the three-level hierarchical model documented by Gustafson (1984). Although variations in the number and names of the factors occur in different studies, one consensus shows an integrated 8-factor Cattell–Horn–Carroll model detailed by Flanagan et al. (2013), and Schneider and McGrew (2012). At the highest level (Stratum 3) is a general intelligence or “g” factor (see Fig. 8.1). At the second level (Stratum 2) are broad factors identified as Fluid Reasoning, Crystallized Ability (or Knowledge/Verbal Comprehension), Short-Term Memory (or Working Memory), Visual-Spatial, Long-term Retrieval, Quantitative Reasoning, Processing Speed, and Auditory Processing. At the bottom level (Stratum 1) is a large number of “primary” factors, nested within the second-level factors. For example, Spatial Relations, Visualization, Perceptual Integration and Closure Flexibility are nested within Visual-Spatial Ability.

**Relationship to Theories of Autism Spectrum Disorder (ASD).** Clearly, autism is an increasingly prevalent condition in the U.S. (CDC 2007; U.S. Department of Education 2016) and often requires nonverbal assessment (Minshew and Goldstein 1998). To meet this



**Fig. 8.1** Cattell–Horn–Carroll model of cognitive abilities (Leiter-3 dimensions shown in boldface type)

need, the Leiter-R (Roid and Miller 1997) was highly recommended by Klinger et al. (2012) for use with children diagnosed with ASD. Because of this and other positive reviews from clinicians, the Leiter-3 was purposely designed for use in ASD intellectual assessment. Theories of ASD were studied in detail during the development of Leiter-3 (e.g., Mayes and Calhoun 2003, 2004; Volkmar et al. 2004), including various grant proposals submitted for research funding by the senior author. For example, the theory proposed by Minschew and Williams (2007) was based on a cognitive theory of ASD called complex information processing (CIP) to explain the observed deficits in autism (Williams et al. 2006). The CIP model arose from two observations of characteristics exhibited by ASD individuals: (a) deficits on neuropsychological tests in conceptualization and complex memory and language, combined with good motor abilities, and (b) patterns of superior function in attention, simple memory and language, and visual-spatial

abilities. The Leiter-3 specifically included measures targeted for complex information processing in ASD (e.g., the subtests “Sequential Order,” and “Repeated Patterns” within the nonverbal fluid reasoning portion). And, former users of Leiter who assess clients with ASD strongly recommended the “hands on” method of using the blocks to show their client responses to test items. Thus, the Leiter-3 provides clinical assessment to supplement ASD evaluations to the CIP model (Mayes and Calhoun 2003, 2004; Bishop et al. 2006). For more information on the importance of cognitive assessment for ASD, consult references such as Kushner et al. (2007) and Mayes and Calhoun (2004).

---

### Description of the Leiter-3

The Leiter-3 includes two sets of subtests—a set of five cognitive ability subtests with four of them providing a nonverbal IQ and a

supplementary set with two attentions, two memories, and one cognitive interference (Stroop 1935) subtests. The sets of subtests can be used separately or together. When used together, they have the advantage of being standardized on exactly the same standardization sample. Hence, the statistical significance of differences between subtest and composite scores can be calculated accurately for strength and weakness analysis.

### Leiter-3 Subtest, Administration and Scoring

Two formats of stimuli and response mode are used in these subtests: (a) colorful pictures in the stimulus book presenting the items with cards used by the examinee to respond and (b) arrays of stimuli on the easel seated above the slotted frame are used with rounded plastic blocks, with printed graphics, which can be moved into slots of the frame.

In the standard subtest order, Figure Ground is first, using the picture/card method. Form Completion is second using a combination of picture/card and block/frame, and the remaining three subtests employ the block/frame method (Classification/Analogies, Sequential Order, and the optional Visual Patterns). Subtests are described below:

#### 1. **Figure Ground (FG).**

FG is a basic visual interference task, but compounded by distractions. The examinee searches for a target object on the Stimulus page that is pictured on a response card. The complete target object is included in the stimulus, but since the object is embedded in increasingly complex backgrounds, the backgrounds can mask the object so that it becomes difficult to recognize. Figure ground perception is a construct which has been widely studied over decades with early studies provided by Gottschaldt (1928). Thurstone and Thurstone (1962) found that performance on this task was associated with visual closure and correlated with freedom from distractibility. Performance on this task

is related to the cognitive flexibility of the individual (i.e., the ability of perceptual shifting) since the individual must shift attention between a discrete figure and complex backgrounds, necessitating a change of perceptual set (Talland 1965). Figure Ground also requires that the individual have adequate visual scanning skills and an effective search strategy. It is a subtest which requires good inhibition from the individual, as impulsivity will result in pointing randomly to similar shapes rather than focusing on the target object embedded in the figures. This subtest lends itself to clinical qualitative observations of process such as perceptual bias to one side of the stimulus page, misidentification of objects, or perseveration (Christensen 1979).

#### 2. **Form Completion (FC).**

This subtest requires organization of disarranged pieces. Cognitive flexibility is tested by requiring the examinee to scan between parts and the “whole” to arrive at a solution, a process that is mostly deductive. As the examinee moves back and forth between the stimulus and response, working memory permits the individual to hold both the stimuli and possible responses in mind simultaneously as the images are constructed and deconstructed. Items are conceptually related to previous research describing Figure Ground processes (Hooper 1983) with easiest items being meaningful familiar objects. This subtest requires perceptual scanning, recognition, and the ability to perceive fragmented percepts as wholes. This task assesses a “higher level of perceptual ability” than a matching task (Dee 1970). Visual organization tasks, such as the type assessed in Form Completion, require synthesizing activities, whereas visual interference tasks, such as the type assessed in Figure Ground, require discrimination of figures from interfering elements.

#### 3. **Classification/Analogies (CA).**

Following extensive scaling studies using item-response theory (Lord 1980) analysis, items from the Leiter-R classification subtest

were combined with the more difficult Design Analogies' items to form the CA subtest. The classification tasks among the easier items of the subtest require pattern recognition and mental shifting of concepts. This subtest progresses into functional classification where objects which "belong together" are grouped because of usage rather than size, shape, or color (Lezak 1995; Wang 1984; Nelson 1976).

For the analogies items, in the higher difficulty level (e.g., above age 6), items are presented in the classical "matrix reasoning" format. The matrix items require solving visual analogy problems presented in tables of two columns and two rows of objects, or more of each dimension, with one "box" of the table missing. The CA items were designed to measure pattern analysis and prediction of "what goes next" in a series of objects, and classifying the most common types of abstraction and concept formation (Lezak 1995; Wang 1984; Nelson 1976). Relationships are induced from concepts (i.e., the bed "goes with" the pillow) or elements (i.e., all the pictures with a shadow "go with" the block showing a shape with a shadow). This subtest is also a measure of matrix reasoning ability which has been widely researched in previous tests (Elliot 2008; Raven et al. 1998; Roid 2003). The individual must select an appropriate response from the possible blocks based upon the perceived relationship between the figures in the matrix. Classification/Analogies measures the ability to generate rules from partial information, and inductively hypothesize what piece would complete the whole pattern. This type of task appears to be a valid measure of general ability (Carroll 1993).

#### 4. **Sequential Order (SO).**

The subtest requires nonverbal reasoning ability and rule generation for analyzing sequential information (Carroll 1993). The individual must understand the relationship between stimuli in order to find the missing elements at the end or in the middle of the series. The ability of the individual to

perceive sequential patterns and determine the rules that govern the relationships between pictures is assessed. On this subtest, the "whole" is the final pattern which is induced from multiple stimuli.

#### 5. **Visual Patterns (VP).**

This subtest was created from the Leiter-R Matching subtest, combined with the Repeated Patterns subtest. At the youngest ages, this optional subtest evaluates the individual's basic ability to match visual stimuli with no memory component. This has been described as perceptual acuity, measuring visual discrimination and awareness of spatial orientation (Elliot 2008). At the youngest age levels, the task is simple, with matching by color or shape and large features being prominent. As the task progresses, attention to detail is required, as the matching dimensions become smaller and less prominent. The subtest does not include rotations or pattern reversals. It requires the ability to scan and make visual comparisons between figures while the individual is tracking several stimuli simultaneously, such as number, orientation of parts, and location of lines. This task requires basic visualization processes, but also necessitates patience and freedom from impulsivity on the part of the individual, as he or she must check different stimuli against the model as the items increase in complexity. The subtest is similar to pattern completion fluid reasoning tasks developed by Thurstone and Turnstone's (1962).

#### 6. **Attention Sustained (AS).**

The subtest relies on a cancellation task designed to assess prolonged visual attention, and requires good visual scanning and motoric inhibition on a rapid repetitive motor task (crossing out stimuli). This classic processing speed task was used clinically by Albert (1973), employed in research on learning disabilities by Rourke (1988), and autism by Goldstein et al. (2001). Under the name "cancellation," it was included in the Wechsler (2003, 2008) scales. Although a motor response is required, the motoric demands are quite limited. During the task,

clinicians can observe the process by which each individual accomplishes the task and if the individual does more poorly on the last item of the set of four, where the stimuli are arrayed in a random manner on the page instead of being presented in straight rows, it may be evidence of visual-spatial inattention (Lezak 1995). Poor performance on this task may reflect an underlying attention problem that is affecting performance on the other cognitive subtests. This subtest can reflect the general slowing of attention due to various brain conditions, stages of dementia, and effects of other insults to neurological function (Lezak 1995).

7. **Forward Memory (FM).**

The FM subtest measures sequential memory span. Also, it requires an organized processing style. Sets of pictures (e.g., boat, car, and shoe) are shown and the examiner touches a sequence of pictures (beginning at one and increasing to seven pictures in the most difficult items). The examinee is taught to touch the pictures in exactly the same way as the examiner. Thus, in addition to remembering the sequence of pictures, the individual must also inhibit the memory of previous sequences of pictures. This subtest, along with Subtest 9, Reverse Memory, is similar to the Digit Span subtest of the Wechsler tests where there are both forward and reverse sequences of digits to be recalled. However, the Leiter-3 version uses visual material without vocalized examiner directions. Another advantage of the nonverbal format is that the examinee does not have to hear and comprehend the name of the picture—only its spatial position, color, and visual features. Digit Span and Forward Memory measure similar constructs—short-term memory, and in the reverse task, working memory where information is stored and manipulated in short-term memory.

8. **Attention Divided (AD).**

The AD subtest is new in Leiter-3 and employs a game-like format. The subtest measures the ability to play a game of slapping targeted cards (marked with a red

triangle) as cards with and without the target are sequentially placed in front of them. Then, examinees must learn to place soft foam disks into a container as quickly as possible. For young children, the task includes only 12 yellow disks, but older children and adults have more disks, sorting red and yellow disks into separate containers. After learning both tasks, the examinee must do both tasks at the same time (within the time it takes to present all the cards). Thus, the subtest measures concentration and executive processing of mental and motor behaviors while completing two different tasks at the same time. If the individual has difficulty with this “double or multiple tracking” (Lezak 1995, p. 551), he or she is likely to slow down or break down during the task. One young adult with diagnosed ADHD in the tryout sample vocalized his difficulty by saying “I can’t do this!” The ability to attend to more than one thing at a time has been found to be very informative for teachers and parents interested in helping children pay attention in noisy classrooms. And, difficulty with dividing attention is a sensitive measure of subtle neurological deficits and, to an extent, autism (Mundy and Crowson 1997). This difficulty may be the only documentable mental change after a head injury (Lezak 1995) or other neurological condition and may be clinically important to observe, particularly, for examinees who deviate from the instructions.

9. **Reverse Memory (RM).**

The RM subtest requires touching pictures as in the Forward Memory subtest. But, this subtest measures working memory because the examinee touches pictures in *reverse* order from that required in Forward Memory. Individuals who have less mental flexibility or become confused easily may have difficulty switching tasks from Forward Memory. For this reason, the RM subtest is administered after a diversion (e.g., another subtest, AD is administered). The RM task is complex, requiring the individual to store and juggle information using mental effort and good working memory. Reverse Memory

does not evaluate the same cognitive processes as Forward Memory and the two scales have been found to be disparate in a number of clinical groups (Banken 1985; Lezak 1995; Reynolds 1997; Sullivan et al. 1989). Raw score differences between forward and reverse memory appear to be varied and not as predictable as those with digits forward and backward (e.g., Wechsler scales typically show a 2-digit difference on average, Wechsler 1991). It appears that the gap between FM and RM on Leiter-R or Leiter-3 may increase as the span increases. For example, younger individuals may do four pictures forward, but three pictures in reverse (87%), while older individuals may do eight pictures forward, but only five pictures reversed.

#### 10. **Nonverbal Stroop (NS).**

The subtest is a nonverbal version of a classic, cognitive interference test. The color-word Stroop test is one of the most widely used tasks for examining cognitive processing. The task is based on Stroop's (1935) original experiment in which he presented the words red, blue, green, brown, and purple twice per row in a  $10 \times 10$  matrix using incongruent ink and asked participants to name the color of the ink (Experiment 2). The time to complete the task was compared to a control condition in which the same colors appeared in a  $10 \times 10$  matrix but as color blocks instead of color words. It took significantly longer to name the colors of the incongruent color words than the colors of the blocks. This difference between the two conditions is commonly referred to as Stroop interference. There have been a number of variations to the task over the years including a picture-word (Rosinski et al. 1975), sorting (Tecce and Happ 1964), and emotional Stroop task (Cha et al. 2010). Of particular importance is the finding that Stroop interference can occur when the color and word are presented together but are not integrated (Dalrymple-Alford and Budayr 1966; Dyer 1973). A color block appearing alongside a color word would be an example of a nonintegrated pair. It is important to note that the color and word need to be

presented close to each other in both time (Dyer and Severance 1973) and space (Kahneman and Chajczyk 1983; Kahneman and Henik 1981) in order to be processed "together." It is generally assumed that the word is processed faster than the color from a color-word pair and must, therefore, be inhibited in order to respond correctly to the color (Dunbar and MacLeod 1984; Posner and Snyder 1975). Carroll (1993) further suggested that performance on the Stroop task is related to naming speed and reading speed. The central role of words in these explanations of the Stroop task suggests that the task itself is not possible to administer nonverbally. That assumption was challenged by presenting two color blocks, instead of a color block and a color word, and asking participants to name the color of a target block. Responses were faster when the two blocks were the same color than when the two blocks were different colors (e.g., Koch and Kubovy 1996). Thus, interference scores similar to the color-word task can be obtained without using words. Performance on this revised Stroop task parallels the performance of a variety of clinical groups with the traditional color-word task (cf., Koch and Roid 2012).

A similar task was developed for the Leiter-3 using two colored circles. To ensure that the two circles were processed together instead of as independent circles, two Gestalt grouping principles were used. First, the two circles in a pair were connected with a line creating a dumbbell (connectedness). An oval was also drawn around the pair of circles (enclosure). A nonverbal response was also needed to make the task entirely nonverbal. Consequently, a test format was selected in which a target stimulus was presented on the left with alternative stimuli presented to the right. The goal of the task is to identify, or cross out, the matching correct response among the alternatives while ignoring the distracters. Participants are given 45 s to identify as many target matches as possible. More items were identified when the circles were the same color compared to when the two circles

within a pair were different colors (Koch et al. 2011; Koch and Barr 2010). The difference between the two conditions of the task is an indicator of interference.

## Scores Provided

The raw score for each of the Leiter-3 subtests is typically the sum of the correct responses marked on the Record Form. For each subtest, the distribution of raw scores, at each age, was converted into normalized scaled scores with a mean of 10 and a standard deviation of 3. To increase the sensitivity of scores in the gifted range, the scaled scores range from 1 to 20 which was used in the subtest profiles of intellectual-ability tests, instead of the standard 1–19 range. The age grouping intervals for the norms vary by age, ranging from 2-month intervals for examinees who range in age from 3 to 10, to 10-year intervals for individuals in the 30–70-year-old range.

To convert raw scores into scaled scores for each of the subtest profile scores, the examiner uses the standard norm tables for the age range that fits the examinee’s chronological age. Scores are then recorded on the front of the Record Form, in a manner similar to other ability tests.

## Nonverbal Stroop Scores

Stroop (1935) required participants to name the colors of 100 color words and recorded the time to complete the task. The Color and Word Stroop Test (Golden et al. 2003) requires examinees to name as many colors as they can within 45 s. The number of correct colors is recorded. Therefore, the task has been presented measuring both time and number (accuracy). However, even if time is held constant (e.g., 45 s), it is still possible to use the total correct score as an index of speed—the more the correct items, the faster the examinee has responded accurately. Consequently, the NS task in the Leiter 3 has norms for the number of correct and number incorrect items for both the congruent and incongruent tasks. The number correct scores are primary and used in the main profile of scaled scores. The numbers

incorrect for both the congruent and incongruent have interpretative norms in the “Supplemental Attention/Memory Scores” section of the Record Form. Other indexes for examinee processing speed can be easily calculated by examiners. For instance, dividing the number of correctly identified colors by 45 s results in a colors/second time that can be used as an indicator of processing speed.

## Supplemental A/M Subtest Scores

There are five Supplemental A/M scores for an in-depth analysis of the accuracy of the individual’s performance, focusing often on errors or incorrect responses. These scores were standardized on the same normative sample as the main subtest scores, but often have more restricted ranges of scores because of the predominance of individuals with zero errors. Thus, the scores are most useful as diagnostic information supplemental to the main profile scores with a profile chart that ranges from 0 to 12 rather than 0 to 20 as done with the main profile scores. Table 8.1 shows the names of the five supplemental scores and a brief explanation of their scoring and purpose.

## Nonverbal IQ and Composite Scores

To simplify scoring and facilitate the profiling, all IQ and Composite scores have been placed on the IQ scale (mean of 100 and standard deviation of 15). However, be aware that the attention and memory (A/M) Composite scores are not included in IQ calculations, for they are not measures of global intellectual ability. Rather, the A/M composites are summative indexes of factors of neuropsychological processes that are more specific than general ability.

## General Ability (Nonverbal IQ) Scores

One general ability score—the nonverbal intelligence quotient (IQ)—is available from the Leiter-3. The foundation for the IQ score is the



**Table 8.1** The supplemental attention/memory subtest scores

Attention sustained errors (ASe)	The number of incorrectly marked objects is a raw score, converted into a scaled score for exploring possible attention-deficit or impulsive responding
Attention divided correct (ADc)	The number of cards slapped and number of foam pieces placed correctly (added together) are converted into a scaled score. The score shows ability to split attention to two tasks simultaneously
Attention divided incorrect (ADi)	The number incorrect on both tasks, added together, form a score that can indicate poor motor ability, coordination issues, slowness of response, or poor executive functioning
Nonverbal Stroop congruent incorrect (NSci)	The number of incorrect markings for color-matched items on the first trial of NS can indicate evidence for a response pattern similar to one obtained from atypical populations (e.g., TBI, ADHD) or reflect other neurological or visual-attention difficulties
Nonverbal Stroop incongruent incorrect (NSii)	The number of incorrect markings for color-mismatched items on the second trial of NS can indicate even greater deficits in motor, visual, or neurological processing than the congruent incorrect score

Fluid Reasoning		Visual-Spatial		(optional Visual)
Classification-Analogies (CA)	Sequential Order (SO)	Figure Ground (FG)	Form Completion (FC)	Visual Patterns (VP)

Processing Speed

Attention Sustained (AS)	Nonverbal Stroop Incongruent—Number correct (NSic)
--------------------------	--

Nonverbal Memory

Forward Memory (FM)	Reverse Memory (RM)
---------------------	---------------------

Additional Attention Subtest

Attention Divided (AD)
------------------------

**Fig. 8.2** Structure of the Leiter-3 subtests and composites nonverbal IQ

sums of the scaled scores for the subtests that compose the IQ estimate. The IQ score provides a measure of “g,” or general nonverbal intelligence. Figure 8.2 shows the subtest composition of the IQ score. There are four subtest-scaled scores included in the calculation of IQ. Preference is given to using the first four cognitive subtests (FG, FC, CA, and SO) and only use Visual Patterns as a substitute if one of the other four subtests is spoiled. The reason for using Visual Patterns as a substitute is that it is slightly

lower in reliability at certain age levels compared to the other four.

To obtain the IQ score, use the scaled score information on the Record Form to organize all the subtests for summing of scaled scores for the IQ or Composite scales. Scaled scores have a mean of 10 and a standard deviation of 3. To provide more precision of measurement at low and high ends of the IQ continuum, the Leiter-3 scaled scores range from zero to 20, instead of the typical 1–19 range. Using the sum of scaled

scores to estimate IQ should be familiar to many school psychologists and other assessment professionals.

The percentile equivalent of each standard score can be found in the appendix of the manual or from any standard percentile table because the IQ and Composite scores are normalized. Confidence intervals are also recommended, especially for the IQ score, because they emphasize to parents, teachers, and other professionals that all such scores have an element of measurement error. Tables for constructing the confidence intervals are provided in the appendix of the test manual. These tables provide the magnitude, based on standard errors of measurement (SEM), to create confidence intervals for the IQ and Composite scores. The user simply subtracts and then adds the confidence interval value, based on the age grouping. For example, a 99% interval is created by multiplying the SEM by 2.58 and using the rounded value to add and subtract from the IQ estimate.

### **AM Battery Composite Scores**

There are two Composite scores available on the Leiter-3 supplemental A/M subtests. Best practice for assessment professionals starts by considering the individual's needs and reason for referral that required the assessment. The Composite scores of Nonverbal Memory and Processing Speed provide a higher degree of test reliability than the individual subtests, and therefore should be the first line of interpretation. To simplify scoring and facilitate the profiling of composites, all scores have been placed on an IQ-type metric (mean of 100 and a standard deviation of 15). However, the metric should not imply that the Composite scores are interpreted as supplemental estimates of nonverbal IQ.

### **Use of the Criterion-Referenced Growth Scale**

The Leiter-3 growth scores have been developed to counteract a well-deserved criticism of

standardized norm-referenced scales—that norm-referenced scaled scores do not always provide detailed information about the skills of an individual or the growth that an individual is achieving. The limitations of norm-referenced scores are particularly present for individuals who are significantly delayed for their age. For example, if a child is functioning at a very low level (scaled score of 1 or 2), he or she may never show an increase in that score due to continual comparison to the normative group for each older age category.

The metric of the growth values and the growth scale scores is similar to the W-scale used in the Woodcock series of tests (e.g., Woodcock and Dahl 1971; Woodcock et al. 2000), with scores centered on a value of 500 set at the beginning of 5th grade (10 years, 0 months). Each task on the Leiter-3 has a value from 380 to 560 along the growth scale located at the top of the Growth scale record form. The estimate of the degree of difficulty for each task is expressed by its location on the growth scale. Using the tables in the manual, it is possible to convert the raw scores for each subtest, each composite, and each IQ estimate into Growth scale scores and to identify the item growth values for each item passed or failed on the Leiter-3. These converted scores are provided to assist with program planning, determining change over time, and explain results to clients, parents, and teachers. To determine an individual's growth value on each item, first locate the item numbers for all items that the individual passes. Next, use the tables in the manual to find the actual growth values for each item. Separate tables are provided for Core Cognitive and A/M items. Those values will explain the relative item difficulty of each item passed and failed by the individual. The individual's item growth values can range from approximately 380 to 560. In addition to values for the items passed by, growth scale scores providing criterion-referenced ability estimates are available for subtests, composites, and Nonverbal IQ (Woodcock 1999). The growth scale is consistent across ages and across different collections of subtests, and provides an "anchor scale" for referencing all subtest and composite scores.

## The Leiter-3 Examiner Rating Scale

The Examiner Rating scale was originally developed for Leiter-R (Roid and Miller 1997) and is repeated in Leiter-3 with the same content. The scale has been widely used by examiners and researchers (e.g., Nordlund 1998) to provide an assessment of test-taking behavior and sensory or social-emotional factors in the examinee. The domains included in the ratings include attention, organization/impulse control, activity level, sociability, energy, feelings, regulation, anxiety, and sensory reaction. Each domain has four to eight items rated on a scale from zero to three and the sum of the ratings provide raw scores. The raw scores can be converted into scaled scores (mean 10, standard deviation 3). Two groupings of the separate domains form the Composites—Cognitive/Social and Emotional/Regulations. The Composite scores have the same metric as the Cognitive Composites—mean 100, standard deviation 15. However, in terms of national norms, most individuals have positive test behavior and typical, adaptive social-emotional behaviors. Therefore, the range of domain and Composite scores for the Examiner Rating scales are somewhat restricted above the mean scores.

## Standardization and Psychometric Properties of the Leiter-3

**Stratification.** Collection of the standardization sample began in 2010 employing the census data (U.S. Census 2009) and was updated in 2011 (U.S. Census 2011). A stratified random sample of individuals was developed, taking into account all the strata of the plan—age, gender, race/ethnicity, educational level (parent or adult individual), and geographic region. After examiners (“field researchers”) were recruited for all four census regions of the United States, each was given a detailed description of the cases needed for their region.

**Examiners.** Training included a detailed description of the sampling strata, which was also printed in the demographic section of the

standardization Record Form completed for all cases. The process of collecting cases for the standardization continued through 2011 and included a number of steps. For example, each field researcher developed a list of major sources for obtaining a sample in his/her vicinity for individuals within the appropriate age range. They then obtained administrative permission, selected participants based on a master list provided by the publisher, administered the Leiter-R, and conveyed the results to the publisher.

**Geographic Representation.** With 150 field researchers selected across all four U.S. Census regions, geographic randomization of the sample was enhanced. These examiners were selected to participate in the study because they were qualified professionally based on work history, education including measurement instruction, and experience with individually administered tests. Examinees were selected based on certain criteria, i.e., they had no severe physical, mental, or emotional impairment (unless included in clinical validity studies separate from norms), or other biological risk factors, and could follow basic directions.

## Description of the Sample

**Age and Gender.** Examinees were recruited at each year of age (including an oversampling of age 2.5–3 for scaling purposes). Also, identification of the sex (‘gender’ in this manual) was required. Some of this information was given by the parent in the cases of younger children or atypical, special cases where self-reporting was impractical. Then, categories of ages (16 categories) were used to select 1603 cases. Details of the categories are described in the Leiter-3 manual, with percentages according to each state, drawn to reflect the U.S. Census, updated as of 2011. For example, the categories for older individuals include more females, given their expected longevity in the population. The final totals, 49.8% male and 50.2% female was, as planned, as close as possible to the 50/50 target. In addition to the 13 categories of age in the

sampling plan, an additional 5-age-group categorization was used for many of the psychometric studies (e.g., factor analyses).

**Race/Ethnicity.** The Leiter-3 standardization sample includes proportions of White (Caucasian and non-Hispanic), African-American, Asian American, Hispanic, Native Americans (those with tribal affiliation or self-identification for this category), and “other/Mixed,” an increasingly important category used in recent Census studies. Also, individuals of Hispanic origin were identified in a special category, and were excluded from other categories (e.g., Anglo-Americans were defined as Caucasian non-Hispanic and African-American as a category excluding those of mixed Hispanic and African-American origin). Because of the primacy of language, the Hispanic category was seen as an important, mutually exclusive category of ethnicity, rather than a separate dimension paralleling racial origin. The match of the Leiter-3 sample data to U.S. Census data was extremely close for all Race/Ethnicity categories, as shown in the manual.

**Educational Level.** Based on the past experience with the estimation of socioeconomic backgrounds of participants in test standardizations (e.g., Roid and Miller 1997; Roid 2003), each examiner collected the educational level of examinees. For practical purposes of keeping the number of sampling strata simple and for the benefit of planning and training of examiners, only four levels of educational attainment were employed with categories similar to U.S. Census reporting categories. The four levels were (1) less than high school (11 years or less and no diploma obtained), (2) completion of 12 years and high school or General Educational Development (GED) programs, (3) completion of 1–3 years of college or post-secondary education without a bachelor degree or equivalent, and (4) any bachelor or higher degree including advanced professional degrees in law, medicine, engineering, business, etc. The proportions of examinees at each level show an extremely close match to comparable Census percentages.

**Geographic Region.** The four U.S. Census regions of the United States were used as sampling categories—Northeast, Midwest, South,

and West, and again show an extremely close match to U.S. Census data. After advertising widely for examiners nationwide, the publisher selected qualified and experienced examiners in each of the four geographic regions. Also, the senior author conducted training sessions for examiners in selected regions, including the West (California, Oregon, Washington, and Colorado) and South (Texas, Oklahoma, and Florida). Examiners were also recruited at regional psychological conferences in locations such as Philadelphia, Orlando, and Chicago. The total number of states sampled in the standardization was 36.

## Other Characteristics of the Sample

**Rural versus Urban Locations.** A demographic that now plays a reduced role in affecting cognitive performance averages is the size of the community in which the examinee lives (Roid 2003). However, the sample is well balanced in terms of this variable, with 24.2% of examinees coming from rural homes (small town less than 2500 population or farm and ranch locations) and 75.8% from urban homes (towns, suburban areas, medium-, and large-population cities).

**Special Groups Sampling.** Examiners were required to report any primary or secondary diagnoses of children, adolescents, or adults with DSM-IV or official special education diagnoses for any of the following conditions: Autism, Alzheimer’s, delays in speech, hearing, or motor abilities, traumatic brain injury, intellectual deficiency or delay, ADHD, giftedness, learning disabilities (Reading or Other), English-as-Second Language (ESL), English Language Learner (ELL), or local designation for second-language speakers. Attempts were made to find “pure” cases with few multiple diagnoses, and this was achieved in 91% of the special cases, with the main exception being a large overlap between Reading and “Other” learning disabilities (e.g., Math, Writing, etc.). Certain examiners were assigned clinical cases for specific validity studies, but most examiners collected “typical” normative cases. For students

in school, academic difficulty areas (spelling, math, reading, and handwriting) were also noted. When no diagnosis was indicated by parents or adult subjects, designation of “typical” was coded for examinees.

Because of the wide-spread implementation of U.S. policies to include individuals with special needs into the standard programs of schools and colleges, 10.7% of cases in the standardization sample had special-group status, including “gifted students” (those enrolled in official school gifted programs) and English-as-Second-Language (ESL) or Emerging Language Learner (ELL) programs.

---

## Reliability

The technical qualities of the Leiter-3 were researched extensively in the standardization process. In addition, the validity of the instrument was bolstered by the validity studies of the Leiter-R, from which much of Leiter-3 was derived. Except for the new subtests within Leiter-3 (AD and NS), the previous validity studies of the Leiter-R provided evidence that the new edition would have high strong validity characteristics. Studies are reviewed in the manual addressing content, construct, and criterion-related validity.

### Internal Consistency Reliability

**Cognitive Subtests.** Estimates of the internal consistency reliability of the Cognitive Battery subtests range from 0.79 to 0.95. Importantly, because Cronbach’s (1951) alpha coefficients were employed, the estimates in figure are estimates of the lower bound of internal reliability (Lord 1980). Because the reliability coefficients are the basis of standard errors of measurement, significance of subtest score differences, and other key elements of test interpretation, conservative estimates were deemed most beneficial to prevent “over interpretation” of small differences between profile scores. The optional subtest Visual Patterns (VP) had the lowest median

reliability (0.78) and the longest subtest, Sequential Order (SO) had the highest (0.95).

**A/M Subtests.** Internal consistency estimates for the A/M subtest range from 0.70 to 0.81. Because the AS number correct score and the NS Effect score have “parts” (different pages for AS and two components for Stroop Effect), alpha coefficients can be calculated for these scores.

The part scores for the Stroop Effect, the Congruent, and Incongruent trials required test-retest coefficients because they are timed tests (see section on Test-Retest Reliability). The AD subtest is a timed subtest also (completed as soon as the cards are distributed to the examinee) and has no “part scores” to calculate alpha coefficients (see section on Decision-Consistency reliability).

**IQ and Composite Scores.** In addition to the nonverbal IQ score there are two composite scores available for the Leiter-3–Nonverbal Memory and Processing Speed. For age groupings, ages 3–6, 7–11, 12–16, 17–29, and 30–75+, Table 8.2 shows the composite reliabilities for these scores. Coefficients were calculated using the formula for a composite of several tests as described by Nunnally (1978, p. 246). Average reliabilities for each age grouping were computed using Fisher’s  $z$ -transformation of the obtained correlation reported above, then summing, averaging, and reconverting the average  $z$ -value into the correlation metric. As shown, the reliability of composites and IQ scores are generally higher than any of the individual subtests because of their increased length and precision.

Many psychometric experts have suggested that reliability coefficients above 0.90 are required for making life-changing decisions about individuals, such as designation as intellectually deficient (Gregory 1996; Salvia and Ysseldyke 1991). In general, the IQ and Composite scores of the Leiter-3 qualify for use in cases of major decisions about individuals.

### Test-Retest Reliability

A total of 156 individuals, ages 3–79 years (mean 21.4), were administered the Leiter-3 on

**Table 8.2** Composite score reliabilities by 5 age group

Age group	Nonverbal IQ		Processing speed		Nonverbal memory	
	<i>N</i>	Rel	<i>N</i>	Rel	<i>N</i>	Rel
3–6	337	0.96	180	0.94	176	0.91
7–11	309	0.98	230	0.95	190	0.93
12–16	352	0.94	310	0.95	298	0.86
17–29	229	0.94	209	0.95	206	0.82
30–75+	371	0.96	328	0.95	329	0.87

*Note* For processing speed, the test–retest correlations (AS = 0.93, NI = 0.91, *N* = 120) were used for calculating composite reliabilities for all age groups

two occasions with an average delay of 7 days. The sample included 50.3% female, 10.1% African-American, and 3.4% Hispanic individuals mostly from the East and South regions of the United States. A good mixture of education levels was obtained (years of schooling completed; by parents for children and adolescents) including 10.6% less than High School, 29.8% High School or GED, 27% some college or post-secondary, and 32.6% college or advanced degree. The sample also included 34 students with conditions requiring special education services and 4 students from gifted programs. The test–retest reliabilities based on the 149 individuals in the total group who had complete data ranged from 0.74 to 0.93.

### Decision-Consistency Reliability for the Supplemental A/M Subtests Scores

The supplementary scores for the A/M set of subtests are largely diagnostic or “error” scores reflecting incorrect responses by examinees on the AS, AD, or NS subtests. The skewed distributions of error scores in psychology and education often are not described well by conventional statistics such as means, standard deviations, or correlations because of the large number of zero or low scores (Guilford and Fruchter 1978, p. 56). Because of skewness, conventional reliability indexes (e.g., test–retest correlations) do not accurately represent the

consistency of these scores. Also, due to the preponderance of zero scores, the decision-consistency methods used for criterion-referenced tests (Berk 1984) also give underestimates because of violations of “cell size” (e.g., very small numbers of examinees who have multiple errors) for statistics such as chi-square or kappa. For these reasons, the indexes of consistency (e.g., percentage of correct decisions) used in decision–classification analysis were calculated for each of the scores. The indexes estimate consistency between first and second testings based on test–retest data.

Index percentages show the consistency of test–retest data for examinees who obtained “average or better scores” (such as zero errors) versus the “clinically meaningful scores” (such as multiple errors). Consistency percentages were defined by cut-off values based on the distributions of standardized scaled scores for each variable. Cut-off scores (expressed as scaled score values whereon SD below average equals 7) were determined from the full standardization sample (*N* = 1603) and then applied to the data in the test–retest sample of 156. Consistency ranged from a percentage of 78.9% (AD, Level 2—ages 6–10) to a high of 94% for the NS incorrect scores using a cut-off score of “less than 7.” The data showed a high degree of zero-incorrect scores obtained on both test and retest administrations. Inconsistent results were often due to practice effects (more errors on the first administration of the test and fewer on retest).

## Validity

Validity is a unitary concept that includes all the sources of evidence that support the interpretations of test scores from a given test and its suggested purposes. According to the technical standards for educational and psychological tests (AERA, APA, NCME 1999), the categories of evidence for validity of a test such as Leiter-3 would include evidence based on (a) test content, (b) response processes, (c) internal structure, (d) relations to other variables, and (e) consequences of testing. This section of the chapter will review the highlights of the extensive evidence for Leiter-3 as documented more completely in the Leiter-3 test manual (Roid et al. 2013, pp. 137–168). Importantly, the validity of interpretations (e.g., identifying intellectual deficiency) with Leiter-3 stand on evidence collected for the current third edition, but also on the long history of studies for the original Leiter (Leiter 1979; Levine 1982) and the Leiter-R (Roid and Miller 1997). The historical data is relevant because all of the editions are measuring the construct of nonverbal intellectual ability and many of the features, and even a number of test items are very similar across editions. The reader desiring more complete information on all validity studies since the 1940s is referred to these previous publications and independent studies published in professional journals on the three editions.

**Evidence Based on Test Content.** Leiter-3 was constructed on a model similar to the Cattell–Horn–Carroll (CHC) theory (Flanagan et al. 2013). Content was selected from the previous edition (Leiter-R) to match two major factors of CHC theory—Fluid Reasoning (Gf) and Visual-Spatial (Gv) for the assessment of nonverbal intellectual ability. Leiter-R Full IQ had correlated 0.86 with the WISC-III Full-scale IQ, confirming that the selected subtests would measure the IQ construct. Also, the A/M composite scores for Processing Speed (attention subtests AS and NS) and Memory (FM and RM) were selected to measure the corresponding CHC factors (Gs and Gsm). Cross-battery correlation studies verified that the Leiter-3 general ability

Growth score (consisting of fluid reasoning subtests CA and SO and visual-spatial subtests FG and FC) was correlated significantly with the corresponding Woodcock-Johnson (WJ-III Cognitive Tests, (Mather & Woodcock, 2001) W-scores. Correlations ranged from 0.77 to 0.92 with median 0.85 between Leiter-3 and WJ-III CHC factors ( $N = 26$ , ages 5–67, median age 11 years, diverse ethnic backgrounds). Other evidence is presented in the section on “Evidence Based on Relations to Other Variables.”

**Evidence Based on Response Processes.** Gregory (1996) stated that validity evidence based on content (or response processes in this case) is determined by the degree to which the tasks or items on a test are representative of the universe of behavior the test was designed to sample (p. 108). Several types of systematic sampling of item content and item writing were employed (Roid and Haladyna 1982). Also, extensive item analyses were conducted including examination of the “fit” of each item to the unidimensional construct underlying the subtest (Bond and Fox 2007), and conventional evidence of item quality (e.g., difficulty at various age levels, item versus total correlations, validity-group differentiation) was obtained during the development of Leiter-3. For example, to measure a nonverbal and nonvocal intellectual ability, the mode of response required of examinee is of vital importance. The Leiter tradition of response mode, since the early versions of the original Leiter in the 1940s, was for the examinee to move blocks (printed with test stimuli on their top side) into slots in a response frame (made of wood in early versions and in plastic in Leiter-3). With the examiner using pantomime instructions to the examinee, the block-in-frame mode of response was clearly nonverbal (except for “subvocal speech”) and nonvocal—a perfect match to the construct being assessed. The block-and-frame mode of response is also valuable for children and adults that benefit from “hands on” testing materials. For the Figure Ground and Form Completion subtests, the examinee points from the response card to the target object on the stimulus booklet a highly effective and reliable response from the Leiter-R.

Also, to assure consistency with nonverbal response processes, Teaching items were employed at the starting points of each subtest. Teaching items ensure and verify that individuals would clearly understand each task prior to completing each type of item. Additionally, the use of two subtests for each CHC factor (CA and SO for Gf, FG and FC for Gv) assured that the main factors in the nonverbal IQ score were well sampled with the block-and-frame or “touch-the-picture” mode of response.

**Evidence Based on Internal Structure.** Both exploratory and confirmatory factor analyses were conducted on the standardization data for Leiter-3. Several types of extraction and rotation methods were explored for the data and the clearest factor matrices across age groups (ages 3–6 with  $N = 148$ ; 7–11 with  $N = 175$ , 12–16 with  $N = 291$ , 17–29 with  $N = 201$ , and 30–75+ with  $N = 309$ ) showed four factors. The patterns of factor loadings for each of the age groups were highly similar (Roid et al. 2013, pp. 157–158). To be concise, Table 8.3 shows the results for the age groups 7–11 and 17–29. The younger children show a tendency to have AS load with the memory subtests and for the Stroop subtest scores to split between factors. The patterns of

loadings for older teens and younger adults are quite representative of ages 17–75+ where four factors are clearly defined, although the Stroop Effect score tends to be alone as a “singleton” defining a smaller factor (Gorsuch 1983).

In addition to the exploratory analyses, a series of confirmatory factor analyses were conducted for each of the five age groupings. The singleton Stroop Effect variable was excluded for clarity and because it violated the rule of having at least two variables defining a factor in the computer program LISREL (Joreskog and Sorbom 1999). Results showed the three factor solutions provided the preferred lower values defined by Browne and Cudeck (1993). The three factors were labeled General Cognitive Ability, Nonverbal Memory, and Processing Speed. The best-fit indexes showed these 3-factor models to have the lowest values on several measures including chi-square per degrees of freedom (chi/df) and root mean-squared error of approximation (RMSEA). The range of Chi-square per degree of freedom values was 5.38–7.77 for the baseline 1-factor model and 1.07–2.73 for the 3-factor model across the five age groups. For the RMSEA index, the values ranged from 0.11 to 0.14 for the 1-factor model and 0.02–0.07 for the

**Table 8.3** Exploratory factor analysis of 11 Leiter-3 profile scores

Name of subtest	Factor loadings for ages 3–6				Factor loadings for ages 17–29			
	“g”	Memory	Attn	Stroop	“g”	Memory	Attn	Stroop
Subtests								
Classification/analogy	<b>0.43</b>	0.32			<b>0.46</b>	0.23		
Form completion	<b>0.92</b>				<b>0.48</b>	0.25		
Figure ground	<b>0.50</b>	0.26			<b>0.74</b>			
Visual patterns	<b>0.35</b>			–0.23	<b>0.62</b>			
Sequential order	<b>0.35</b>	0.28			<b>0.70</b>	0.24		
Forward memory	0.36	<b>0.58</b>			0.30	<b>0.59</b>		
Reverse memory		<b>0.69</b>			0.28	<b>0.78</b>	0.28	
Stroop effect				<b>0.95</b>				<b>0.99</b>
NS congruent			0.97		0.22	0.22	0.84	
NS incongruent			<b>0.77</b>	–0.64	0.27	0.25	<b>0.88</b>	
Attention sustained	0.25	0.40	<b>0.28</b>				<b>0.50</b>	

*Note* Factor loadings with near-zero values (–0.20 to +0.20) excluded for clarity of factor patterns. Subtests intended to measure a construct (factor) are shown in boldface type



3-factor model. More detailed analyses are included in the manual (Roid et al. 2013).

Finally, in terms of internal structure, all subtests in the three composite scores (IQ, Nonverbal memory, and Processing Speed) were examined for general ability (“g”) loading, and subtest general, specific, and error variance. Among the nine subtests examined, the “g” loadings ranged from 0.38 for the Stroop Incongruent score to 0.69 for Form Completion. In the tradition promoted by Kaufman (1990), these g-loadings were estimated from the first unrotated factor loading (Jensen 1980) in an exploratory factor analysis. According to the tradition described by Sattler (2001), the variance components of subtests are derived from factor analyses and data on the reliability of subtests. General variance is derived from the communality of a subtest based on the sum of squared loadings on the common factors for a particular subtest. Specific variance is obtained by subtracting the general variance from the reliability (e.g., internal consistency) index for that subtest. The remaining variance is considered error. Again, in the tradition described by Sattler (2001), the ideal pattern of variance should show general variance highest and specific variance higher than error variance. The average pattern of variance values for Leiter-3 (averaged across age groups) was 45, 37, and 18, close to the ideal pattern.

**Evidence Based on Relationships with Other Variables.** Several other well-known intellectual-ability tests were correlated with Leiter-3 in the standardization studies. First, the previous edition, Leiter-R, correlated well with

the third edition, for a sample of 60 individuals, ages 3–87 (median 20) as shown in Table 8.4. All the validation samples represented in Table 8.4 had a balance of genders but wide ranges of ethnicity, race, and educational level (see Roid et al. 2013 for more details of each sample). From comparisons of Leiter-R and Leiter-3, Growth scales correlated 0.89, and IQs, 0.78. The Leiter-3 and Stanford-Binet Fifth Edition (SB5, Roid 2003) were compared for a sample of 26 individuals, ages 4–35 with a median of 8 years. Nonverbal IQ scores from the Leiter-3 and SB5 correlated 0.77 with means of 95.9 and 103.8, respectively (perhaps explained partially by the 10-year difference in standardization dates, estimated to be 0.30 IQ points higher per year, by Flynn 1987, 2012). The item-response theory scores (Change-Sensitive Nonverbal score in SB5 and Growth score in Leiter-R) were correlated 0.85 with means of 493.3 and 499.3, very similar given standard deviations of 16–27.

The Woodcock-Johnson Cognitive Abilities Tests (WJ-III, Mather and Woodcock 2001) was administered to 26 individuals, ages 5–38 (median 11), along with the Leiter-3. The most relevant score measuring fluid reasoning in the WJ-III was found to correlate 0.74, as shown in Fig. 11 (Roid et al. 2013). The WJ-III W-score for the Fluid Reasoning factor-score cluster correlated 0.92 with the Leiter-3 Growth score. The two prominent Wechsler scales available at the time of the standardization (WISC-IV and WAIS-IV) also correlated 0.73 and 0.72 with the Leiter-3 IQ score, on samples of 50 and 53 individuals, respectively.

**Table 8.4** Correlations between Leiter-3 IQ scores and related intellectual-ability scales

Intellectual-ability scales	Leiter-3 IQ score	Leiter-3 growth score
Leiter-R nonverbal IQ	0.78	–
Leiter-R growth score	–	0.89
SB5 nonverbal IQ	0.77	–
SB5 change-sensitive NV scale	–	0.85
WJ-III fluid reasoning cluster	0.74	0.92
WISC-IV perceptual reasoning index	0.73	–
WAIS-IV perceptual reasoning index	0.72	–

Related Stroop measures were also correlated with the Stroop scores of the Leiter-3 (see Fig. 12, Roid et al. 2013). The traditional Stroop Color-Word Test (SCWT, Golden et al. 2003) requires the examinee to inhibit the word and respond to the color. In the NS task, the examinee must identify one color while inhibiting or ignoring the second color. Even though there is an underlying similarity between the tasks, the question of whether or not the nonverbal version is related to the color-word version is important. To address this concern, a sample of 75 college students ( $M_{\text{age}} = 21.12$ ) completed the Stroop Color and Word Test and the NS test from the Leiter-3. The order of the two tests was randomly varied across participants. Correlations between both tests were statistically significant providing evidence of concurrent validity, and ranged from 0.24 to 0.62.

Attention-related deficits, including inhibition, have been associated with reading disabilities (Ackerman et al. 1986; Kelly et al. 1989). Fifty-four individuals from the standardization sample had been diagnosed with a reading-related learning disability. Compared to the typically developing sample, these individuals scored significantly lower on both the congruent ( $t(1264) = 4.82, p < 0.001, d = 1.79$ ) and incongruent ( $t(1264) = 6.32, p < 0.001, d = 1.71$ ) NS tasks, providing evidence for the construct validity of the nonverbal task.

**Evidence Concerning Consequences of Testing with the Leiter-3.** Three lines of evidence for the effectiveness and fairness of the Leiter-3 were presented in detail in Roid et al. (2013): (a) Effective separation or fairness of differences among mean scores between typical and atypical individuals (e.g., those with disabilities), (b) Low error rates for using Leiter-3 scores for serious decisions such as identifying intellectual deficiency, and (c) fairness and lack of differential item functioning (DIF, Holland and Wainer 1993) among ethnic and racial groups.

Studies of mean scores for 11 specialized groups of individuals (e.g., those with deafness, traumatic brain injury, intellectual deficiency, ADHD, or giftedness) were reported in detail in the Leiter-3 test manual (Roid et al. 2013).

Individuals with medically diagnosed traumatic brain injury ( $N = 28$ , ages 5–87 with median 21) had mean scores ranging 4.6–6.8 among the cognitive subtests and IQ mean of 77.6, as expected. The validation group for intellectual deficiency (as diagnosed with a history of special education) showed subtest score means as low as 2.5 on Form completion and an IQ mean of 71 ( $N = 47$ , ages 3–35, mean 13). Similar patterns of expected mean scores were found for individuals with learning disabilities, ADHD, and autism spectrum disorder.

The classification accuracy of Leiter-3 IQ scores was examined by comparing a sample of 53 individuals with diagnosed intellectual deficiency to a random sample of 500 typical cases (with no diagnosed conditions, medical, emotional, or educational) from the normative sample. Total correct identification (“hit rate”) ranged from 95.4% (using the traditional standard of IQ less than or equal to 70) to 97.1 (using 75 as the standard). False negative rates (classified typical when actually atypical) ranged from 6.6% using a standard of 65–2.4% using a standard of 75. False positive rates (classifying as atypical when truly typical) were all extremely low (0.2%). Similar results were found when using the Nonverbal Memory and Processing Speed composite scores to classify individuals with ADHD using cut-off scores of 80 to 90 (Hit rates 93.0–95.1%, false negative 3.0–3.6%, and false positive rates of 1.3% (using scores of 80–85).

In terms of fairness of measurement, several group-mean studies showed similarity of special groups to the normative sample. For example, a sample of 46 (ages 3–66) individuals with deafness or hard-of-hearing conditions showed mean scores very near 10 (the average score) for each of the cognitive subtests and near the average of 100 for IQ (97.2), indicating fairness of measurement for this group. Also, two groups of individuals with dominant spoken language other than English were administered the Leiter-3. Those with Spanish language dominance showed cognitive subtest score means ranging from 9.8 to 11.0 and an IQ mean of 101.9 ( $N = 22$ , ages 5–47, median 15). A second group with a variety of non-English language dominance (although 63% Asian

language) showed mean scores of 10.2–12.5 on cognitive subtests and a mean IQ of 105.8, as often found in samples of individuals with Asian educational backgrounds (e.g., Roid 2003).

Finally, extensive studies of item differential functioning (DIF, potential item bias) were conducted on contrasting samples of gender and ethnicity/race (Holland and Wainer 1993). All items within the five cognitive subtests used in the composite IQ score for Leiter-3 were calibrated using the WINSTEPS (Lincacre and Wright 2000) program for the one-parameter logistic (Rasch 1966, 1980) model, separately in normative and contrast groups. The goal was to explore the predicted item-difficulty invariance between groups (Bond and Fox 2007) by plotting the difficulties of items on graphic scatter plots (expecting a pattern of difficulties aligning together on a 45-degree line, indicating invariance). The calibrations of items were conducted on relatively large groups (as shown in Figs. 13 and 14, Roid et al. 2013) including 197 African-American, 248 Hispanic, and 1040 Caucasian non-Hispanic individuals. Out of a total of 152 items, only two items (introductory Teaching items calibrated on Anglo versus Hispanic samples) showed slight departures from the linear trend in the scatter plots. Similar analyses were conducted on items in the Nonverbal Memory subtests with similar results. Items in the Processing Speed subtests are “speeded” (timed subtests) and could not be used in this type of DIF study. Thus, evidence is compelling to conclude that the Leiter-3 is quite free from DIF, potential bias, in measurement of nonverbal intellectual ability and nonverbal memory.

### Deriving Norm-referenced Standardized Scores

**Raw Scores.** General directions for recording and scoring the individual on each subtest are provided in the manual. All Leiter-3 items are easy to administer and score; however, some subtests, such as AS and AD, require special scoring directions. For most of the subtests, rules for obtaining raw scores for items are

straightforward. Each subtest has a stop rule (e.g., stop after 5 cumulative errors). When the stop rule is reached, testing is terminated for that subtest. Each item response that is correct is counted as 1 raw score point. Incorrect item responses receive no credit. Some subtests use special procedures that differ from the general patterns for determining a subtest raw score, i.e., subtests contained in the A/M Battery. Two subtests have slightly more involved rules for obtaining subtest raw scores (Attention Sustained and or Attention Divided). The availability of very easy Teaching Trial items for each age group should assist examiners in identifying individuals who understand the nature and the expectations of a subtest. Also, the Teaching Trial items (where as many as three trials are allowed) are very useful in providing some degree of measurement or a “basal” for with cognitive delays. Since some children may have difficulty with the Teaching items it is not uncommon for an individual to initially provide two or more responses.

---

### Leiter-3 Interpretation

The following discussion regarding interpretation of the Leiter-3 is based upon the recommendations presented in the manual (Roid et al. 2013). The test authors recommend a hierarchical method of interpretation that begins with a thorough review of developmental, clinical, and academic history, presenting concerns, and information collected via rating scales. Next, score examination begins with the most global estimates and proceeds to more specific estimates provided by the Growth scale scores, Composites, and Subtest scores. For special interpretations and adaptations of the Leiter-3 results for individuals with deafness or hard-of-hearing conditions, please consult Appendix K of the Leiter-3 manual (Hardy-Braz 2013).

The first stage of interpretation involves examining the global estimates of nonverbal intelligence as represented by the Full-scale IQ and the Brief IQ screener of the VR Battery. Such global estimates reflect the definition of

intelligence as measured on the Leiter-3 and is defined as “the general ability to perform complex nonverbal mental manipulations related to conceptualization, inductive reasoning, and visualization” (Roid and Miller 1997, p. 103). Although attention, speed of processing, and memory are regarded as pre-cursors or substrates of cognitive performance, they are also highly associated with disorders such as attention-deficit and other neuropsychological processing deficits (Hale et al. 2012). For these reasons, subtests of the AM Battery are not included in the nonverbal IQ. Instead, the AM subtests serve as “rule out” measures to assist in interpreting the effects of processing disorders on the level of global IQ scores. Performance is interpreted using standard scores, confidence intervals, percentile ranks, and descriptive classification.

The examination of Growth scores, especially for individuals who function at a low level of ability and/or are expected to be re-tested on the Leiter-3, is the next stage of interpretation. Designed using item-response theory (IRT, Lord 1980; Hambleton et al. 1991; Bond and Fox 2007), Growth scores provide an opportunity to measure small increments of growth (or decline) along the continuum of general ability as measured by the Leiter-3 (Roid and Woodcock 2000; Woodcock 1999).

The next stage of interpretation involves examination of the Composite scores of the cognitive and attention/memory batteries. The Cognitive Battery has one main composite—Nonverbal IQ. The other battery contains two composites—A/M. The Examiner Rating scales each contain two composites: Cognitive/Social and Emotional/Regulation. Composite scores, Confidence intervals, percentile ranks, and classifications are provided in the manual to describe performance.

Leiter-3 subtests were designed to measure unitary constructs, are reliable, and unbiased based on gender, race, ethnicity, and socioeconomic factors. Thus, interpretation of performance and abilities at the subtest level is supported and is the next stage of interpretation.

**Interpretation of Nonverbal Stroop Results.** Because the Stroop task is often absent from major Cognitive Batteries, it is important to

explain some ideas for the interpretation of this unique new subtest. It is commonly known that attention tests, such as the Stroop, are useful for identifying processing differences but are not particularly useful for specifying a condition or disorder. Therefore, it is necessary to use Stroop scores in conjunction with other scores when making a diagnosis. Furthermore, the strong reliability of the congruent and incongruent scores and moderate reliability of the interference score suggests that it may also be beneficial to include all three scores in the clinical decision-making process. The congruent score can be viewed as an indicator of naming speed (cf., Carroll 1993) while the incongruent score may represent naming or processing speed with noise. The Stroop effect, or interference, score may represent the ability to inhibit or selectively ignore meaningful but irrelevant information.

In the case of reading-related learning disabilities, a stepwise logistic regression analysis was conducted on the Leiter-3 standardization data to determine the test scores associated with a learning disability for reading. This analysis included all three Stroop scores (i.e., congruent, incongruent, and Stroop effect). The results indicate that a model including classification/analogies, AS, NS congruent, and figure ground scores best fit the data ( $X^2(4) = 35.43, p < 0.001$ ). Therefore, an attention task like the NS task may be useful for ruling in an attention-related cognitive disorder but should be combined with other construct-related measures to determine the specific disorder. Additionally, the relevant measure from the Stroop task may vary across disorders.

---

## Strengths and Limitations of the Leiter-3

Based on the evidence presented in this chapter, the Leiter-3 has several positive strengths and few negative weaknesses. The main strength of the Leiter-3 would be the true nonverbal character of the test, requiring no spoken directions by the examiner and no spoken responses by the examinee. In addition, the subtests have a great

deal of color, are game-like, and are more engaging than some of the other nonverbal instruments that use primarily black-and-white illustrations without manipulatives such as the blocks used in the Leiter-3 cognitive subtests. Also, the Leiter-3 provides a nonverbal IQ with just four subtests, and, for differential analysis of cognitive process deficits, a companion set of memory and attention nonverbal subtests that are optional. The technical qualities are strong in terms of reliability, evidence of validity, decision-consistency accuracy, and fairness, as presented in this chapter and, more extensively in the test manual (Roid et al. 2013). One of the reviews in *Buros Mental Measurements Yearbook* (online at [www.Buros.org](http://www.Buros.org) by Martin Wiese, Buros Center for Testing 2014) included several positive comments about the test, stating “The Leiter-3 authors have succeeded in their goal of constructing a reliable and valid nonverbal measure of intellectual ability and Attention/Memory.”

Weaknesses are few based on the reviews to date (e.g., Buros reviews by Wiese and Ward, Buros Center for Testing 2014) but include the fact that the global nonverbal IQ score is not supplemented by factor index scores for fluid reasoning and visual-spatial abilities separately. The Buros review by Susan Ward (Buros Center for Testing 2014) was largely positive but pointed to the relatively small number of subjects in criterion-group studies and lower test–retest coefficients compared to internal consistency estimates. Also, Ward’s review mentioned the need for new users to study the Training DVD (available from [www.stoeltingco.com](http://www.stoeltingco.com)) given the challenges of learning different pantomime instructions for most of the subtests. Finally, the authors are aware, and emphasize in training sessions on Leiter-3, that some individuals with disabilities (e.g., Autism) remain quite verbal and this may need additional verbal encouragement between subtests.

---

## Summary

The Leiter-3 is a totally nonverbal, individually administered test battery that does not require

spoken directions by the examiner or vocal responses by the examinee. The wide age range (from 3 years to 75+) and combination of nonverbal general intellectual ability (IQ) and nonverbal A/M subtests makes the Leiter-3 stand out among alternative nonverbal batteries. Also, the addition of a NS effect (Stroop 1935; Golden 1976) subtest within the battery allows the assessment of more neuropsychological factors than other nonverbal batteries. Hands-on movement of blocks and cards for children with autism or other attention-impaired conditions have proven to be more engaging as a “game-like” approach to testing (M. Wiese, Buros Review online 2014).

Based on user requests, the “block and frame” response mode was restored in the third edition (where the examinee moves printed blocks into alignment with printed illustrations on the Easel attached to the frame). At the same time, the successful features and many items from the second edition (Leiter-R, Roid and Miller 1997) were retained in the new format or intact with updated stimulus materials. The scoring system for the Leiter-3 is largely unchanged from the Leiter-R and matches the style of scaled score and composite score methods (and metric) of the other instruments measuring IQ (e.g., Stanford-Binet Fifth Edition and Wechsler scales). A supplemental scoring method called “Growth Scores” is based on item-response theory (IRT) and very similar in metric to those of the Stanford-Binet 5 Change-Sensitive scores (Roid 2003) and Woodcock-Johnson W-score metrics. These IRT (Rasch Model) scores function as criterion-referenced and developmental scales helpful in tracking intellectual abilities across many years of follow-up for children and adults with continuing or chronic conditions.

The technical qualities of the Leiter-3, with extensive reliability, validity, classification accuracy, and fairness studies (summarized briefly in this chapter and more extensively in the Leiter-3 test manual), have been praised by recent reviews in the Buros Mental Measurement series (S. Ward and M. Wiese reviews online, 2014). Each area of evidence for validity (content, response processes, internal structure,

relationships with other variables, and consequential validity) has been highlighted in this chapter and will surely be supplemented by more recent, independent research studies.

Finally, the Leiter-3 provides examiners with a modern, nonverbal test battery for relatively quick assessment of IQ by an individually administered instrument using a game-like format with only four subtests'. The Leiter-3 assessment of a wide range of cognitive deficits, delays, autism, learning disabilities, ADHD, or brain-injury cases as well as a fair assessment for non-English speakers and those with deafness and hard-of-hearing conditions.

## References

- Ackerman, P. T., Anhault, J. M., Dykman, R. A., & Holcomb, P. J. (1986). Effortful processing deficits in children with reading or attention disorders. *Brain and Cognition*, *5*, 22–40.
- Albert, M. L. (1973). A simple test of visual neglect. *Neurology*, *23*, 658–664.
- American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (1999). *Standards for educational and psychological testing*. Washington, DC: APA.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-V-TR)* (5th ed.). Washington, DC: Author.
- Banken, J. A. (1985). Clinical utility of considering digits forward and digits backward as separate components of the Wechsler Adult Intelligence Scale-Revised. *Journal of Clinical Psychology*, *41*, 686–691.
- Berk, R. A. (1984). *A guide to criterion-test construction*. Baltimore, MD: Johns Hopkins University Press.
- Bishop, S. L., Richler, J., & Lord, C. (2006). Association between restricted and repetitive behaviors and non-verbal IQ in children with ASD. *Child Neuropsychology*, *12*, 247–267.
- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ: Routledge/Erlbaum.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.
- Buros Center for Testing. (2014). *Wise review of Leiter-3*. Lincoln, NE: Buros Institute, University of Nebraska-Lincoln. [www.buros.org](http://www.buros.org)
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.
- CDC (Centers for Disease Control). (2007). *Press release: CDC releases new data on autism spectrum disorders (ASDs) from multiple communities in the United States*. Retrieved September 6, 2007, from the Centers for Disease Control Web site: <http://www.cdc.gov/od/oc/media/pressrel/2007/r070208.htm>
- Cha, C. B., Najmi, S., Park, J. M., Finn, C. T., & Nock, M. K. (2010). Attentional bias toward suicide-related stimuli predicts suicidal behavior. *Journal of Abnormal Psychology*, *119*, 616–622.
- Christensen, A. L. (1979). *Luria's neuropsychological investigation* (2nd ed.). Copenhagen: Munksgaard.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334.
- Dalrymple-Alford, E. C., & Budayr, B. (1966). Examination of some aspects of the Stroop color-word test. *Perceptual and Motor Skills*, *23*, 1211–1214.
- Dee, H. L. (1970). Visuoconstructive and visuoceptive deficit in patients with unilateral cerebral lesions. *Neuropsychologia*, *8*, 305–314.
- Dunbar, K. N., & MacLeod, C. M. (1984). A horse race of a different color: Stroop interference patterns with transformed words. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 622–639.
- Dyer, E. N. (1973). The Stroop phenomenon and its use in the study of perceptual, cognitive, and response processes. *Memory and Cognition*, *1*, 106–120.
- Dyer, F. N., & Severance, L. J. (1973). Stroop interference with successive presentations of separate incongruent words and colors. *Journal of Experimental Psychology*, *98*, 438–439.
- Elliott, C. D. (2008). *Differential Ability Scales—Second Edition (DAS-II)*. San Antonio, TX: Pearson/Psychological Corp.
- Flanagan, D. P., Ortiz, S. O., & Alfonso, V. (2013). *Essentials of cross-battery assessment* (3rd ed.). New York: Wiley.
- Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, *101*, 171–191.
- Flynn, J. R. (2012). *Are we getting smarter? Rising IQ in the 21st century*. Cambridge, UK: Cambridge University Press.
- Golden, C. J. (1976). The diagnosis of brain damage by the Stroop test. *Journal of Clinical Psychology*, *32*, 654–658.
- Golden, C. J., Freshwater, S. M., & Golden, Z. (2003). *Stroop color and word test, children's version for ages 5–14*. Wood Dale, IL: Stoelting Co.
- Goldstein, G., Johnson, C. R., & Minshew, N. J. (2001). Attentional processes in autism. *Journal of Autism and Developmental Disorders*, *31*, 433–440.
- Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Gottschaldt, K. (1928). Hidden figures test. In G. Talland (Ed.), *Deranged memory* (p. 1965). New York: Academic Press.
- Gregory, R. J. (1996). *Psychological testing* (2nd ed.). Boston, MA: Allyn & Bacon.

- Guilford, J. P., & Fruchter, B. (1978). *Fundamental statistics in psychology and education*. New York, NY: McGraw-Hill.
- Gustafsson, J. E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence*, 8, 179–203.
- Hale, J. B., Yim, M., Schneideer, A. N., Wilcox, G., Henzel, J. N., & Dixon, S. G. (2012). Cognitive and neuropsychological assessment of ADHD. In D. Flanagan & P. Harrison, *Contemporary intellectual assessment* (pp. 687–707). New York, NY: Guilford Press.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.
- Hardy-Braz, S. (2013). General guidelines for using the Leiter-3 with deaf/hard-of-hearing (Appendix K). In G. Roid, L. Miller, M. Pomplun, & C. Koch (Eds.), *Leiter-3 manual* (pp. 277–278). Wood Dale, IL: Stoelting Co.
- Holland, P., & Wainer, H. (Eds.). (1993). *Differential item functioning*. Mahwah, NJ: Routledge/Erlbaum.
- Hooper, H. E. (1983). *Hooper visual organization test*. Los Angeles, CA: Western Psychological Services.
- Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology*, 57, 253–270.
- Jensen, A. R. (1980). *Bias in mental testing*. New York: The Free Press.
- Joreskog, K. G., & Sorbom, D. (1999). *LISREL 8: User's reference guide*. Chicago: Scientific Software.
- Kahneman, D., & Chajczyk, D. (1983). Tests of the automaticity of reading: Dilution of Stroop effects by color-irrelevant stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, 9, 497–509.
- Kahneman, D., & Henik, A. (1981). Perceptual organization and attention. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 181–211). Hillsdale, NJ: Erlbaum.
- Kaufman, A. S. (1990). *Assessing adolescent and adult intelligence*. Boston, MA: Allyn & Bacon.
- Kelly, M. S., Best, C. T., & Kirk, U. (1989). Cognitive processing deficits in reading disabilities: A prefrontal cortical hypothesis. *Brain and Cognition*, 11, 275–293.
- Klinger, L. G., O'Kelly, S. E., Mussey, J. L., Goldstein, S., & DeVries, M. (2012). Assessment of intellectual functioning in autism spectrum disorder. In D. Flanagan & P. Harrison, *Contemporary intellectual assessment*, 3rd Ed. (pp. 670–686). New York, NY: Guilford Press.
- Koch, C., & Barr, B. (2010). Can Task Demands Override Connectedness? *Abstracts of the Psychonomic Society*, 15, 56.
- Koch, C., & Kubovy, M. (1996, June). Interference in a color-color Stroop task. Poster presented at the *Eighth Annual Convention of the American Psychological Society*, San Francisco, CA.
- Koch, C., Lowen, J., & McWilliams, M. (2011, November). *Developmental differences using a nonverbal Stroop task*. Poster presented at the Object Perception, Attention, & Memory 19th Annual Conference, Seattle, WA.
- Koch, C., & Roid, G. (2012). *Manual for the nonverbal Stroop card sorting task*. Wood Dale, IL: Stoelting Company.
- Kuschner, E. S., Bennetto, L., & Yost, K. (2007). Patterns of nonverbal cognitive functioning in young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 37(5), 795–807.
- Leiter, R. G. (1938). *A comparative study of the general intelligence of Caucasian and Asian children as measured by the Leiter international performance scale*. Unpublished doctoral dissertation, University of Southern California, Los Angeles, CA.
- Leiter, R. G. (1979). *Instruction manual for the Leiter international performance scale*. Wood Dale, IL: Stoelting Co.
- Levine, M. N. (1982). *Leiter international performance scale: A handbook*. Los Angeles, CA: WPS.
- Lezak, M. D. (1995). *Neuropsychological assessment* (3rd ed.). New York: Oxford University Press.
- Lincacre, J. M., & Wright, B. D. (2000). *WINSTEPS v. 3.00: Rasch item analysis computer program manual*. Chicago: MESA Press.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Mahwah, NJ: Erlbaum.
- Mather, N., & Woodcock, R. W. (2001). *Woodcock-Johnson tests of cognitive ability third edition: Technical manual*. Chicago: Riverside.
- Mayes, S. D., & Calhoun, S. L. (2003). Ability profiles in children with autism. *The National Autistic Society*, 6, 65–80.
- Mayes, S. D., & Calhoun, S. L. (2004). Influence of IQ and age in childhood autism: Lack of support for DSM-IV Asperger's disorder. *Journal of Developmental and Physical Disabilities*, 3, 257–272.
- Minshew, N. J., & Goldstein, G. (1998). Autism and a disorder of complex information processing. *Mental Retardation and Developmental Disabilities*, 4, 129–136.
- Minshew, N. J., & Williams, D. L. (2007). The new neurobiology of autism: Cortex, connectivity, and neuronal organization. *Neurological Review*, 64, 945–950.
- Mundy, P., & Crowson, M. (1997). Joint attention and early social communication: Implications for research on intervention with autism. *Journal of Autism and Developmental Disorders*, 27, 653–675.
- Nelson, H. E. (1976). A modified card sorting test sensitive to frontal lobe defects. *Cortex*, 12, 313–324.
- Nordlund, C. B. (1998). *An examination of behavior ratings and rater differences of ADHD subjects on the Leiter-R rating scales*. Unpublished doctoral dissertation, George Fox University, Newberg, Oregon.
- Nunnally, J. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.
- Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 55–85). Hillsdale, NJ: Erlbaum.

- Rasch, G. (1966). An item analysis which takes individual differences into account. *British Journal of Mathematical and Statistical Psychology*, 19, 49–57.
- Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago, IL: University of Chicago Press (Translated from the Danish original of 1960).
- Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's progressive matrices and vocabulary scales*. Oxford, UK: Oxford Psychologists Press.
- Reynolds, C. R. (1997). Forward and backward memory span should not be combined for clinical analysis. *Archives of Clinical Neuropsychology*, 12, 29–40.
- Roid, G. H. (2003). *Stanford-Binet intelligence scales—Fifth edition, examiner & technical manuals*. Austin, TX: Pro-Ed.
- Roid, G. H., & Haladyna, T. M. (1982). *A technology for test-item writing*. New York: Academic Press.
- Roid, G. H., & Miller, L. J. (1997). *Leiter international performance scale—Revised*. Wood Dale, IL: Stoelting.
- Roid, G. H., Miller, L. J., Pomplun, M., & Koch, C. (2013). *Leiter international performance scale-third edition*. Wood Dale, IL: Stoelting Company.
- Roid, G. H., Pomplun, M., & Martin, J. (2009). Nonverbal cognitive assessment with the Leiter-R. In J. Naglieri & S. Goldstein (Eds.), *A practitioner's guide to assessment of intelligence and achievement*. NY: Wiley.
- Roid, G. H., & Woodcock, R. W. (2000). Uses of Rasch scaling in the measurement of cognitive development and growth. *Journal of Outcome Measurement*, 4(2), 579–594.
- Rosinski, R. R., Golinkoff, R. M., & Kukish, K. S. (1975). Automatic semantic processing in a picture-word interference task. *Child Development*, 46, 247–253.
- Rourke, B. P. (1988). The syndrome of non-verbal learning disabilities. *Clinical Neuropsychologist*, 2, 293–330.
- Salvia, J., & Ysseldyke, J. E. (1991). *Assessment*. Boston, MA: Houghton Mifflin.
- Sattler, J. M. (2001). *Assessment of children: Cognitive applications* (4th ed.). La Mesa, CA: Author.
- Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. In D. Flanagan & P. Harrison, *Contemporary intellectual assessment*, 3rd Ed. (pp. 99–144). New York, NY: Guilford Press.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643–662.
- Sullivan, E. V., Sagar, H. J., & Gabrieli, J. D. (1989). Different cognitive profiles on standard behavioral tests in Parkinson's disease and Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology*, 11, 799–820.
- Talland, G. A. (Ed.). (1965). *Deranged memory*. New York: Academic Press.
- Teccie, J. J., & Happ, S. J. (1964). Effects of shock-arousal on a card-sorting test of color-word interference. *Perceptual and Motor Skills*, 19, 905–906.
- Thurstone, L. L., & Thurstone, T. G. (1962). *Primary Mental Abilities* (Rev.). Chicago, IL: Science Research Associates.
- U.S. Bureau of Census. (2009). *Population statistics: 2008 update*. Washington, DC: Author.
- U.S. Bureau of Census. (2011). *Census 2010 summary file 1 of the United States*. Washington, DC: Author.
- U.S. Department of Education. (2016). National Center for Education Statistics, ED*Facts* file 141, Data Group 678, extracted May 13, 2016, from the ED*Facts* Data Warehouse (internal U.S. Department of Education source); Common Core of Data (CCD), “State Non-fiscal Survey of Public Elementary and Secondary Education,” 2008–09 through 2013–14. [http://nces.ed.gov/programs/digest/d15/tables/dt15\\_204.27.asp](http://nces.ed.gov/programs/digest/d15/tables/dt15_204.27.asp)
- U.S. Department of Education. (2007). *Data analysis system OMB #1820-0043: “Children with disabilities receiving special education under part B of the individuals with disabilities education act”*. Data updated as of July 15, 2008. Washington DC: Author, Office of Special Education Programs. Retrieved from: [https://www.ideadata.org/arc\\_toc9.asp#partBCC](https://www.ideadata.org/arc_toc9.asp#partBCC)
- Volkmar, F. R., Lord, C., Bailey, A., Schultz, R. T., & Klin, A. (2004). Autism and pervasive developmental disorders. *Journal of Child Psychology and Psychiatry*, 45, 135–170.
- Wang, P. L. (1984). *Modified Vygotsky concept formation test manual*. Chicago, IL: Stoelting Co.
- Wechsler, D. (1991). *Wechsler intelligence scale for children—Third Edition (WISC-III)*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (2003). *Wechsler intelligence scale for children—Fourth edition*. (WISC-IV) San Antonio, TX: Pearson/PsychCorp.
- Wechsler, D. (2008). *Wechsler adult intelligence scale—Fourth edition (WAIS-IV)*. San Antonio, TX: Pearson/PsychCorp.
- Williams, D. L., Goldstein, G., & Minshew, N. J. (2006). Neuropsychologic functioning in children with autism: Further evidence for disordered complex information-processing. *Child Neuropsychology*, 12, 279–298.
- Woodcock, R. W. (1999). What can Rasch-based scores convey about a person's test performance? In S. E. Embretson & S. L. Hershberger (Eds.), *The new rules of measurement: What every psychologist and educator should know* (pp. 105–128). Mahwah, NJ: Erlbaum.
- Woodcock, R. W., & Dahl, M. N. (1971). *A common scale for the measurement of person ability and test item difficulty*. (AGS Paper No. 10). Circle Pines, MN: American Guidance Service.
- Woodcock, R. W., McGrew, K., & Mather, N. (2000). *Woodcock-Johnson tests of cognitive ability-third edition: Examiner's manual*. Chicago: Riverside.